# DISCUSSIÓ DE RESULTATS: CAPÍTOL 3

# <u>QSAR</u>

# PART I: Conformació activa i QSAR.

# 1. Introducció.

L'activitat dels BRs com a promotors del creixement vegetal, està íntimament relacionada amb la seva estructura. Els estudis de relació estructura-activitat quantitatius (QSAR) pretenen obtenir un model matemàtic que sigui capaç d'explicar aquesta relació. La interpretació d'aquest model permet establir els requisits necessaris per a que els BRs desencadenin la seva activitat. Fins i tot, es pot intentar explicar com es produeix la unió substrat/receptor (*virtual receptor site* o *VRS*). A més a més, el model es pot fer servir per predir l'activitat de nous anàlegs, encara sense sintetitzar.

En el capítol anterior s'ha vist com l'equip ha anat desenvolupant fins a quatre conformacions actives dels BRs (*SQR, IZR, MVP* i *HIP*). A cadascuna d'elles li ha seguit el seu corresponent model QSAR.<sup>\*</sup> El fet de tenir quatre conformacions actives amb els seus quatre corresponents models QSAR, fa que hi hagi quatre maneres diferents de: a) interpretar els requisits necessaris per a que els BRs desencadenin activitat, b) explicar la unió substrat/receptor, i c) predir l'activitat de nous anàlegs.

S'ha vist que les quatre conformacions actives són estructuralment diferents i en alguns casos és difícil establir una relació estructura-activitat *qualitativa* (SAR). No obstant, els models *quantitatius* (QSAR) no són tan diferents com es podria esperar i presenten una bona correlació malgrat les diferències estructurals. Aquest fet es pot justificar, a priori, argumentant que les variacions estructurals es centren en una part reduida de la molècula. Les principals diferències afecten: 1) a l'extrem final de la cadena lateral, el qual representa només el 24% de l'estructura total, i 2) a un dels cinc grups funcionals polars, l'hidroxil en C23.

A continuació es pretén comparar entre si els quatre models QSAR per: a) quantificar fins a quin punt són similars i/o diferents, i b) establir si els models són independents de la conformació activa o no. No obstant, la comparació directa no és la més raonable per diversos motius:

- 1. El nombre de BRs presents en cada model és diferent ja que a mesura que s'han sintetitzat nous BRs s'han incorporat al model.
- 2. Les dades d'activitat dels BRs no són iguals en cada model. Primer, perquè en alguns models s'ha fet servir el " $-\log(dosi_{45^2})$ " i en altres el " $-\log(dosi_{50\%})$ ". Després, perquè al llarg de la tesi l'equip ha revisat l'activitat de la brassinolida i la castasterona (pàgines 213–219).
- 3. En la realització dels models es fa un pretractament de les dades per tal de millorar la seva qualitat. Aquest pretractament està íntimament relacionat amb l'estructura del conjunt de BRs del model (*data set*). Degut a que l'estructura de cadascuna de les conformacions actives és diferent i a que també ho és el nombre de BRs del "*data set*", el pretractament de les dades en cada model és necessariament diferent.

Així doncs, el primer que cal fer és unificar criteris.

<sup>&</sup>lt;sup>\*</sup> El nom que es va assignar a la conformació activa es trasllada als models estadístics i, per exemple, el model *SQR* és el model QSAR realitzat amb la conformació activa *SQR*.

# 1.1. BRs dels models (data set) i dades d'activitat.

Seria molt interessant actualitzar els models fent servir tots els BRs actius que actualment es disposen. Aquest desig, però, no és possible ja que la conformació activa *IZR* actua com a limitant. Recordar que en aquesta conformació el diol de les cadenes laterals dels 22R,23R-BRs presentava una conformació eclipsada amb els hidroxils quasi solapats. Això comportava un problema alhora de construir els nous BRs. Inicialment s'assignava al diol de la cadena lateral la conformació eclipsada, però l'enllaç evolucionava cap a posicions alternades en optimitzar l'energia i la conformació final no coincidia amb la de la seva cadena lateral tipus. Aquest fet obliga a prendre els 20 compostos presents en el model *IZR*. El valor d'activitat s'expressa a través del " $-log(dosi_{50\%})$ " (Taula 8).

BRs	–log(dosi <sub>50%</sub> )	]	BRs	_log(dosi <sub>50%</sub> )
bl	3,86		hbl	3,32
CS	3,33		hbl_cis	2,45
epibl	2,67		hcs	2,78
epics	2,29		hcs_cis	1,41
epics_2b3b	0,37		hcs_2n3a	2,02
epics_2n3n	0,47		hcs_2n3b	1,52
S_hbl	2,05		hcs_60Ha	1,60
S_hcs	1,27		hcs_60Hb	1,70
S_epibl	2,24		hcs_50H	1,71
S epics	1,12	]	hcs eter	0,71

*Taula 8*: Conjunt de BRs presents en els models ("data set").

#### 1.2. Estratègia i pretractament de les dades.

Per tal de garantir el mateix tractament a cadascuna de les quatre conformacions actives, es procedeix de la següent forma (*Figura 128*):

- S'apleguen les 80 molècules (20 BRs per 4 conformacions actives) en un únic fitxer.
- Sobre aquest fitxer es calculen els mapes de GRID fent servir una caixa que engloba les 80 molècules i un pas de malla de mig amstrong.
- El resultat s'importa a GOLPE on es fa el següent pretractament: a) es suprimeixen els valors positius per considerar només les interaccions favorables, b) es realitza un escalat per bloc de les sondes per donar-les el mateix pes, c) s'aplica un tall d'energia de 0,1 kcal/mol per suprimir les variables amb més soroll, d) s'aplica un tall en la desviació estàndard de 0,1 kcal/mol per suprimir les variables amb menys informació, i e) es suprimeixen les variables a 2, 3 i 4 nivells, les quals poden distorsionar el model.
- Primer es realitza un model CONJUNT amb les 4 conformacions actives dels 20 BRs. Posteriorment es separen els compostos segons la seva conformació activa i es realitzen els models SQR, IZR, MVP i HIP.



Figura 128: Esquema del procés d'obtenció dels models.

Per cada model es realitza: 1) un anàlisi de components principals (PCA) on s'estudien les semblances i diferències estructurals entre BRs, i 2) un anàlisi de mínims quadrats parcials (PLS) per correlacionar l'estructura amb l'activitat i obtenir el model QSAR.

# 2. Model CONJUNT.

# 2.1. Anàlisi de Components Principals (PCA).

#### a) Agrupació segons les conformacions actives.

Els tres primers components principals expliquen amb un 34% de la variància, les quatre conformacions actives (*Figura 129*). S'observa com la conformació activa *SQR* és la més diferent de totes. Per la resta, s'observa una gradació en la qual les conformacions *IZR* i *HIP* són les més diferents entre si, i la conformació *MVP* queda enmig. Cal destacar que hi ha dos compostos que malgrat tenir la conformació activa *IZR*, es troben en el grup dels confórmers *SQR*. Aquest són els dos 22*S*,23*S*-homoBRs (S\_hbl i S\_hcs), en els quals no hi ha gaire diferència estructural entre les conformacions *SQR* i *IZR* (veure *Figura 87* i *Figura 88*, pàgina 106).



Figura 129: Gràfic d'objectes dels 3 primers components principals del model CONJUNT.

El primer component principal posa de manifest la diferència existent entre la conformació activa *SQR* i la resta. Aquesta es deguda a: 1) l'hidroxil en C23 s'orienta en el pla de l'esquelet enlloc de fer-ho cap a la cara  $\alpha$ , i 2) l'extrem final s'orienta cap a la cara  $\alpha$ .

El segon component principal explica principalment les diferències en els anells A i B dels BRs, però comença a diferenciar les conformacions actives *IZR* de les *HIP*.

El tercer component principal acaba d'explicar les diferències entre les conformacions actives *IZR* i les *HIP*. Aquesta es deguda a: 1) la diferent orientació de l'hidroxil en C23, i 2) el volum de l'extrem final de la cadena lateral. Les conformacions actives *MVP* queden enmig ja que l'orientació de l'hidroxil en C23 coincideix amb la *HIP*, però el volum de l'extrem final de la cadena lateral coincideix millor amb la *IZR*.

#### b) Altres agrupacions.

El segon component principal explica majoritàriament les diferències en els anells A i B dels BRs segons una tendència que, a priori, no sembla lògica. Habitualment, s'acostuma a tractar per les funcionalitats de l'anell A (diol  $2\alpha,3\alpha$ , diol  $2\beta,3\beta...$ ) i l'anell B (lactona, cetona...) per separat. No obstant, el model les tracta conjuntament diferenciant entre  $2\alpha,3\alpha$ -lactones,  $2\alpha,3\alpha$ -cetones<sup>\*</sup>, i cetones amb funcionalitat diversa a l'anell A<sup>†</sup> (*Figura 130a*).

Estudiant el perquè d'aquesta observació, es detecta una interacció estructural que afecta als anells A i B. Tots els BRs amb funcionalitat lactona en l'anell B presenten el diol  $2\alpha,3\alpha$  en l'anell A. En canvi, els BRs amb funcionalitat cetona en l'anell B presenten o bé el diol  $2\alpha,3\alpha$  o bé altres funcionalitats en l'anell A. No hi ha cap BRs amb funcionalitat lactona en l'anell B i amb funcionalitat diversa a l'anell A.<sup>‡</sup> Aquesta interacció afecta considerablement els resultats del PCA ja que el segon component principal no pot explicar l'anell A sense l'ajut de l'anell B, ni viceversa. La màxima diferència estructural no es dóna entre el diol  $2\alpha,3\alpha$  o  $2\beta,3\beta$ , ni entre lactona o cetona sinó entre  $2\alpha,3\alpha$ -lactones i cetones amb funcionalitat diversa a l'anell A, quedant les  $2\alpha,3\alpha$ -cetones en una situació intermèdia.

El quart component principal explica els BRs amb unió A/B cis (Figura 130b).

Els coeficients del cinquè component principal no es poden interpretar d'acord amb cap dels BRs sintetitzats. Per fer-ho, cal separar l'anell A de la cadena lateral: En l'anell A explica les diferències entre els  $2\alpha$ , $3\alpha$ -BRs i els  $3\beta$ -BRs. En la cadena lateral explica les diferències entre els 22R,23R-BRs i els 22S,23S-BRs de les conformacions actives SQR i *IZR*.

El sisè component principal explica les diferències entre els 22*R*,23*R*-BRs i els 22*S*,23*S*-BRs de les conformacions actives *MVP* i *HIP* d'acord amb els volums de la cadena lateral.

<sup>&</sup>lt;sup>\*</sup> Aquest grup inclou els tres compostos amb una funcionalitat oxigenada en l'anell B enlloc del carbonil (hcs\_6OHa, hcs\_6OHb i hcs\_eter). Aquest fet es repetirà en tots les models. Per tant, si no es diu el contrari el grup de les  $2\alpha$ , $3\alpha$ -cetones inclourà els tres compostos anteriors. Igualment, si no es diu el contrari el grup de les  $2\alpha$ , $3\alpha$ -cetones inclourà l'anàleg amb un sol hidroxil en posició  $3\alpha$  (hcs\_2n3a). Els BRs amb unió A/B cis i diol  $2\beta$ , $3\beta$  també acostumen a entrar dins d'aquest grup, no obstant, en el model *CONJUNT* es comporten com cetones amb funcionalitat diversa a l'anell A.

<sup>&</sup>lt;sup>+</sup> S'entén per funcionalitat diversa en l'anell A, qualsevol que no presenti un grup funcional en la cara  $\alpha$  de l'anell A.

<sup>&</sup>lt;sup>\*</sup> L'anàleg BR hbl\_cis entraria dins d'aquesta categoria. No obstant, l'anàleg es comporta com si la funcionalitat a l'anell B fora una cetona enlloc d'una lactona. Aquest fet es repetirà en tots les models i per tant la hbl\_cis mai no es considerarà com una lactona.



*Figura 130*: *a)* Gràfic d'objectes dels 2 primers components principals del model *CONJUNT. b)* Gràfic d'objectes del tercer i quart component principal del model *CONJUNT.* 

#### 2.2. Anàlisi de mínims quadrats parcials (PLS).

#### a) Paràmetres estadístics.

S'han calculat 10 components principals. El model s'ha validat mitjançant mètodes de validació creuada. Es fan servir tres tècniques: LOO (Leave One Out), LTO (Leave Two Out) i classificació en grups definits. En l'últim cas es defineixen quatre grups segons la conformació activa. Així doncs, es construeixen tres models reduïts. En cadascun es deixa una família de conformers actius fora. Posteriorment, el compostos de cada família s'interpolen en el model reduït del qual n'han quedat fora i els contrasta l'activitat predita amb l'experimental.

A la *Taula* 9 es mostren els paràmetres estadístics del model. Es presenta: a) la variable latent, b) la variància tant l'explicada pel component (XVarExp) com l'acumulada (XAccum), c) la desviació estàndard de l'error de càlcul (SDEC), d) el coeficient de correlació ( $r^2$ ), e) la desviació estàndard de l'error de predicció (SDEP), i f) el coeficient de correlació de la predicció ( $q^2$ ) calculat amb cadascuna de les tres tècniques (LOO, LTO i classificació en grups definits).

Els mètodes de validació LOO i LTO donen valors molt optimistes de predictibilitat, els quals augmenten component rere component. Això es deu a que aquests mètodes no són capaços d'extreure tots els compostos d'una mateix classe o família i en conseqüència la informació de les classes és ben present en els models reduïts. Per contra el mètode de validació per grups ha permès eliminar les famílies de confórmers actius en els models reduïts la qual cosa ha provocat que l'estimació de la predictibilitat hagi estat més conservadora i més realista. Més conservadora perquè arriba a un màxim del 60% i més realista perquè la seva evolució coincideix amb la interpretació dels coeficients.

components	XVarExp	XAccum	SDEC	r <sup>2</sup>	SDEP	q <sup>2</sup> (LOO & LTO)****	SDEP	q <sup>2</sup> (Grups)
0	0,00	0,00	0,92	0,00	0,93	-0,03	0,92	0,00
1	7,76	7,76	0,56	0,62	0,66	0,48	0,62	0,54
2	5,52	13,27	0,44	0,77	0,57	0,61	0,58	0,60
3	6,51	19,78	0,33	0,87	0,54	0,65	0,61	0,56
4	12,10	31,88	0,27	0,91	0,49	0,72	0,56	0,63
5	5,16	37,04	0,22	0,94	0,46	0,75	0,56	0,63
6	4,47	41,51	0,17	0,96	0,41	0,80	0,57	0,61
7	5,11	46,62	0,14	0,98	0,39	0,82	0,57	0,62
8	4,24	50,85	0,12	0,98	0,37	0,84	0,58	0,60
9	3,16	54,01	0,10	0,99	0,35	0,85	0,57	0,61
10	5,19	59,20	0,09	0,99	0,34	0,86	0,58	0,60

Taula 9: Paràmetres estadístics del model CONJUNT.

#### b) Interpretació dels coeficients.

El primer component principal correlaciona l'estructura dels BRs amb la seva activitat. Amb un 8% de la variància estructural, corresponent a les zones d'interacció dels anells A i B, el model és capaç d'explicar un 62% de l'activitat amb un error de càlcul de  $\pm 0,56$ , i predir-ne un 54% amb un error de predicció de  $\pm 0,62$ .

El segon component principal corregeix a la baixa els 22*S*,23*S*-BRs independentment de la seva conformació activa. Aquest fet és sorprenent ja que els 22*S*,23*S*-BRs presenten una gran diversitat estructural segons la conformació activa i, a priori, no sembla que hagin d'haver gaires similituds. Els mapes de coeficients mostren un volum que és comú a tots els 22*S*,23*S*-BRs. No obstant, mentre que a les conformacions actives *SQR* i *IZR* correspon al volum de C22 i descriu el fet que la cadena lateral s'orienta cap a la cara  $\alpha$ , a les conformacions actives *MVP* i *HIP* correspon al volum de l'extrem final de la cadena lateral i dels anells B i D. Com els mapes de coeficients no es corresponen amb un únic patró que caracteritzi l'estructura dels 22*S*,23*S*-BRs independentment de la seva conformació activa, es suposa que es tracta d'una correlació a l'atzar.

El tercer component principal corregeix l'activitat segons la conformació activa. Corregeix a la baixa la conformació *SQR* i a l'alça les conformacions *MVP* i *HIP*. Els coeficients corresponen a les zones d'interacció de l'hidroxil en C23 i del volum de la cadena lateral. Tot i que el segon i tercer components principals comporten una millora del 25% en la capacitat de càlcul, no comporten una millora significativa en el predictibilitat.

# c) <u>Contribucions a l'activitat.</u>

En el model *CONJUNT* només té sentit considerar el primer component principal ja que és independent de la conformació activa i no se li atribueix cap correlació a l'atzar. Aquest explica un 62% de l'activitat amb un error de càlcul de  $\pm 0,56$ , i prediu un 54% amb un error de predicció de  $\pm 0,62$ , fent servir només un 8% de la variabilitat estructural.

En la *Figura 131* es mostren els coeficients del primer component principal a  $\pm 0,001$ . De la interpretació dels mapes de coeficients es conclou:

- Contribucions positives a l'activitat: les  $2\alpha$ , $3\alpha$ -lactones, i l'absència de grups funcionals en  $3\beta$ .
- Contribucions negatives a l'activitat: les cetones sense grups funcionals en  $2\alpha$  i  $3\alpha$ , l'hidroxil  $3\beta$ , i l'èter en C6.

<sup>\*\*\*\*</sup> En ambdós casos s'obté el mateix valor.

• La cadena lateral no presenta cap contribució a l'activitat.

Així doncs, independentment de la cadena lateral tipus i/o la seva conformació, el 54% de la predictibilitat de l'activitat depèn exclusivament de les funcionalitats dels anells A i B.



Sonda d'aigua Sonda de metil Figura 131: Coeficients del model CONJUNT. En vermell es mostren les contribucions positives a l'activitat i en blau les contribucions negatives.

# 2.3. Conclusions del model CONJUNT.

Anàlisi de components principals:

- S'ha demostrat quantitativament que les cadenes laterals tipus de les quatre conformacions actives són diferents entre si. De totes quatre la més diferent és la conformació *SQR*.
- Pel que fa a la diversitat estructural dels BRs:
  - a) La presència d'interaccions estructurals fa que no sigui possible explicar la diversitat en l'anell B sense explicar la diversitat en l'anell A.
  - b) Quantitativament, la diversitat en els anells A i B és de l'ordre del de la diversitat de les conformacions actives. En el següent ordre de d'importància està la diversitat deguda als BRs amb unió A/B cis.
  - c) Finalment s'ha demostrat quantitativament que els 22*R*,23*R*-BRs són estructuralment diferents als 22*S*,23*S*-BRs. Aquesta diferència és més acusada en les conformacions actives *SQR* i *IZR* que en les *MVP* i *HIP*.

Mínims quadrats parcials (PLS):

- La influència de les cadenes laterals tipus de les quatre conformacions actives en l'activitat es dóna en el tercer component principal com un factor de correcció, després d'una correlació a l'atzar en el segon component principal i no aporta una millora significativa en la predictibilitat. Per tant, es pot afirmar que les quatre conformacions actives són prou diferents entre si com per no presentar trets comuns que es puguin correlacionar amb l'activitat.
- Només amb les variacions estructurals dels anells A i B ja s'obté un model acceptable, capaç de predir l'activitat en un 54% amb un error de predicció de  $\pm 0,62$ .

# 3. Model SQR.

# 3.1. Anàlisi de Components Principals (PCA).

#### a) Agrupació segons les cadenes laterals tipus.

Entre el primer, el quart i el cinquè components principals s'expliquen les diferències entre les cadenes laterals tipus. Enmig es troben també els BRs amb unió A/B cis<sup>++++</sup> (*Figura 132*). S'observa com les cinc cadenes laterals tipus són diferents entre si. La principal diferència dóna entre els 22*R*,23*R*-BRs (cadenes tipus *bl, homo* i *epi*) i els 22*S*,23*S*-BRs (cadenes tipus *S\_homo* i *S\_epi*). No obstant, mentre els 22*R*,23*R*-BRs són bastant similars entre si, els 22*S*,23*S*-BRs són molt diferents entre si.

El primer component principal (*Figura 133a*) explica principalment la diferència entre els 22*R*,23*R*-BRs (cadenes tipus *bl*, *homo*<sup>‡‡‡‡</sup> i *epi*) i els 22*S*,23*S*-BRs (cadenes tipus *S\_homo* i *S\_epi*). Aquesta diferència es deguda a: 1) les zones d'interacció del diol de la cadena lateral, i 2) les cadenes laterals dels 22*S*,23*S*-BRs s'orienten cap a la cara  $\alpha$  abans que ho facin les dels 22*R*,23*R*-BRs. També explica per als 22*R*,23*R*-BRs la diferencia entre les cadenes tipus *homo* i *epi*, quedant enmig les *bl*. Aquesta es deguda als volums del carboni C28 que en el cas de les cadenes tipus *epi* coincideix amb una part del volum dels 22*S*,23*S*-BRs.

El quart component principal explica la diferència entre la cadena lateral tipus *epi* i els 22*S*,23*S*-BRs (cadenes tipus *S\_homo* i *S\_epi*). Els BRs amb unió A/B cis també tenen influència en aquest component (*Figura 133b*).

Finalment, el cinquè component principal explica les diferències entre les cadenes laterals tipus *S\_homo* i *S\_epi* (*Figura 133c*).



Figura 132: Gràfic d'objectes del primer, quart i cinquè components principals del model SQR.

<sup>&</sup>lt;sup>++++</sup> Els BRs amb unió A/B cis es comporten de forma diferent al marge de que la seva cadena lateral tipus és *homo*.

<sup>&</sup>lt;sup>++++</sup> En el primer component principal els BRs amb unió A/B cis es comporten segons la seva cadena lateral tipus (*homo*).



*Figura 133*: Gràfic d'objectes del primer (*a*), quart (*b*) i cinquè (*c*) components principals del model *SQR*. La coloració dels compostos és fa segons la cadena lateral tipus.

#### b) Altres agrupacions.

El segon component principal explica les diferències en els anells A i B d'acord amb la seqüència  $2\alpha$ , $3\alpha$ -lactones,  $2\alpha$ , $3\alpha$ -cetones, i cetones amb funcionalitat diversa a l'anell A (*Figura 134a*).

El tercer component principal explica els BRs amb unió A/B cis (Figura 134b).



*Figura 134*: *a*) Gràfic d'objectes dels 2 primers components principals del model *SQR*. *b*) Gràfic d'objectes del segon i tercer component principal del model *SQR*.

Finalment, destacar que les diferències entre les cadenes laterals són d'un ordre de magnitud similar a les diferències en els anells A i B.

# 3.2. Anàlisi de mínims quadrats parcials (PLS).

#### a) Paràmetres estadístics.

En el primer model amb els 20 compostos, tres compostos (cs, epics i hcs\_eter) presenten un comportament anòmal i es suprimeixen. En el segon model, amb 17 compostos, només els dos primers components principals són interpretables des del punt de vista estructural. A la *Taula 10* es mostren els paràmetres estadístics del model.

components	XVarExp	XAccum	SDEC	r <sup>2</sup>	SDEP	q <sup>2</sup> (LOO)	
0	0,00	0,00	0,88	0,00	0,93	-0,13	
1	23,45	23,45	0,36	0,83	0,48	0,70	
2	20,12	43,57	0,30	0,88	0,45	0,73	
3	0,74	44,31	0,17	0,96	0,53	0,63	
4	10,42	54,73	0,14	0,97	0,54	0,63	
5	7,06	61,79	0,11	0,98	0,55	0,61	
6.214 variables							

Taula 10: Paràmetres estadístics del model SQR.

#### b) interpretació dels coeficients.

El primer component principal correlaciona l'estructura dels BRs amb la seva activitat (*Figura 135a*). Amb un 23% de la variància estructural corresponent a les zones d'interacció dels anells A i B, i en menor grau a la cadena lateral, el model és capaç d'explicar un 83% de l'activitat amb un error de càlcul de ±0,36, i predir-ne un 70% amb un error de predicció de ±0,48.

El segon component principal corregeix l'activitat segons la cadena lateral tipus. Coincideix amb el primer component principal del PCA i corregeix a la baixa les cadenes laterals tipus *epi*, *S\_homo* i *S\_epi*, i a l'alça les cadenes tipus *bl* i *homo* (*Figura 135b*).



*Figura 135*: *a*) Gràfic de correlació interna del primer component principal del model *SQR*. *b*) Gràfic de correlació interna del segon component principal del model *SQR*.

#### c) <u>Contribucions a l'activitat.</u>

En el model *SQR*, cal considerar fins al segon component principal, ja que es pot interpretar la relació existent entre l'estructura i l'activitat. El model final explica un 88% de l'activitat amb un error de càlcul de  $\pm 0,30$  i prediu un 73% amb un error de predicció de  $\pm 0,45$ , fent servir el 44% de la variabilitat estructural.

En la *Figura 136* es mostren els coeficients per al segon component principal a  $\pm 0,001$ . De la interpretació dels mapes de coeficients es conclou:

- Contribucions positives a l'activitat: les  $2\alpha$ , $3\alpha$ -lactones, l'absència de grups funcionals en  $3\beta$ ; i en menor grau el diol 22R,23R en les zones més allunyades de la zona intermèdia del diol i la presència d'un metil o etil amb configuració *S* en C24 (cadenes laterals tipus *bl* i *homo*).
- Contribucions negatives a l'activitat: les cetones sense grups funcionals en  $2\alpha$  i  $3\alpha$ , i les cadenes laterals que a partir de C22 es situen cap a la cara  $\alpha$  (cadenes laterals tipus *S\_homo* i *S\_epi*).



Sonda d'aiguaSonda de metilFigura 136: Coeficients del model SQR. En vermell es mostren les contribucions<br/>positives a l'activitat i en blau les contribucions negatives.Sonda de metil

#### d) <u>Predictibilitat.</u>

En la *Figura 137a* es mostra el gràfic d'activitat predita davant l'activitat experimental. En conjunt la predictibilitat del model és bona (73%, amb un error de  $\pm$ 0,45). No obstant, quan es representa la predictibilitat en funció de l'estructura s'observa que hi ha conjunts de BRs que es prediuen millor que altres. Per a la discussió cal diferenciar entre els anells A i B i la cadena lateral.

La predictibilitat segons les cadenes laterals tipus:

- Prediu bé les cadenes tipus *epi*, *S\_homo* i *S\_epi* (*Figura 137b*, en verd, blau i cian respectivament).
- No és capaç d'explicar la bl (*Figura 137b*, en vermell), subestima el seu valor i no la considera pas diferent de la hbl.
- Promedia l'activitat de les cadenes tipus homo (*Figura 137c*, en groc). A excepció de hbl i hcs\_2n3b (en gris), el valor d'activitat predita per la resta de homoBRs oscil·la al voltant de 2,00  $\pm$ 0,17.

L'últim punt es degut a que en els homoBRs hi ha un elevat nombre de compostos amb modificacions estructurals en els anells A i B que no tenen cap influència en les contribucions a l'activitat.



*Figura 137*: Gràfics d'activitat predita davant d'activitat experimental: (a) sense coloració, (b) coloració segons les cadenes laterals tipus, i (c) detall de la cadena tipus homo.

La predictibilitat segons els anells A i B (Figura 138):

- Prediu bé les cetones amb funcionalitat diversa a l'anell A (en verd).
- Acceptable predicció de les 2α,3α-lactones (en vermell). L'excepció és la bl, però el seu comportament es deu més a la cadena lateral que no pas als anells A i B.
- Comportament divers de les 2α,3α-cetones (en groc). Com a conjunt no es pot treure una conclusió. El seu comportament respon més aviat a la cadena lateral. Bona predicció de les 2α,3α-cetones dels 22*S*,23*S*-BRs, però promedia les cadenes tipus *homo*.
- El model no és capaç de diferenciar entre els dos BRs amb unió A/B cis (en taronja) i els prediu el mateix valor d'activitat.



*Figura 138*: Gràfics d'activitat predita davant d'activitat experimental: coloració segons els anells A i B.

# 4. Model IZR.

# 4.1. Anàlisi de Components Principals (PCA).

#### a) Agrupació segons les cadenes laterals tipus.

Entre el primer, el quart i el cinquè components principals s'expliquen les diferències entre les cadenes laterals tipus, enmig es troben també els BRs amb unió A/B cis (*Figura 139*). S'observa que, a excepció de les cadenes tipus *bl* i *homo*, la resta de les cadenes laterals són diferents entre si. Igual que en el model *SQR*, la principal diferència es dóna entre els 22*R*,23*R*-BRs (cadenes tipus *bl*, *homo* i *epi*) i els 22*S*,23*S*-BRs (cadenes tipus *S\_homo* i *S\_epi*), i mentre els 22*R*,23*R*-BRs són bastant similars entre si els 22*S*,23*S*-BRs són molt diferents entre si.



Figura 139: Gràfic d'objectes del primer, quart i cinquè components principals del model IZR.

El primer component principal explica la diferència entre la cadena tipus *S\_homo* i la resta. Aquesta diferència es deu a: 1) la zona d'interacció de l'hidroxil en C23 de la cadena *S\_homo*, i 2) la cadena lateral dels *S\_homo* s'orienta cap a la cara  $\alpha$  quan en la resta s'estén lluny de l'esquelet esteroide. En aquest component principal, s'observa una altre interacció estructural que dificulta la seva interpretació: Tots els 22*S*,23*S*-BRs (cadenes laterals tipus *S\_homo* i *S\_epi*) presenten el diol 2 $\alpha$ ,3 $\alpha$ .

El quart component principal és complex ja que explica diverses coses alhora. En l'anell A explica els BRs amb unió A/B cis. En la cadena lateral explica coses diferents pels 22*R*,23*R*-BRs i pels 22*S*,23*S*-BRs. Pels primers explica la diferència entre les cadenes tipus *homo* i *epi* d'acord amb el volum de la cadena lateral. Pels segons explica la diferència entre les cadenes tipus *S\_homo* i *S\_epi* d'acord amb les zones d'interacció dels hidroxils de cadena.

Finalment, el cinquè component principal explica la cadena lateral tipus  $S_{epi}$  d'acord amb la zona d'interacció del diol de cadena i el volum de la cadena.

#### b) Altres agrupacions.

El segon component principal explica les diferències en els anells A i B d'acord amb la seqüència  $2\alpha$ , $3\alpha$ -lactones,  $2\alpha$ , $3\alpha$ -cetones, i cetones amb funcionalitat diversa a l'anell A (*Figura 140a*). No obstant, la diferència entre les  $2\alpha$ , $3\alpha$ -cetones i les cetones amb funcionalitat diversa a l'anell A és molt petita. Això permet que el tercer component principal pugui explicar les diferències en l'anell A sense tenir en compte l'anell B. Així doncs, explica la diferència entre els BRs que tenen un o dos hidroxils en  $\alpha$  en l'anell A i els que no els tenen o els tenen en  $\beta$  (*Figura 140b*).



*Figura 140*: *a)* Gràfic d'objectes dels 2 primers components principals del model *IZR*. *b)* Gràfic d'objectes del segon i tercer component principal del model *IZR*.

# 4.2. Anàlisi de mínims quadrats parcials (PLS).

#### a) Paràmetres estadístics.

En el primer model amb els 20 compostos, un compost (hcs\_eter) presenta un comportament anòmal i es suprimeix. En el segon model, amb 19 compostos, només els tres primers components principals són interpretables des del punt de vista estructural. A la *Taula 11* es mostren els paràmetres estadístics del model.

components	XVarExp	XAccum	SDEC	r <sup>2</sup>	SDEP	q <sup>2</sup> (LOO)	
0	0,00	0,00	0,90	0,00	0,95	-0,11	
1	11,97	11,97	0,42	0,78	0,61	0,54	
2	20,36	32,34	0,33	0,86	0,50	0,69	
3	21,06	53,39	0,28	0,90	0,51	0,68	
4	1,55	54,95	0,17	0,96	0,52	0,66	
5	7,53	62,48	0,13	0,98	0,53	0,65	
6.897 variables							

Taula 11: Paràmetres estadístics del model IZR.

#### b) interpretació dels coeficients.

El primer component principal correlaciona l'estructura dels BRs amb la seva activitat (*Figura 141a*). Amb un 12% de la variància estructural corresponent principalment a les zones d'interacció de l'hidroxil en  $3\alpha$  i al volum de l'extrem final de la cadena lateral, el model és capaç d'explicar un 78% de l'activitat amb un error de càlcul de  $\pm 0,42$ , i predir-ne un 54% amb un error de predicció de  $\pm 0,61$ .

El segon i el tercer components principals corregeixen l'activitat segons la cadena lateral tipus (*Figura 141b*). El segon component principal del PLS, coincideix bastant bé amb el primer component principal del PCA i el tercer component principal del PLS comparteix característiques del quart i del cinquè components principals del PCA.



*Figura 141*: *a*) Gràfic de correlació interna del primer component principal del model *SQR*. *b*) Gràfic d'objectes del segon i tercer components principals del model *IZR*.

c) <u>Contribucions a l'activitat.</u>

En el model *IZR*, cal considerar fins al tercer component principal, ja que es pot explicar la relació existent entre l'estructura i l'activitat. El model explica un 90% de l'activitat amb un error de càlcul de ±0,28, i prediu un 68% amb un error de predicció de ±0,51, fent servir el 53% de la variabilitat estructural.

En la *Figura 142* es mostren els coeficients per al segon component principal a  $\pm 0,001$ . De la interpretació dels mapes de coeficients es conclou:

- Contribucions positives a l'activitat: l'hidroxil en  $3\alpha$ , la presència d'un metil amb configuració *S* en C24 (cadena lateral tipus *bl*); i en menor grau l'absència de grups funcionals en  $3\beta$ , les lactones i el diol 22R, 23R.
- Contribucions negatives a l'activitat: l'absència de grup funcional en  $3\alpha$ , les cetones, la cadena lateral quan es situa cap a la cara  $\alpha$  (cadena lateral tipus *S\_homo*), i les cadenes que tenen l'isopropil terminal diferent al de la bl (cadenes laterals tipus *epi* i *S\_epi*).



*Figura 142*: Coeficients del model *IZR*. En vermell es mostren les contribucions positives a l'activitat i en blau les contribucions negatives.

# d) <u>Predictibilitat.</u>

En la *Figura 143a* es mostra el gràfic d'activitat predita davant l'activitat experimental. En conjunt la predictibilitat del model és bona (68%, amb un error de  $\pm$ 0,51). No obstant, quan es representa la predictibilitat en funció de l'estructura s'observa que hi ha conjunts de BRs que es prediuen millor que altres. Per a la discussió es torna a diferenciar entre els anells A i B i la cadena lateral.

La predictibilitat segon les cadenes laterals tipus:

- No és capaç de predir la cadena lateral tipus epi (Figura 143b, en verd).
- Prediu molt bé les cadenes tipus *S\_homo* i *S\_epi*, i acceptablement la tipus *bl* (*Figura* 143b, en blau, cian i vermell respectivament).
- Promedia l'activitat de les cadenes tipus homo (*Figura 143c*, en groc). A excepció de hbl i hcs\_2n3b (en gris), el valor d'activitat predit per la resta de homoBRs oscil·la al voltant de 2,10  $\pm$ 0,21.

La predictibilitat segons els anells A i B (Figura 144):

- Prediu molt bé les 2α,3α-lactones (en vermell).
- No és capaç de predir les cetones amb funcionalitat diversa a l'anell A (en verd). L'activitat predita mínima es situa al voltant de 1,0, sobreestimant el valor d'activitat dels compostos menys actius.
- Comportament divers de les  $2\alpha$ ,  $3\alpha$ -cetones (en groc). El salt d'activitat experimental entre lactona i cetona, no és uniforme i és més acusat en els 22S, 23S-BRs que en els 22R, 23R-BRs. En canvi, el salt d'activitat predita és bastant uniforme i més proper al valor dels 22S, 23S-BRs. Per aquest motiu, subestima les  $2\alpha$ ,  $3\alpha$ -cetones dels 22R, 23R-BRs, mentre que prediu correctament les dels 22S, 23S-BRs.
- Es prediu bé la  $3\alpha$ -cetona (en rosa), però es sobreestima l'activitat dels  $2\alpha$ , $3\alpha$ -BRs que no tenen una funcionalitat carbonílica en l'anell B (en morat).
- El model no és capaç de diferenciar entre els dos BRs amb unió A/B cis (en taronja).



*Figura 143*: Gràfics d'activitat predita davant d'activitat experimental: *(a)* sense coloració, *(b)* coloració segons les cadenes laterals tipus, i *(c)* detall de la cadena tipus *homo*.



*Figura 144*: Gràfics d'activitat predita davant d'activitat experimental: coloració segons els anells A i B.

# 5. Model MVP.

#### 5.1. Anàlisi de Components Principals (PCA).

Aquest model és el més complex ja que és molt sensible a les interaccions estructurals del conjunt de BRs i són necessaris diversos components principals per explicar una característica estructural. Això dificulta enormement la interpretació dels coeficients i complica el model.

#### a) Agrupació segons les cadenes laterals tipus.

Entre el primer, el segon i el quart components principals s'expliquen les diferències entre les cadenes laterals tipus (*Figura 145*). S'observa com les cinc cadenes laterals tipus s'apleguen en tres grups: a) cadenes tipus *bl* i *homo*, b) cadena tipus *epi*, i c) cadenes dels

22*S*,23*S*-BRs (cadenes tipus *S\_homo* i *S\_epi*). Entre el cinquè i sisè components principals s'explicaran les diferencies entre les cadenes tipus *S\_homo* i *S\_epi*.



*Figura 145*: Gràfic d'objectes del primer, segon i quart components principals del model *MVP*: Coloració segons la cadena lateral tipus.

Entre el primer i el segon components principals s'explica la diferència entre les cadenes laterals *bl* i *homo* per una banda i *epi*, *S\_homo* i *S\_epi* per l'altre (*Figura 146a*). En aquests dos components principals, s'observa una nova interacció estructural: El 50% dels BRs amb cadena lateral tipus *epi* representen el 66% de les cetones sense cap funcionalitat en  $\alpha$  en l'anell A.

Entre el primer i el quart components principals s'explica la diferència entre les cadenes laterals *epi* per una banda i *S\_homo* i *S\_epi* per l'altre (*Figura 146b*).



*Figura 146*: *a)* Gràfic d'objectes dels dos primers components principals del model *MVP*. *b)* Gràfic d'objectes del primer i quart components principals del model *MVP*.

Al contrari que passa amb els dos models anteriors, la diferència entre les diferents cadenes laterals tipus és del mateix ordre de magnitud. Això té molt de sentit ja que el canvi més gran en la conformació activa *MVP* es dóna en la geometria dels 22*S*,23*S*-BRs, la qual s'aproxima molt a la dels 22*R*,23*R*-BRs. Tan els hidroxils de cadena (22*R*,23*R* o

225,235) com l'extrem final de la cadena lateral coincideixen força bé en l'espai tridimensional. Això és possible degut a un petit gir de l'esquelet que provoca un creuament de les cadenes laterals en l'enllaç C22–C23 (veure *Figura 110*, pàgina 124). Els coeficients indiquen que la principal diferència estructural entre les cadenes dels 22*R*,23*R*-BRs i dels 22*S*,23*S*-BRs es dóna entre els anells C, D, i el metil C21 i no pas entre els hidroxils a C22,C23. Precísament, els anells C, D, i el metil C21 són les les zones més afectades pel gir de l'esquelet.

#### b) Altres agrupacions.

El primer, el segon i el quart components principals no només expliquen les principals diferències entre les cadenes laterals, sinó que també expliquen les diferències en els anells A i B d'acord amb la seqüència típica:  $2\alpha$ , $3\alpha$ -lactones,  $2\alpha$ , $3\alpha$ -cetones, i cetones amb funcionalitat diversa a l'anell A (*Figura 147*). Aquest fenomen és degut a les interaccions estructurals.



*Figura 147*: Gràfic d'objectes del primer, segon i quart components principals del model *MVP*: Coloració segons els anells A i B.

El primer component principal explica, majoritàriament, les diferències en els anells A i B d'acord amb la seqüència  $2\alpha$ , $3\alpha$ -lactones,  $2\alpha$ , $3\alpha$ -cetones, i cetones amb funcionalitat diversa a l'anell A (*Figura 148*).

El tercer component principal explica exclusivament els BRs amb unió A/B cis (Figura 149a).

A partir del quart component principal, es tornen a observar interaccions estructurals que compliquen la interpretació dels coeficients. Aquestes són: 1) totes les modificacions estructurals en l'anell B, diferents dels arquetips  $2\alpha$ , $3\alpha$ -lactona i  $2\alpha$ , $3\alpha$ -cetona, tenen la cadena lateral tipus *homo*, 2) tots els BRs que no tenen un carbonil (lactona o cetona) en C6 tenen el diol  $2\alpha$ , $3\alpha$ , i 3) tots els 22*S*,23*S*-BRs presenten el diol  $2\alpha$ , $3\alpha$ .

Degudes a aquestes interaccions, el cinquè i sisè components principals expliquen coses diferents en funció de la cadena lateral. Per la cadena lateral tipus *homo*, explica les diferències en l'anell B (*Figura 149b*). Per les cadenes laterals tipus *S\_homo* i *S\_epi* explica, amb l'ajuda del quart component principal, les diferències entre elles i amb la cadena tipus *epi* (*Figura 150*).



Figura 148: Gràfic d'objectes dels 2 primers components principals del model MVP.



*Figura 149*: Model *MVP*: *a*) Gràfic d'objectes del primer i tercer components principals. *b*) Gràfic d'objectes del quart i cinquè components principals pels BRs amb cadena lateral tipus *homo*.



**Figura 150**: Model *MVP*: Gràfic d'objectes del quart, cinquè i sisè components principals pels BRs amb cadena lateral tipus *epi*, *S\_homo* i *S\_epi* amb diol  $2\alpha$ , $3\alpha$  en l'anell A.

A partir del setè component principal, cada component principal explica les particularitats entre dos compostos. El setè component principal explica la diferència entre la S\_hcs i la S\_epics, el vuitè entre la epics\_2n3n i la epics\_2b3b, el novè entre la hbl\_cis i la hcs\_cis i el desè entre la hcs\_6OHa i la hcs\_6OHb.

Curiosament, tot i que el model *MVP* és el més complex d'interpretar, també és el que més informació estructural explica. S'arriben a interpretar 10 components principals, el que suposa el 73% de la variància. A més a més, és l'únic capaç d'explicar les modificacions estructurals de l'anell B.

# 5.2. Anàlisi de mínims quadrats parcials (PLS).

# a) Paràmetres estadístics.

Aquest és l'únic model on cap compost presenta un comportament anòmal. A la *Taula 12* es mostren els paràmetres estadístics del model.

components	XVarExp	XAccum	SDEC	r <sup>2</sup>	SDEP	q <sup>2</sup> (LOO)	
0	0,00	0,00	0,92	0,00	0,97	-0,11	
1	14,48	14,48	0,40	0,81	0,57	0,62	
2	6,70	21,17	0,24	0,93	0,49	0,71	
3	10,93	32,10	0,17	0,96	0,50	0,71	
4	7,73	39,83	0,12	0,98	0,50	0,71	
5	3,99	43,83	0,08	0,99	0,50	0,71	
5.271 variables							

Taula 12: Paràmetres estadístics del model MVP.

# b) Interpretació dels coeficients.

El primer component principal correlaciona l'estructura dels BRs amb la seva activitat (*Figura 151a*). Amb un 14% de la variància estructural corresponent principalment amb les zones d'interacció dels anells A i B, el volum de l'extrem final de la cadena lateral, i en menor grau el volum dels anells C, D i el metil C21, el model és capaç d'explicar un 81% de l'activitat amb un error de càlcul de  $\pm 0,40$ , i predir-ne un 62% amb un error de predicció de  $\pm 0,57$ .

El segon component principal corregeix a la baixa els 22*S*,23*S*-BRs i els BRs amb cadena tipus *homo* que presenten la unió A/B cis, o que no tenen un carbonil en l'anell B (*Figura* 151b).

El tercer component principal corregeix a la baixa els BRs amb cadena tipus homo (Figura 151c).

Com s'ha vist en el PCA, degut a les interaccions estructurals, són necessaris diversos components per explicar una característica estructural. El quart component principal del PLS fa aquesta funció. Així doncs, el segon, tercer i quart components principals del PLS comparteixen característiques amb el primer, segon i quart components del PCA. Per tant:

- Entre els segon i el quart es corregeix l'activitat dels 22*S*,23*S*-BRs en front a la dels 22*R*,23*R*-BRs (*Figura 152a*). També es corregeix l'activitat de les cadenes tipus *homo*, segons si tenen un tenen un carbonil a l'anell B, si no el tenen, o tenen la unió A/B cis (*Figura 152b*).
- Entre el tercer i el quart es corregeix l'activitat dels 22*R*,23*R*-BRs segons si la cadena és del tipus *homo* o *epi* (*Figura 152c*).

Entre els segon i el quart es components principals també es corregeix l'activitat de les cadenes tipus *homo*, segons si tenen un tenen un carbonil a l'anell B, si no el tenen, o tenen la unió A/B cis.



*Figura 151*: Model *MVP*: *a*) Gràfic de correlació interna del primer component principal. *b*) Gràfic de correlació interna del segon component principal. *c*) Gràfic de correlació interna del tercer component principal.



(a) i (b) • BRs que no participen en la discussió



*Figura 152*: Model *MVP*: *a*) Gràfic d'objectes del segon i quart components principals: coloració segons la cadena lateral. *b*) Gràfic d'objectes del segon i quart components principals: Detall dels BRs amb cadena tipus *homo* i coloració segons els anells A i B. *c*) Gràfic d'objectes del tercer i quart components principals.

#### c) <u>Contribucions a l'activitat.</u>

En el model *MVP*, té sentit considerar fins al quart component principal, ja que es pot explicar la relació entre l'estructura i l'activitat. Per tant, el model final *MVP* explica un 98% de l'activitat amb un error de càlcul de  $\pm 0,12$ , i prediu un 71% amb un error de predicció de  $\pm 0,50$ , fent servir el 40% de la variabilitat estructural.

En la *Figura 153* es mostren els coeficients per al segon component principal a  $\pm 0,001$ . La interpretació dels mapes de coeficients es conclou:

- Contribucions positives a l'activitat: les  $2\alpha$ , $3\alpha$ -lactones, l'absència de grups funcionals en  $3\beta$ , anell C i metil en C21 dels 22R,23R-BRs, la presència d'un metil o etil amb configuració S en C24 (cadenes laterals tipus *bl* i *homo*) i en menor grau, el diol de cadena.
- Contribucions negatives a l'activitat: les cetones sense grups funcionals en  $2\alpha$  i  $3\alpha$ , els anells B, D i l'extrem final de la cadena lateral dels 22*S*,23*S*-BRs.

Destacar que la contribució del diol de la cadena lateral és molt baixa. Aquest fet tot i ser contrari als requisits estructurals necessaris per desencadenar activitat descrits anteriorment,<sup>95,107</sup> no és sorprenent. Els hidroxils de cadena (22*R*,23*R* o 22*S*,23*S*) coincideixen força bé en l'espai tridimensional (veure *Figura 110*, pàgina 124). L'anàlisi de mínims quadrats parcials busca quina és la màxima dispersió de les dades capaç de correlacionar amb l'activitat i com la variabilitat estructural del diol és molt baixa, també ho ha de ser la seva contribució a l'activitat.



Sonda d'aiguaSonda de metilFigura 153: Coeficients del model MVP. En vermell es mostren les contribucions<br/>positives a l'activitat i en blau les contribucions negatives.

#### e) Predictibilitat.

En la *Figura 154a* es mostra el gràfic d'activitat predita davant l'activitat experimental. En conjunt la predictibilitat del model és bona (71%, amb un error de  $\pm$ 0,50). No obstant, quan es representa la predictibilitat en funció de l'estructura s'observa que hi ha conjunts de BRs que es prediuen millor que altres. Per a la discussió es torna a diferenciar entre els anells A i B i la cadena lateral.

La predictibilitat segons les cadenes laterals tipus:

- No és capaç de predir la cadena lateral tipus *S\_epi* (*Figura 154b*, en cian) ja que tan a la *S\_epibl* com a la *S\_epics* els prediu el mateix valor d'activitat.
- No prediu gaire bé la cadena tipus *epi* (*Figura 154b*, en verd), tot i que els seu comportament es deu més als anells A i B que no pas a la cadena lateral.

- Prediu bé la cadena tipus *S\_homo* i acceptablement la cadena tipus *bl* (*Figura 154b*, en blau).
- És l'únic model que no promedia l'activitat de les cadenes tipus *homo* (*Figura 154c*, en groc). Malgrat aquest fet, la dispersió en les activitats predites és força gran.



*Figura 154*: Gràfics d'activitat predita davant d'activitat experimental: (a) sense coloració, (b) coloració segons les cadenes laterals tipus, i (c) detall de la cadena tipus homo.

La predictibilitat segons els anells A i B (Figura 155):

- Prediu molt bé les  $2\alpha$ ,  $3\alpha$ -lactones (en vermell).
- No és capaç de predir les cetones amb funcionalitat diversa a l'anell A (en verd). L'activitat predita mínima es situa al voltant de 1,0, sobreestimant el valor d'activitat dels compostos menys actius.
- El model no és capaç de diferenciar entre els dos BRs amb unió A/B cis (en taronja) ja que tan a la hbl\_cis com a la hcs\_cis els prediu el mateix valor d'activitat.
- Es prediu molt bé la 3α-cetona (en rosa), però les 2α,3α-cetones (en groc) presenten un comportament divers.
- Acceptable predicció de l'activitat dels 2α,3α-BRs que no tenen una funcionalitat carbonílica en l'anell B (en morat). Dels tres BRs el que pitjor es prediu és l'èter, però és degut a que el model sobreestima els compostos menys actius.



*Figura 155*: Gràfics d'activitat predita davant d'activitat experimental: coloració segons els anells A i B.

# 6. Model HIP.

#### 6.1. Anàlisi de Components Principals (PCA).

El primer component principal explica les diferències entre les cadenes laterals (*Figura* 156). Aquesta diferència és deguda principalment al volum de les cadenes laterals. El volum dels anells C, D i metil en C21 permet separar els 22*R*,23*R*-BRs dels 22*S*,23*S*-BRs. El volum del metil o etil en C24 permet en el cas dels 22*R*,23*R*-BRs separar les cadenes tipus *bl* i *homo* de les *epi*, i en el cas dels 22*S*,23*S*-BRs separar entre les cadenes *S*\_*homo* i *S*\_*epi*. S'observa una relació gairebé lineal entre les cadenes laterals, sent les cadenes tipus *bl* i *homo* bastant similars entre si.



*Figura 156*: Model *HIP*: Gràfic d'objectes dels 2 primers components principals. Coloració segons la cadena lateral.

El primer component principal presenta coeficients al voltant dels grups funcionals polars. No obstant, aquests no contribueixen a explicar les diferències "reals" entre les cadenes, sinó que són fruit de la interacció estructural deguda al fet que tots els 22*S*,23*S*-BRs tenen el diol  $2\alpha$ , $3\alpha$ . En la regió del diol de cadena presenta hi ha tan coeficients positius com negatius, els quals no són atribuïbles a cap cadena lateral sinó que sorprenentment són conseqüència de la funcionalitat present en l'anell B. Curiosament, la contribució de l'anell B és molt petita en comparació amb la resta de models. En la *Figura 157* es mostren les contribucions de la brassinolida ( $2\alpha$ , $3\alpha$ -lactona) i de la castasterona ( $2\alpha$ , $3\alpha$ -cetona) al primer component principal. S'observa que el diol de cadena de la brassinolida presenta majoritàriament coeficients negatius (en blau) mentre que el de la castasterona presenta coeficients positius (en vermell). També s'observa com la influència del carbonil de l'anell B és petita i coincideix amb el signe (i el color) dels coeficients del diol de cadena. Per tant es pot concloure que l'anell B queda parcialment descrit través del diol de la cadena lateral.



*Figura 157*: Contribucions al primer component principal de la brassinolida i la castasterona.

El segon component principal explica les diferències en els anells A i B d'acord amb la seqüència  $2\alpha$ , $3\alpha$ -lactones,  $2\alpha$ , $3\alpha$ -cetones,  $3\alpha$ -cetona i cetones amb funcionalitat diversa a l'anell A (*Figura 158a*).

El tercer component principal explica els BRs amb unió A/B cis (Figura 158b).



*Figura 158*: Model *HIP*: (a) Gràfic d'objectes dels 2 primers components principals. Coloració segons els anells A i B. (b) Gràfic d'objectes del segon i tercer components principals.

Destacar que entre el primer i el segon components principals les grans característiques estructurals dels BRs queden perfectament definides i permeten encasellar els compostos segons la cadena lateral i els anells A i B. A més a més, queden una sèrie de forats on es poden situar compostos no inclosos en el model (veure pàgina 188).

Sorpren que el model HIP no pugui explicar les les modificacions estructurals en l'anell B, ja que la principal diferència amb el MVP és la cadena lateral tipus *epi* i no pas els anells A i B.

A efectes pràctics, l'explicació està en el fet que el carbonil de l'anell B s'explica parcialment a través del diol de la cadena lateral. Per tant, no ha d'estranyar que el model no pugui explicar els BRs que no tenen aquest carbonil, ja que els explica majoritàriament a través dels coeficients de la cadena lateral i per això els tracta com si fossin  $2\alpha$ , $3\alpha$ -cetones.

#### 6.2. Anàlisi de mínims quadrats parcials (PLS).

a) Paràmetres estadístics.

En el primer model amb els 20 compostos, un compost (hcs\_eter) presenta un comportament anòmal i es suprimeix. En el segon model, amb 19 compostos, només els dos primers components principals són interpretables des del punt de vista estructural. A la *Taula 13* es mostren els paràmetres estadístics del model.

components	XVarExp	XAccum	SDEC	$r^2$	SDEP	q <sup>2</sup> (LOO)	
0	0,00	0,00	0,90	0,00	0,95	-0,11	
1	18,83	18,83	0,34	0,86	0,47	0,73	
2	4,95	23,78	0,24	0,93	0,52	0,66	
3	8,55	32,33	0,17	0,96	0,55	0,62	
4	9,59	41,92	0,13	0,98	0,55	0,62	
5	2,40	44,31	0,09	0,99	0,57	0,60	
6.225 variables							

Taula 13: Paràmetres estadístics del model HIP.

# b) Interpretació dels coeficients.

El primer component principal correlaciona l'estructura dels BRs amb la seva activitat (*Figura 159a*). Els coeficients coincideixen amb els del primer i segon components principals del PCA. Així doncs, amb un 19% de la variància estructural corresponent a les zones d'interacció dels anells A i B, i de la cadena lateral, el model és capaç d'explicar un 86% de l'activitat amb un error de càlcul de ±0,34, i predir-ne un 73% amb un error de predicció de ±0,47.

El segon component principal corregeix a la baixa l'activitat dels BRs amb modificacions estructurals en C5 (unió A/B cis i l'hidroxil en 5 $\alpha$ ) i els BRs sense funcionalitat carbonílica en C6 (2 $\alpha$ ,3 $\alpha$  no cetona) (*Figura 159b*).



*Figura 159*: Model *HIP*: *a*) Gràfic de correlació interna del primer component principal i *b*) Gràfic de correlació interna del segon component principal.

#### c) <u>Contribucions a l'activitat.</u>

En el model *HIP*, cal considerar fins al segon component principal, ja que es pot interpretar la relació existent entre l'estructura i l'activitat. El model final explica un 93% de l'activitat amb un error de càlcul de  $\pm 0,24$ , i prediu un 66% amb un error de predicció de  $\pm 0,52$ , fent servir el 24% de la variabilitat estructural.

En la *Figura 160* es mostren els coeficients per al segon component principal a  $\pm 0,001$ . De la interpretació dels mapes de coeficients es conclou:

- Contribucions positives a l'activitat: les  $2\alpha$ ,  $3\alpha$ -lactones<sup>§§§§</sup> amb una major contribució de l'hidroxil  $3\alpha$  en front del  $2\alpha$ , absència de grups funcionals en  $3\beta$ , anell C i metil en C21 dels 22R, 23R-BRs, i la presència d'un metil o etil amb configuració S en C24 (cadenes laterals tipus *bl* i *homo*).
- Contribucions negatives a l'activitat: les cetones sense grups funcionals en  $2\alpha$  i  $3\alpha$ , els anells B, D i l'extrem final de la cadena lateral dels 22*S*,23*S*-BRs, i la presència d'un metil amb configuració S en C24 (cadena lateral tipus *epi*).



*Figura 160*: Coeficients del model *HIP*. En vermell es mostren les contribucions positives a l'activitat i en blau les contribucions negatives.

d) <u>Predictibilitat.</u>

En la *Figura 161a* es mostra el gràfic d'activitat predita davant l'activitat experimental. En conjunt la predictibilitat del model és bona (66%, amb un error de  $\pm$ 0,52). No obstant, quan es representa la predictibilitat en funció de l'estructura s'observa que hi ha conjunts de BRs que es prediuen millor que altres. Per a la discussió es torna a diferenciar entre els anells A i B i la cadena lateral.

La predictibilitat segon les cadenes laterals tipus:

- Prediu molt bé les cadenes tipus *S\_homo* i *S\_epi* (*Figura 161b*, en blau i cian).
- Malgrat que l'activitat està lleugerament subestimada, prediu força bé la cadena tipus *bl (Figura 161b,* en vermell).
- Prediu de forma acceptable la cadena tipus epi (Figura 161b, en verd).
- Promedia l'activitat de les cadenes tipus homo (*Figura 161c*, en groc). A excepció de hbl i hcs\_2n3b, el valor d'activitat predit per la resta de homoBRs oscil·la al voltant de 1,99  $\pm$ 0,37.

<sup>&</sup>lt;sup>\$\$\$\$</sup> Recordar que l'anell B s'explica a través del diol de la cadena lateral.



*Figura 161*: Gràfics d'activitat predita davant d'activitat experimental: (*a*) sense coloració, (*b*) coloració segons les cadenes laterals tipus, i (*c*) detall de la cadena tipus *homo*.

La predictibilitat segons els anells A i B (Figura 162):

- Prediu molt bé les  $2\alpha$ ,  $3\alpha$ -lactones (en vermell).
- Prediu bé les cetones amb funcionalitat diversa a l'anell A (en verd).
- Comportament divers de les  $2\alpha$ , $3\alpha$ -cetones (en groc). El salt d'activitat experimental entre lactona i cetona, no és uniforme i és més acusat en els 22S,23S-BRs que en els 22R,23R-BRs. En canvi el salt d'activitat predita és bastant uniforme i més proper al valor dels 22S,23S-BRs. Per aquest motiu, subestima les  $2\alpha$ , $3\alpha$ -cetones dels 22R,23R-BRs, mentre que prediu correctament les dels 22S,23S-BRs.
- Es prediu bé la  $3\alpha$ -cetona (en rosa), però es sobreestima l'activitat dels  $2\alpha$ ,  $3\alpha$ -BRs que no tenen una funcionalitat carbonílica en l'anell B (en morat).
- El model no és capaç de diferenciar entre els dos BRs amb unió A/B cis (en taronja).



*Figura 162*: Gràfics d'activitat predita davant d'activitat experimental: coloració segons els anells A i B.

# 7. Conclusions.

#### 7.1. Quantificació de la diversitat estructural.

El model *CONJUNT* confirma que les quatre conformacions actives són diferents entre si. De totes quatre la més diferent és la conformació *SQR* ja que la cadena lateral es posiciona cap a la cara  $\alpha$ . Les conformacions *IZR*, *MVP* i *HIP* són diferents degut a l'orientació de l'hidroxil en C23 i al volum de l'extrem final de la cadena lateral.

Tant el model *CONJUNT* com els models *SQR* i *IZR* confirmen que per les conformacions actives *SQR* i *IZR* els 22*S*,23*S*-BRs són molt diferents dels 22*R*,23*R*-BRs, i inclús són molt diferents entre si. En les conformacions *MVP* i *HIP* les cinc cadenes laterals tipus són diferents entre si, però dins d'un mateix ordre de magnitud.

Es posa de manifest un gran nombre d'interaccions estructurals que afecten als models en major o menor grau. Aquestes són:

- 1. Totes les lactones presenten el diol  $2\alpha$ ,  $3\alpha$ .
- 2. Tots els 22*S*,23*S*-BRs presenten el diol  $2\alpha$ ,  $3\alpha$ .
- 3. El 50% dels BRs amb cadena lateral tipus *epi* representen el 66% de les cetones sense cap funcionalitat en  $\alpha$  en l'anell A.
- 4. Tots els BRs amb modificacions estructurals en l'anell B presenten la cadena lateral tipus *homo*.
- 5. Tots els BRs sense carbonil en l'anell B presenten el diol  $2\alpha_3\alpha$ .

De totes, la que més influència té és la primera. S'observa en tots els models i obliga a veure els anells A i B de forma conjunta d'acord amb la seqüència  $2\alpha$ , $3\alpha$ -lactona,  $2\alpha$ , $3\alpha$ -cetona i cetones sense funcionalitat en la cara  $\alpha$  de l'anell A.

Fins ara sempre s'havia considerat que els BRs tenien tres zones d'influència: 1) l'anell A, 2) l'anell B i 3) la cadena lateral. No obstant, quantitativament, només hi ha dos zones d'influència: 1) els anells A i B conjuntament i 2) la cadena lateral.

Tots els models coincideixen en senyalar que les característiques estructurals que diferencien els BRs entre si són:

- Els anells A i B d'acord amb la seqüència  $2\alpha$ , $3\alpha$ -lactona,  $2\alpha$ , $3\alpha$ -cetona i cetones sense funcionalitat en la cara  $\alpha$  de l'anell A.
- La unió dels anells A i B (A/B cis o A/B trans).
- Les cadenes laterals tipus. No obstant les semblances i diferències entre elles s'expliquen de forma diferent per cada conformació activa.

Algunes característiques estructurals no queden explicades per igual en tots els models. Aquests són els casos de:

- a) Les diferències estructurals en l'anell A. Només el model *CONJUNT* i el model *IZR* diferencien si els hidroxils de l'anell A estan en la cara  $\alpha$  o en la cara  $\beta$ . En la resta de models la diversitat estructural de l'anell A s'explica d'acord amb la interacció amb l'anell B.
- b) Els BRs sense funcionalitat cetona a l'anell B només són explicats pel model MVP.
- c) La diferència entre lactona o cetona a l'anell B per als BRs amb unió A/B cis només és explicada en el model *IZR*.
- d) El diol de la cadena lateral perd pes en les conformacions *MVP* i *HIP*. És més, en la conformació activa *HIP*, l'anell B queda descrit parcialment a través del diol de la cadena lateral.

En absència d'explicació d'algunes funcionalitats, els models fan les següents aproximacions estructurals:

- Els BRs sense funcionalitat carbonílica a l'anell B es comporten com  $2\alpha$ , $3\alpha$ -cetones.
- Abans de formar un grup independent, els BRs amb unió A/B cis es comporten com  $2\alpha$ , $3\alpha$ -cetona encara que l'anell B tingui funcionalitat lactona.
- Els BRs amb un o dos hidroxils en  $\beta$  en l'anell A, es comporten com els que no tenen cap funcionalitat.
- La hcs\_2n3a es comporta de forma diferent segons el model. En el model *CONJUNT* és comporta com una cetona sense el diol  $2\alpha$ , $3\alpha$ , en els models *SQR*, *IZR* i *MVP* com una  $2\alpha$ , $3\alpha$ -cetona i només en el model *HIP* com una cetona a mig camí entre l'absència de funcionalitat en  $\alpha$  i la doble funcionalitat.

#### 7.2. Relacions estructura-activitat quantitatives segons la conformació activa.

El model *CONJUNT* indica que només amb les variacions estructurals dels anells A i B es pot explicar un 62% de l'activitat i predir-ne un 54% amb un error de predicció de  $\pm 0,62$ . La cadena lateral no presenta cap contribució. En la resta de models (*SQR*, *IZR*, *MVP* i *HIP*) en els quals la cadena lateral sí influeix, s'arriba a explicar més d'un 85% de l'activitat i a predir-ne al voltant del 70% amb un error de predicció al voltant de  $\pm 0,50$ .

En general els models coincideixen en explicar les contribucions a l'activitat. No obstant, hi ha diferències en la forma d'explicar la influència de la cadena lateral segons sigui la conformació activa.

Al contrari del que es creia, aquestes no es deuen a tres arees d'influència (anell A, anell B i cadena lateral), sinó que només es deuen a dos: 1) els anells A i B conjuntament i 2) la cadena lateral.

Resumint, les contribucions a l'activitat són:

- Contribucions positives a l'activitat:
  - 1.  $2\alpha$ ,  $3\alpha$ -lactones. En el model *HIP*, però sobretot en el model *IZR* l'hidroxil  $3\alpha$  presenta una major contribució que el  $2\alpha$ .
  - 2. l'absència de grups funcionals en 3β.
  - 3. 22*R*,23*R*-BRs. Però mentre en els models *SQR* i *IZR* la contribució positiva és a través diol de cadena, en els models *MVP* i *HIP* ho és a través volum de l'anell C i del metil C21.
  - 4. cadenes laterals tipus *bl* i *homo* degut al volum de metil o etil amb configuració *S* en C24. No obstant, la relació d'aquest volum amb el seu entorn varia segons la conformació activa.
- Contribucions negatives a l'activitat:
  - 1. les cetones sense grups funcionals en  $2\alpha$  i  $3\alpha$ .
  - 2. 22*S*,23*S*-BRs. Però mentre en els models *SQR* i *IZR* la contribució negativa és perquè la cadena gira cap a la cara  $\alpha$ , en els models *MVP* i *HIP* ho és a través del volum dels anells B, D, i de l'extrem final de la cadena lateral.
  - 3. cadenes tipus *epi* i *S\_epi*, però només en els models *IZR* i *HIP*. Però mentre en el model *IZR* la contribució negativa és a través del volum de l'isopropil terminal, en el model *HIP* ho és a través del volum del metil amb configuració *R* en C24.

En conjunt, la predictibilitat dels models és bona (al voltant del 70% amb un error de predicció al voltant de  $\pm 0,50$ ). No obstant, varia considerablement en funció de l'estructura i dels models:

- Predictibilitat segons la cadena lateral:
  - a) el model *IZR* no ha estat capaç de predir correctament la cadena tipus *epi*.
  - b) el model *MVP* no ha estat capaç de predir correctament els 22*S*,23*S*-BRs, en concret falla en la cadena tipus *S\_epi*.
  - c) a excepció del model *MVP*, els models promedien l'activitat dels BRs amb cadena lateral tipus *homo*.
- Predictibilitat segons els anells A i B:
  - a) tots els models prediuen molt bé les  $2\alpha$ ,  $3\alpha$ -lactones. L'única excepció és la bl en el model *SQR*.
  - b) els models *IZR* i *MVP* sobreestimen els valors de les cetones amb funcionalitat diversa en l'anell A. En canvi, els models *SQR* i *HIP* són capaços de predir-les correctament.
  - c) el model *MVP* ha estat l'únic capaç de predir els BRs sense funcionalitat cetona a l'anell B.
  - d) cap model no ha estat capaç de predir correctament els BRs amb unió A/B cis.

# 7.3. Conclusions finals.

Conformacions actives estructuralment diferents han donat lloc a models quantitativament similars, però qualitativament diferents. Quantitativament similars perquè les parts dels BRs que correlacionen amb l'activitat són bàsicament les mateixes: Anells A i B conjuntament, i cadena lateral tipus. Qualitativament diferents perquè: 1) la manera amb que cada part i especialment la cadena lateral, contribueix a l'activitat és diferent, i 2) perquè la predictibilitat varia en funció de l'estructura.

Les conformacions actives *SQR* i *IZR* presenten el problema que no es pot justificar en termes de complementarietat, el reconeixement molecular dels 22*S*,23*S*-BRs per part del receptor. Per tant, els seus models QSAR es descarten.

El model *MVP* presenta l'avantatge que és l'únic que explica els BRs sense funcionalitat carbonílica en l'anell B, tots ells amb el diol  $2\alpha$ , $3\alpha$  i cadena tipus *homo*. Això fa que no promediï l'activitat dels BRs amb cadena lateral tipus *homo*. No obstant, aquesta millora és a costa dels 22*S*,23*S*-BRs i de sobreestimar les activitats inferiors a 1,0.

El model *HIP* té com a únic "inconvenient" que promedia l'activitat dels BRs amb cadena lateral tipus *homo*. No obstant, aquest fet es deu a que hi ha un bon nombre de compostos amb cadena lateral tipus *homo*, les modificacions estructurals dels quals en els anells A i B no tenen cap influència en les contribucions a l'activitat. A més, cal afegir que els valors d'activitat d'aquests compostos es troben en la zona intermèdia del marge d'activitats. Per tant, el promig en la predicció dels seus valors d'activitat no és tan crític com en un principi pot semblar.

Per tot plegat, es recomana fer servir la conformació activa i el model *HIP* per explicar i predir l'activitat dels BRs.

# <u>PART II: Models Reduïts.</u>

# 1. Introducció.

En la discussió dels models QSAR (PLS) s'ha posat molt d'èmfasi en l'avaluació de la qualitat dels models en termes de predictibilitat. Aquesta s'ha obtingut mitjançant mètodes de validació creuada, concretament el mètode L.O.O. (*Leave One Out*). Els paràmetres que regeixen la discussió són la desviació estàndard de l'error de predicció (SDEP) i el coeficient de correlació de la predicció (q<sup>2</sup>). Aquests mètodes són molt útils quan es disposa d'un reduït nombre de compostos. No obstant, tendeixen a donar estimacions optimistes de la predictibilitat. Una altre forma d'avaluar la predictibilitat és fent servir el mètode de validació externa. Aquest mètode és útil quan es té un elevat nombre de compostos i consisteix en dividir els compostos en dos grups: 1) *"data set"* i 2) *"test set"*. El *"data set"* es fa servir per realitzar el model i el *"test set"* per validar-lo. Aquest mètode avalua la predictibilitat de forma més realista. No obstant, hi ha diversos factors que s'han de tenir en compte per aplicar-lo correctament: a) Els compostos s'han de distribuir uniformement dins del marge d'activitat per ambdós grups i b) el resultat depèn de quants i quins compostos s'han assignat a cada grup.

En el tema anterior s'han comparat els models obtinguts amb cadascuna de les conformacions actives. Els models han treballat amb 20 BRs. A continuació es proposa fer servir el resultat de l'anàlisi de components principals per dividir els 20 BRs en dos grups de 10 BRs. Un es farà servir com a *"data set"* per fer un model QSAR reduït i l'altre es farà servir com a *"test set"* per validar-lo externament i avaluar-ne la predictibilitat. El *"data set"* el formen els 10 compostos més descriptius considerant els components principals més rellevants del PCA i els 10 compostos restants formen el *"test set"* (*Figura 163*).



Figura 163: Esquema de la realització dels models reduïts per cada conformació activa.

# 2. Compostos del "data set" i "test set".

La *Taula 14* mostra els 10 compostos que formen part del "*data set*" i els 10 del "*test set*" per cadascun dels quatre models reduïts. Com els resultats del QSAR depenen dels compostos del "*data set*" cal veure com es distribueix la informació estructural en cada model reduït en relació amb el model inicial de 20 compostos (*Taula 15*).

Destacar que el model inicial, ja d'entrada, no és un model equilibrat ja que les característiques estructurals no es troben en igual proporció. Els fets més rellevants són: 1) dels sis tipus de modificacions estructurals en l'anell A, el diol  $2\alpha$ , $3\alpha$  acapara el 70% dels compostos, 2) dels cinc tipus de modificacions estructurals en l'anell B, la funcionalitat cetona acapara el 55% dels compostos i la lactona el 30%, i 3) de les cinc cadenes laterals tipus, la cadena tipus *homo* acapara el 50% dels compostos.

Malgrat que el model inicial no està equilibrat, els models reduïts mantenen la informació estructural en una proporció similar a la del model inicial. Només destacar que

en el model reduït *HIP* no hi ha cap representant de la cadena tipus *bl* ja que es substitueix la brassinolida (és l'únic model reduït que no la selecciona) per la homobrassinolida (és l'únic model reduït que la selecciona).

DDo		log(dosi )			
DKS	SQR	IZR	MVP	HIP	-10g(uosi <sub>50%</sub> )
bl	data set	data set	data set		3,86
CS					3,33
epibl	data set			data set	2,67
epics		data set	data set		2,44
epics_2b3b	data set				0,37
epics_2n3n		data set	data set	data set	0,47
S_hbl					2,05
S_hcs	data set	data set		data set	1,27
S_epibl	data set	data set	data set	data set	2,24
S_epics		data set	data set		1,12
hbl				data set	3,07
hbl_cis	data set		data set	data set	2,45
hcs	data set	data set			2,78
hcs_cis		data set			1,30
hcs_2n3a				data set	2,02
hcs_2n3b		data set	data set	data set	1,52
hcs_60Ha	data set		data set	data set	1,60
hcs_60Hb			data set		1,70
hcs_50H	data set		data set	data set	1,71
hcs_eter	data set	data set			0,71

\* Els compostos que no estan marcats com a "data set" pertanyen al "test set".

Taula 14: Conjunt de BRs que per cada model reduït formen part del "data set" i del "test set".

Característica	Model		Model reduït						
estructural	inicial	SQR	IZR	MVP	HIP				
Anells A i B									
$2\alpha$ , $3\alpha$ -lactona	5	3	2	2	3				
$2\alpha$ , $3\alpha$ -cetona*	7	3	4	3	3				
$2\alpha$ , $3\alpha$ no cetona	3	2	1	2	1				
A/B cis	2	1	1	1	1				
cetona no $2\alpha$ , $3\alpha$	3	1	2	2	2				
Cadena lateral									
bl	2	1	1	1	0				
homo	10	5	4	5	6				
ері	4	2	2	2	2				
S_homo	2	1	1	0	1				
S_epi	2	1	2	2	1				

\* Inclou la 3α-cetona".

Taula 15: Distribució de la informació estructural del model inicial i dels models reduïts.

També és important que l'activitat es distribueixi uniformement tant en el "data set" com en el "test set". La Figura 164 mostra la distribució de l'activitat de cada grup per cadascun dels models reduïts. S'observa que en el model *IZR-reduït* l'activitat dels compostos del "data set" s'acumula en els extrems del marge d'activitats mentre que en el model *MVP-reduït* s'acumula en el centre. En els models *SQR-reduït* i *HIP-reduït* l'activitat es distribueix de forma bastant homogènia entre els dos grups.


*Figura 164*: Distribució de l'activitat en cadascun dels models reduïts. En verd es mostra l'activitat dels compostos del *"data set"* i en blanc la dels del *"test set"*.

# 3. Discussió de resultats.

Per a la discussió dels models, només es considera el primer component principal ja que és el que millor correlaciona l'estructura i l'activitat. Els models *SQR-reduït*, *IZR-reduït* i *MVP-reduït* expliquen al voltant del 70% de l'activitat amb un error de càlcul que oscil·la entre  $\pm 0,46$  i  $\pm 0,53$ . El model *HIP-reduït* és millor ja que explica el 86% de l'activitat amb un error de càlcul de  $\pm 0,27$ .

Per avaluar la predictibilitat es fan servir dos mètodes: 1) Validació interna mitjançant el mètode de validació creuada L.O.O., i 2) Validació externa fent servir el *"test set"*.

En termes de validació interna, els models *SQR-reduït*, *IZR-reduït* i *MVP-reduït* no han mostrat cap capacitat predictiva ja que tots ells presenten valors del coeficient de correlació de la predicció (q<sup>2</sup>) negatius. Per contra, el model *HIP-reduït* es capaç de predir un 32% de l'activitat. Aquest valor podria ser superior ja que la major part de l'error de predicció es deu al compost epics\_2n3n. Aquest és el de menor activitat i l'únic que no presenta un hidroxil en l'anell A. Degut al propi mètode de validació creuada, el model on aquest compost es queda fora no considera l'absència de funcionalitats com un fet rellevant i per tant li sobreestima enormement la seva activitat.

En termes de validació externa, els models *SQR-reduït*, *MVP-reduït* i *HIP-reduït* no han mostrat cap capacitat predictiva ja que tots ells presenten desviacions estàndard de l'error de predicció molt elevades (SDEP oscil·la entre  $\pm 0,71$  i  $\pm 0,89$ ). Tots ells tendeixen a promediar l'activitat dels compostos del "test set". Per contra, el model *IZR-reduït* presenta un valor de SDEP de  $\pm 0,52$ , comparable a l'obtingut en els models amb 20 compostos. Aquest bon resultat, però, té trampa. Aquest model s'ha realitzat amb compostos dels extrems del marge d'activitat. La major part dels compostos del "test set" són els que tenen un valor d'activitat més proper al promig, el qual és el valor que el model prediu. Per tant, el valor de SDEP és enganyosament baix.

#### 4. Conclusions.

#### 4.1. Conclusions particulars de l'estudi.

El model *HIP-reduït* és el millor dels quatre models reduïts tant en l'explicació de les dades d'activitat com en la seva predicció mitjançant validació interna. Amb només 10 compostos, explica un 86% l'activitat dels compostos i comença a predir-la de forma acceptable amb un 32%. Quan es dobla el nombre de compostos, i per tant se li aporta més informació estructural, el model (Model *HIP*) arriba a explicar el 93% de l'activitat i a predir el 66%.

A mesura que els models han augmentat el nombre de compostos, la predictibilitat ha augmentat considerablement. No obstant, la informació estructural present en el model inicial i en el model reduït s'ha mantingut proporcionalment. Això implica que l'augment en la predictibilitat no ha de ser igual per tots els compostos, la qual cosa encaixa perfectament amb les observacions fetes en el tema anterior on la predictibilitat variava en funció de determinades característiques estructurals.

#### 4.2. Conclusions generals de l'estudi.

Són necessaris els 20 compostos per a que el model QSAR presenti una bona predictibilitat. Això comporta que la predictibilitat només es pugui avaluar amb mètodes de validació interna, amb les limitacions que això comporta.

Per millorar la predictibilitat del model és necessari augmentar el nombre de compostos presents en el *"data set"*. No obstant, aquest augment de compostos ha d'anar orientat a equilibrar estructuralment el model per tal que la predictibilitat estigui homogèniament repartida per totes les característiques estructurals dels BRs i evitar que en la predicció hi hagi conjunts de BRs que es prediguin millor que altres.

# <u>PART III: Model HOMO.</u>

# 1. Introducció.

En el tema anterior s'ha vist com de les cinc cadenes laterals tipus, el 50% dels BRs del model té la cadena tipus *homo*. Aquest fet no és casualitat, ja que la Secció d'Esteroides de l'IQS s'ha dedicat especialment a la síntesi d'anàlegs homobrassinoesteroides (cadena tipus *homo*) i 22*S*,23*S*-homobrassinoesteroides (cadena tipus *S\_homo*). Tots dos s'obtenen a partir l'estigmasterol (*Figura 165a*) mitjançant síntesi parcial.

La síntesi d'anàlegs epibrassinoesteroides (cadena tipus *epi*) i 22*S*,23*S*-epibrassinoesteroides (cadena tipus *S\_epi*) es realitza a partir de l'ergosterol (*Figura 165b*), el qual també és comercial. La síntesi és similar a la dels homobrassinoesteroides, però amb una etapa més, la reducció selectiva del doble enllaç  $\Delta^7$  de l'anell B.

La síntesi parcial de la brassinolida i els seus anàlegs (cadena tipus *bl*) hauria de partir del brassicasterol (*Figura 165c*). No obstant, aquest compost no és un compost abundant i no és comercial. La major part de les síntesis descrites parteixen de l'estigmasterol i degraden la cadena lateral per obtenir l'aldehid en C22. A partir d'aquest es construeix específicament la cadena de la brassinolida i posteriorment es funcionalitzen els anells A i B.



Anteriorment s'ha vist que, en general, els models tendeixen a promediar l'activitat dels homoBRs. Això es deu a que els models calculen i prediuen l'activitat segons la cadena lateral tipus i els anells A i B en la seqüència típica ( $2\alpha$ , $3\alpha$ -lactona,  $2\alpha$ , $3\alpha$ -cetona i cetones sense funcionalitat en la cara  $\alpha$  de l'anell A) sense contemplar ni els BRs amb unió A/B cis ni els  $2\alpha$ , $3\alpha$ -BRs sense carbonil en l'anell B. Només el model *MVP* es capaç d'explicar i predir l'activitat dels homoBRs.

A continuació es realitza un model on només es consideren els BRs amb cadena tipus *homo*. Això perme treure la influència que la cadena lateral exerceix en el model i incorporar nous anàlegs amb una gran diversitat funcional en l'anell A, amb la qual cosa s'espera trencar la tendència de  $2\alpha$ , $3\alpha$ -lactona,  $2\alpha$ , $3\alpha$ -cetona i cetones sense funcionalitat en la cara  $\alpha$  de l'anell A.

# 2. Model HOMO.

#### 2.1. BRs del "data set".

Per raons cronològiques, en el model *HOMO* es va fer servir la conformació activa *MVP*. No obstant, i sense que serveixi de precedent, en la realització d'aquest model la conformació activa és irrellevant ja que només hi ha una cadena lateral tipus comuna per tots els compostos.

El model *HOMO* conta amb 25 homoBRs, els quals només es diferencien entre si per les modificacions estructurals dels anells A i B (*Figura 166*). Dels 25 homoBRs, 10 ja estaven presents en el model *MVP*, mentre que 15 són nous. Aquests incorporen noves funcionalitats com són cetona,<sup>117</sup> fluor,<sup>114-118</sup> brom<sup>199</sup> i azida.<sup>115,117,119,120</sup>

En la *Figura 167* es representa l'activitat dels BRs del model *HOMO* en ordre decreixent i la distribució d'aquests valors dins del marge d'activitats. S'observa que la distribució dels valors d'activitat no és uniforme i bona part d'aquests valors es concentren al voltant de 1,50.



\* Valor d'activitat expressat com –log(dosi<sub>45°</sub>).

*Figura 166*: Anells A i B dels BRs del "data set" dels model *HOMO* i valor d'activitat expressat com –log(dosi<sub>50%</sub>). En verd s'indiquen els nous BRs i en granat els antics.



*Figura 167*: Activitat dels BRs del model *HOMO* i distribució dins del marge d'activitats.

# 2.2. Anàlisi de components principals (PCA).

Entre el primer, el segon components principals s'expliquen principalment els BRs amb funcionalitat azida, les quals són molt diferents a la resta de BRs. També s'explica l'anell B diferenciant les funcionalitats lactona i cetona entre si (*Figura 168a*). En total s'observen quatre grups: a)  $3\alpha$ -azides amb funcionalitat cetona, b)  $3\alpha$ -azides amb funcionalitat lactona, c) BRs amb funcionalitat cetona, i d) BRs amb funcionalitat lactona més BRs amb unió A/B cis. Aquests es troben explicats dos a dos en els components principals. Destacar certes particularitats:

- Igual que passa en el model *HIP*, l'anell B queda descrit a través del diol de la cadena lateral.
- Al contrari que passa en els models amb les cinc cadenes laterals tipus, els BRs amb unió A/B cis són més similars als BRs amb funcionalitat lactona, que no pas amb funcionalitat cetona.

El tercer component explica l'anell A dels BRs d'acord amb la seqüència  $2\alpha$ ,  $3\alpha$ -cetona, cetona amb una funcionalitat en la cara  $\alpha$  de l'esteroide, ja sigui  $2\alpha$  o  $3\alpha$ , i cetona sense grups funcionals en la cara  $\alpha$  de l'esteroide (*Figura 168b*).

El quart component principal explica, precisament, la diferencia entre els BRs amb funcionalitat lactona a l'anell B i els BRs amb unió A/B cis (*Figura 169a*).

Entre el cinquè i el sisè components principals s'explica la diversitat de l'anell A (*Figura* 169b). Concretament s'expliquen les diferències en la cara  $\beta$  de l'esteroide d'acord amb la seqüència 2 $\beta$ ,3 $\beta$ , 3 $\beta$  i sense funcionalitat en l'anell A. El sisè component també explica les diferències en la cara  $\alpha$  d'acord amb si té una funcionalitat en 2 $\alpha$  o en 3 $\alpha$ .



*Figura 168*: Model *HOMO*: *(a)* Gràfic d'objectes dels 2 primers components principals. *(b)* Gràfic d'objectes del tercer i quart components principals.



*Figura 169*: Model *HOMO*: *(a)* Gràfic d'objectes del primer i quart components principals. *(b)* Gràfic d'objectes del cinquè i sisè components principals.

De les noves funcionalitats afegides (cetona, fluor, brom i azida) el model només és capaç d'explicar les azides. No obstant, les tracta com una funcionalitat especial. Formen un grup per si soles i només es diferencien entre si per l'anell B. Els fluors en  $2\alpha$  o  $3\alpha$  els tracta com hidroxils, el brom en  $3\beta$  el tracta com hidrogen, i la cetona en C2 la tracta com l'hidroxil  $2\alpha$ . Malgrat haver fet servir la conformació activa *MVP*, el model *HOMO* no és capaç d'explicar els BRs sense funcionalitat carbonílica en l'anell B i els tracta com si tinguessin una cetona.

S'ha realitzat un anàlisi de components principals suprimint del "*data set*" els 4 BRs amb funcionalitat azida. El primer component explica les diferències en l'anell B i diferència dos grups: a) BRs amb funcionalitat lactona a l'anell B més BRs amb unió A/B cis, i b) BRs amb funcionalitat cetona a l'anell B. Els següents components expliquen els BRs amb unió A/B cis, i la distribució de funcionalitats en l'anell A de la mateixa forma que en el model que s'acaba d'explicar. Això reforça la impressió que les azides són un grup estrany.

# 2.3. Anàlisi de mínims quadrats parcials (PLS).

#### a) Paràmetres estadístics.

A la Taula 16 es mostren els paràmetres estadístics del model HOMO.

components	XVarExp	XAccum	SDEC	r <sup>2</sup>	SDEP	$q^2$ (LOO)
0	0,0	0,0	0,79	0,00	0,83	-0,09
1	12,7	12,7	0,58	0,47	0,81	-0,04
2	7,9	20,6	0,40	0,74	0,90	-0,27
3	8,4	29,1	0,24	0,91	0,87	-0,21
4	11,2	40,2	0,19	0,95	0,81	-0,04
5	6,1	46,3	0,15	0,96	0,74	0,13
18.862 variables						

Taula 16: Paràmetres estadístics del model HOMO.

El primer component només és capaç d'explicar un 47% de l'activitat amb un error de càlcul de  $\pm 0,58$ . El model té una capacitat predictiva nul·la ja que el coeficient de correlació de la predicció (q<sup>2</sup>) és negatiu.

#### b) <u>Contribucions a l'activitat.</u>

En la *Figura 170* es mostren els coeficients per al primer component principal a  $\pm 0,001$ . De la interpretació dels mapes de coeficients es conclou:

- Contribucions positives a l'activitat: les  $2\alpha$ ,  $3\alpha$ -lactones, absència de grups funcionals en  $2\beta$  però sobretot en  $3\beta$ , absència de funcionalitat en  $5\alpha$ .
- Contribucions negatives a l'activitat: les cetones sense grups funcionals en  $2\alpha$  i  $3\alpha$ .



*Figura 170*: Coeficients del model *HOMO*. En vermell es mostren les contribucions positives a l'activitat i en blau les contribucions negatives.

Exceptuant alguna petita diferència, les contribucions a l'activitat dels anells A i B són les mateixes que s'obtenien en els models amb les cinc cadenes laterals tipus. El model no és capaç d'explicar l'activitat ni de les noves funcionalitats afegides, ni de la distribució de funcionalitats en l'anell A, ni dels BRs sense funcionalitat carbonílica en l'anell B.

# c) <u>Predictibilitat.</u>

En la *Figura 171* es mostra el gràfic d'activitat predita davant l'activitat experimental. Com ja s'ha comentat, la predictibilitat és nul·la. De fet, el model promedia les dades d'activitat al voltant de 1,5. Aquest fet és conseqüència de que l'activitat experimental dels BRs no és homogènia i està distribuïda al voltant d'aquest valor (*Figura 167*, pàgina 183). Llavors, el model no té la necessitat d'explicar bona part de la diversitat estructural incorporada pels nous BRs. Posteriorment, com no explica aquesta diversitat estructural, el model promedia l'activitat predita al voltant de 1,5 i es tanca el cercle.



*Figura 171*: Model *HOMO*: Activitat predita davant activitat experimental.

# 3. Conclusions.

# 3.1. Conclusions particulars de l'estudi.

De les noves funcionalitats afegides (cetona, fluor, brom i azida), el model només explica els BRs amb funcionalitat azida. No obstant, els explica com un grup a banda, independent. El comportament de les funcionalitats cetona i fluor és similar a un hidroxil, i el de la funcionalitat brom a un hidrogen. El comportament dels BRs sense carbonil a l'anell B és similar als que tenen un carbonil.

El model *HOMO* ha aconseguit explicar la variabilitat dels anells A i B per separat, sense interaccions. Primer, explica la diferència en l'anell B segons els BRs tinguin funcionalitat lactona o cetona. En segon lloc, explica les cetones segons tinguin dos, una o cap funcionalitat en la cara  $\alpha$  de l'anell A. En tercer lloc, explica la diferència entre els BRs amb funcionalitat lactona i els que tenen la unió A/B cis. Finalment, explica la distribució de grups funcionals en l'anell A independentment de la funcionalitat en l'anell B.

L'augment de la variabilitat en l'anell A porta associat un augment de valors d'activitat situats en la zona intermèdia. Això fa que, més enllà de la seqüència  $2\alpha$ , $3\alpha$ -lactona,  $2\alpha$ , $3\alpha$ -cetona i cetones sense funcionalitat en la cara  $\alpha$  de l'anell A, no hi hagi cap relació quantitativa, entre la variabilitat en l'anell A i l'activitat.

# 3.2. Conclusions generals de l'estudi.

Aquest estudi dins del conjunt global de la tesi suggereix que per tenir una millor informació sobre l'anell A és necessari augmentar racionalment el nombre de compostos. Cal incloure BRs amb modificacions estructurals a l'anell A per les cinc cadenes laterals tipus, evitant interaccions estructurals entre els anells i les cadenes laterals.

# <u>PART IV: Model 1 µa/planta.</u>200

# 1. Introducció.

En la bibliografia, sovint es confonen els termes "activitat" i "resposta" i es parla d'activitat quan s'hauria de parlar de resposta. La resposta és l'efecte que s'aconsegueix en realitzar el bioassaig a una concentració o dosi determinada. L'activitat és quelcom més complex i està relacionada amb la dosi a la qual la resposta és el 50% de la resposta màxima, LD<sub>50%</sub> (pàgina 53). La confusió entre els dos termes radica en el fet que els compostos que presenten una resposta elevada a una determinada dosi són actius i per tant, hi ha la tendència d'associar resposta a activitat.

El estudis QSAR es recomana fer servir la dada d'activitat (LD<sub>50%</sub>) per establir la correlació. Aquest valor, però, no sempre està disponible i en substitució, s'utilitza la *resposta* a una dosi concreta. Com s'explicarà tot seguit, considerar l'activitat o la *resposta* no és indiferent al QSAR ja que segons si es fa servir una dada o l'altre es donen solucions a preguntes diferents. Per il·lustrar-ho, a continuació s'exposen dos casos en el camp dels BRs on activitat i *resposta* presenten incompatibilitats entre si.

# • <u>Cas 1:</u>

Els BRs amb la cadena lateral saturada, o amb un sol hidroxil (*Figura 172*) serien uns bons candidats per ampliar el conjunt de BRs dels models, ja que augmentarien la diversitat estructural en la cadena lateral, ampliant el marge de coneixement en aquesta zona. Cap d'aquestes modificacions de cadena es troba present en els models, per tant la predicció de l'activitat d'aquests anàlegs comporta una extrapolació.



Figura 172: BRs amb cadena lateral saturada o amb un sol hidroxil.

En el cas de la hcs\_2n3n\_22n23n, epics\_22n23n i S\_hcs\_22n el model *HIP* els prediu unes *activitats* (–log(dosi<sub>50%</sub>)) entre 0,9 i 1,7. Aquests valors signifiquen que a 1 µg/planta les *respostes* haurien d'estar entre 80º i 90º. No obstant, les seves *respostes* oscil·len entre 1º i 28º. Per tant, el model QSAR prediu com actius compostos que no ho son.

En el cas de la hcs\_22n23n, el model *HIP* li prediu una *activitat* ( $-\log(dosi_{50\%})$ ) de 2,35. La seva activitat experimental és de 0,73. Tot i que el compost és actiu i el model el prediu com a tal, la seva activitat predita expressada com a dosi al 50% és 40 vegades superior a la real.

El cas de la S\_hcs\_23n, és l'excepció. El model *HIP* li prediu una *activitat* ( $-\log(dosi_{50\%})$ ) de 0,82 i la seva activitat experimental és de 0,50. Tot i que la diferència és moderada (±0,32), està dins dels marges d'error de la predicció del model (±0,52).

En aquest primer cas, en el qual la predicció comporta una extrapolació, s'observa que en general, el model tendeix a sobreestimar l'activitat. Aquesta sobreestimació porta a considerar actius els tres compostos que són inactius i com a molt actiu un compost que és moderadament actiu.

#### • <u>Cas 2:</u>

Igual que en el cas anterior, els BRs S\_epics\_2b3b i S\_epics\_2n3n també serien uns bons candidats per ampliar el conjunt de BRs dels models, ja que ajudarien a reduir la proporció de cadenes laterals tipus *homo* i augmentarien el nombre de BRs amb funcionalitat diversa a l'anell A. Aquests dos anàlegs seran fàcilment acceptats pel model ja que tenen un lloc assignat dins l'espai dels objectes del PCA del model *HIP* (*Figura 173*), per tant la predicció de l'activitat d'aquests anàlegs comporta una interpolació en el model.

L'activitat predita de tots dos ( $-\log(dosi_{50\%})$ ) està al voltant de -0,20. Tot i que l'activitat és inclús inferior a l'activitat més baixa present en el model (0,71 de l'hcs\_eter) és coherent amb la interpretació dels coeficients. El valor predit significa que a 1,5 µg/planta les *respostes* de tots dos compostos haurien d'estar al voltant de 45°. Les seves *respostes* a 1 µg/planta estan al voltant dels 5°. Degut a la limitació del bioassaig (pàgina 198) es desconeixen les respostes a 1,5 µg/planta i en conseqüència no es pot corroborar si la predicció feta pel model és encertada o no.



*Figura 173*: Situació dels BRs S\_epics\_2b3b i S\_epics\_2n3n dins dels gràfics d'objectes dels 2 primers components principals del model *HIP*: (*a*) Coloració segons cadena lateral tipus i (*b*) coloració segons els anells A i B.

En aquest segon cas, en el qual la predicció comporta una interpolació, el model prediu un valor d'activitat coherent. No obstant, la predicció xoca amb la limitació del bioassaig i experimentalment no és possible corroborar si el compost és actiu tal com es preveu segons el model, o és inactiu tal com es dedueix d'unes dades experimentals limitades. En els dos casos anteriors, s'ha vist com el model QSAR no és capaç d'explicar l'*activitat*  $(-\log(dosi_{50\%}))$  de BRs que tenen una baixa *resposta* (1 µg/planta). La raó és que al model se li està preguntant quelcom que no pot resoldre.

Els models QSAR anteriors correlacionen l'estructura amb el valor d'activitat  $(-\log(dosi_{50\%}) \circ LD_{50\%})$ . Aquesta dada descriu l'afinitat per substrat del receptor. Per tant, els coeficients dels models expliquen els requeriments estructurals per a que un compost actiu sigui més o menys actiu. Tots els compostos presents en els models QSAR són actius, en major o menor grau, i tots ells presenten una *resposta* màxima a una dosi suficientment elevada. Els compostos amb una *resposta baixa* o *inactius* són ignorats pel simple fet que no es pot determinar el seu valor d'activitat. Per tant, a aquests models no se'ls pot preguntar per la *resposta* dels compostos *inactius* perquè no tindrà "resposta".

L'activitat o inactivitat d'un compost ve determinada per l'especificitat per substrat del receptor vers aquest compost, la qual es descriu a través de la *resposta* a la concentració màxima aplicable. Un model que correlacioni l'estructura amb la resposta serà capaç d'explicar els requeriments estructurals que fan que un compost sigui actiu o inactiu.

Fins aquest moment, tots els models QSAR han correlacionat l'estructura dels BRs amb el seu valor d'activitat ( $-\log(dosi_{50\%})$ ) donant explicació a l'*afinitat per substrat*. A continuació es realitzarà un model QSAR que correlacioni l'estructura dels BRs amb la *resposta* a 1 µg/planta per donar explicació a l'*especificitat per substrat* (*Figura 174*).





*Figura 174*: Corbes dosi resposta dels BRs: Resposta a 1  $\mu$ g/planta i dades d'activitat expressades com a  $-\log(dosi_{45^{\circ}})$ .

# 2. Model a 1 µg/planta.

#### 2.1. BRs del "data set" i resposta.

La *Taula 17* mostren els BRs que formen part del model organitzats en ordre decreixent segons la resposta a  $1 \mu g/planta$ . Per a la realització del model QSAR s'utilitza la conformació activa *HIP*.

Fer servir la resposta a 1 µg/planta enlloc de l'activitat ( $-\log(dosi_{50\%})$ ) permet incloure 53 BRs en el model. S'augmenta el nombre de compostos amb cadenes laterals tipus *epi*, *S\_homo* i *S\_epi*. S'augmenta també el nombre de BRs sense el diol 2 $\alpha$ ,3 $\alpha$  en l'anell A. S'incorporen BRs amb un hidroxil  $\alpha$  en C11 i BRs amb cadenes laterals sense el diol en C22,C23. De moment no s'incorporen els BRs amb fluor o azida.

BRs	1 μg/planta	BRs	1 µg∕planta	BRs	1 μg/planta
hbl	116	S_hcs	75	hcs_2n3n_22n23	3n 28
bl	114	S_epibl	74	hcs_2n3Br_5OH	H 24
CS	103	hcs_60Hb	74	epics_2n3b_23	า 24
epics_2n3b	101	hcs_2n3b	74	S_hcs_22n	22
hcs_2CO3n	99	hcs_2n3n	73	S_epics_2n3b_110	OH 22
hbl_cis	98	hcs_2b3b	72	S_epics_2n3b_5C	DH 21
S_hcs_23n	93	epics_2n3n	72	S_hcs_2CO3n	18
hcs	92	S_epics	67	S_hcs_2n3Br	9
hcs_2a3n	88	epibl_11OH	65	S_epics_11OH	8
epibl	88	S_epibl_11OH	64	epics_2n3b_22i	า 7
hcs_2n3a	85	epics_2b3b	64	S_epics_2b3b	6
hcs_cis	84	hcs_60Ha	62	S_epics_2n3n	5
epics_11OH	82	hcs_eter	57	S_hcs_2n3Br_5C	H 4
hcs_50H	81	S_hcs_2a3n	56	hbl_2n3n_22n23	Bn 4
S_hbl	80	S_hcs_5OH	45	S_hcs_2n3b_5O	H 3,7
hcs_22n23n	79	hcs_2n3Br	42	epics_22n23n	1
epics	79	hbl_2b3b	36	S_hcs_eter	-2
epics_2n3b_11OH	76	hcs_2b3b_5OH	31		

*Taula 17*: Conjunt de BRs presents en els models. Nom i resposta a 1 µg/planta.

#### 2.2. Reducció de compostos.

En el model inicial, dos compostos presenten un comportament anòmal i es suprimeixen del model. En reduccions successives es suprimeixen fins a 12 BRs els quals es poden classificar en sis grups:

- a) 22*R*,23*R*-BRs que sense tenir cap grup funcional en la cara  $\alpha$  de l'anell A presenten una resposta molt elevada: epics\_2n3b (101<sup>o</sup>), epics\_2n3b\_11OH (76<sup>o</sup>) i hcs\_2n3n (73<sup>o</sup>).
- b) BRs que sense tenir el diol de cadena presenten una resposta molt elevada: S\_hcs\_23n (93<sup>2</sup>).
- c) BRs que tenint el diol  $2\alpha$ ,  $3\alpha$  presenten una resposta molt baixa: S\_epics\_11OH (8<sup>o</sup>) i epics\_22n23n (1<sup>o</sup>).
- d) BRs amb una funcionalitat èter en l'anell B: hcs\_eter (57<sup>o</sup>) i S\_hcs\_eter (-2<sup>o</sup>).
- e) BRs on la funcionalitat lactona en l'anell B presenta una resposta menor que la funcionalitat cetona: epibl\_11OH (65<sup>o</sup>), hbl\_2b3b (36<sup>o</sup>) i hbl\_2n3n\_22n23n (4<sup>o</sup>).
- f) BRs que simplement es surten de la tendència:  $hcs_2n3Br (42^{\circ})$ .

# 2.3. Anàlisi de mínims quadrats parcials (PLS).

a) Paràmetres estadístics.

Finalment es realitza un model amb els 41 compostos. Només els dos primers components principals són interpretables des del punt de vista estructural. A la *Taula 18* es mostren els paràmetres estadístics del model.

components	XVarExp	XAccum	SDEC	r <sup>2</sup>	SDEP	q <sup>2</sup> (LOO)
0	0,0	0,0	33,8	0,00	34,6	-0,05
1	29,0	29,0	9,9	0,91	11,2	0,89
2	11,3	40,3	7,3	0,95	9,6	0,92
3	5,4	45,6	5,6	0,97	9,5	0,92
4	3,1	48,7	4,1	0,98	9,8	0,92
5	3,9	52,7	3,2	0,99	9,8	0,92
2.578 variables						

*Taula 18*: Paràmetres estadístics del model a 1 µg/planta.

#### b) Interpretació dels coeficients.

El primer component principal correlaciona l'estructura dels BRs amb la seva resposta a 1 µg/planta (*Figura 175a*). Amb un 29% de la variància estructural corresponent a les zones d'interacció dels anells A i B, el diol de cadena lateral i el volum dels anells C, D i metil C21, el model és capaç d'explicar un 91% de la resposta a 1 µg/planta amb un error de càlcul de ±10°, i predir-ne un 89% amb un error de predicció de ±11°.

El segon component principal corregeix a l'alça els  $2\alpha$ , $3\alpha$ ,225,235-BRs i a la baixa els 22R,23R-BRs sense cap funcionalitat polar en  $\alpha$  de l'anell A (*Figura 175b*).

En el gràfic d'objectes dels dos primers components principals, s'observen 6 grups ben definits en funció del nombre d'hidroxils en  $\alpha$  de l'anell A i de si són 22*R*,23*R* o 22*S*,23*S*. En funció del nombre d'hidroxils en  $\alpha$  de l'anell A es formen tres grups: a dalt a l'esquerra els BRs amb cap hidroxil en  $\alpha$ , a baix a la dreta els BRs amb dos hidroxils en  $\alpha$ , i enmig queden els BRs que tenen un hidroxil en  $\alpha$  i els BRs amb unió A/B *cis*. En funció del diol de la cadena lateral s'agrupen a dalt a la dreta els 22*R*,23*R*-BRs i a baix a l'esquerra els 22*S*,23*S*-BRs. Malauradament, el grup de BRs sense diol en la cadena lateral no queda ben explicat i els compostos es distribueixen aleatòriament.



**Figura 175**: Model a 1 µg/planta: **a**) Gràfic de correlació interna del primer component principal i **b**) Gràfic d'objectes per als dos primers components principals.

#### c) <u>Contribucions a l'activitat.</u>

En el model a 1 µg/planta, cal considerar fins al segon component principal, ja que es pot interpretar la relació existent entre l'estructura i l'activitat. El model final explica un 95% de la resposta a 1 µg/planta amb un error de càlcul de  $\pm 7^{\circ}$ , i prediu un 92% amb un error de predicció de  $\pm 10^{\circ}$ , fent servir el 40% de la variabilitat estructural.

En la *Figura 176* es mostren els coeficients per al segon component principal a  $\pm 0,001$ . De la interpretació dels mapes de coeficients es conclou:

- Contribucions positives a la resposta a 1  $\mu$ g/planta: el diol 2 $\alpha$ ,3 $\alpha$ , un carbonil a l'anell B amb una major contribució del carbonil de la lactona, el diol de la cadena lateral, l'absència de grups funcionals en 3 $\beta$ , l'absència de grups funcionals en 5 $\alpha$ , i l'anell C i metil en C21 dels 22*R*,23*R*-BRs.
- Contribucions negatives a la resposta a 1  $\mu$ g/planta: l'absència de grups funcionals en 2 $\alpha$  i 3 $\alpha$ , algunes parts del carbonil de les cetones en l'anell B, els anells B i D dels 22*S*,23*S*-BRs, i l'absència de grup funcional en C22.
- Contribucions nul·les a la resposta a 1 μg/planta: la zona d'interacció de l'hidroxil en 5α, l'hidroxil en C11, i les cadenes laterals tipus més enllà de la distinció entre 22*R*,23*R*-BRs i 22*S*,23*S*-BRs.



Sonda d'aiguaSonda de metilFigura 176: Coeficients del model a 1 μg/planta. En vermell es mostren les<br/>contribucions positives a la resposta a 1 μg/planta i en blau les contribucions negatives.

d) <u>Predictibilitat.</u>

En la *Figura 177* es mostra el gràfic de resposta predita a 1  $\mu$ g/planta davant de resposta experimental. La predictibilitat del model és molt bona (92%, amb un error de ±10°). No s'observa cap dependència entre la predictibilitat i l'estructura dels BRs.



*Figura 177*: Model a 1  $\mu$ g/planta: Gràfic de resposta predita davant de resposta experimental.

# 3. Conclusions.

El model a 1 µg/planta descriu l'especificitat per substrat, explicant els requeriments estructurals que fan que un BRs sigui actiu o inactiu. El model HIP descriu l'afinitat per substrat, explicant els requeriments estructurals que fan que un BRs actiu sigui més o menys actiu.

Es proposa el següent protocol per predir l'activitat dels BRs: Amb el model a 1 µg/planta és prediu la *resposta* a 1 µg/planta. Si aquesta és superior a  $30^\circ$ , es considera que el BR pot presentar activitat com a promotor del creixement vegetal. En aquest cas, amb el model *HIP* es prediu l'*activitat* expressada com el  $-\log(dosi_{50\%})$ . Caldrà desconfiar del valor en el cas que la *resposta* sigui baixa i l'*activitat* elevada.

# DISCUSSIÓ DE RESULTATS: CAPÍTOL 4

# <u>ACTIVITAT</u>

#### 1. Introducció.

Com s'ha vist en el tema dels androstans, un dels objectius dels models QSAR és poder predir l'activitat de compostos no sintetitzats. Si l'activitat predita resulta interessant, s'aborda la seva síntesi i sinó es descarta. Això permet estalviar un gran esforç sintètic. Com s'ha vist en el tema anterior, la capacitat predictiva dels models es mesura a través de dos paràmetres: el coeficient de correlació de la predicció (q<sup>2</sup>) i la desviació estàndard de l'error de predicció (SDEP). Precisament aquest últim correspon a l'error de predicció del model.

En el cas dels BRs es tenen dos models QSAR complementaris. El model a 1 µg/planta que prediu la resposta a 1 µg/planta amb un error de  $\pm 10^{\circ}$  i el model *HIP* que prediu l'activitat com  $-\log(dosi_{50\%})$  amb un error de  $\pm 0,52$ . Seria molt interessant conèixer l'error experimental tan de la resposta a 1 µg/planta com de l'activitat, per comparar-lo tan amb l'error de càlcul (desviació estàndard de l'error de càlcul, SDEC) com amb l'error de predicció dels dos models.

A continuació s'explica l'obtenció de l'activitat experimental dels BRs fent servir el test d'inclinació de la làmina d'arròs.

# 2. Test d'inclinació de la làmina d'arròs.

#### 2.1. Protocol.

El protocol del test d'inclinació de la làmina d'arròs (RLIT) que actualment es segueix al Laboratori d'Esteroides de l'IQS consta de quatre etapes (*Figura 178*):<sup>90,97,118,201-205</sup>

- 1. Germinació: Es submergeixen les llavors d'arròs en aigua desionitzada i es tenen 48 hores en un una cambra de cultiu a 30º i un fotoperíode de 16 hores de llum i 8 de foscor.
- 2. Sembra: Es planten les llavors germinades en un gel d'agar-agar 0,5% posant la radícula cap amunt en contacte amb la superfície. Es col·loquen en un recipient que manté la humitat relativa al voltant del 90% i es tenen 48 hores en un una cambra de cultiu a 30º i un fotoperíode de 16 hores de llum i 8 de foscor.
- 3. Aplicació: En el nus de desenvolupament de la segona làmina, s'injecten 0,5  $\mu$ l d'una dissolució de BR en etanol 95%. Les plantes es tenen 48 hores en un una cambra de cultiu a 30º i un fotoperíode de 24 hores de foscor.
- 4. Mesura: Es talla la planta i de seguida es dibuixa sobre un paper l'angle format per l'anvers de la fulla i la tija. Amb un transportador d'angles es realitza la mesura final.

En cada experiment, el nombre de plantes a les quals s'aplica una mateixa dosi de BR oscil·la entre 10 i 15. A més a més, es reserven un nombre similar de plantes on s'aplica homocastasterona (hcs) a 1 µg/planta que serveixen com a referència interna, i un altre nombre similar de plantes sense tractar que serveixen de control.



Figura 178: Esquema del test d'inclinació de la làmina d'arròs.

#### 2.2. Fonts d'error i limitacions del test.

Existeixen una sèrie de factors que poden influir en la dispersió dels resultats. Aquests poden ser deguts a la pròpia metodologia del bioassaig (factors extrínsecs) i/o propis de la planta d'arròs (factors intrínsecs):

- a) Factors extrínsecs: experimentador, envelliment de les llavors, condicions de la cambra de cultiu (llum, temperatura i humitat relativa), aplicació de la dosi correcte (errors en la pesada, dissolució, pressa de mostra, injecció i variació de la concentració per evaporació del dissolvent), mesura de l'angle, etcètera.
- b) Factors intrínsecs: la dispersió deguda al propi creixement de les plantes.

Els factors extrínsecs es controlen amb la resposta de la homocastasterona a 1  $\mu$ g/planta, mentre que els intrínsecs es controlen amb la resposta del control.

La gran limitació del bioassaig és que difícilment es poden aplicar dosis superiors a 1  $\mu$ g/planta. Això és degut a la falta d'assimilació del BR per la planta, arribant a observar-se una taca blanca en el punt d'aplicació. Ocasionalment s'han arribat a aplicar dosis de 2  $\mu$ g/planta.

#### 2.3. Tractament estadístic de les dades.

Es realitza a dos nivells: 1) en cada experiment i 2) en cada repetició de l'experiment.

#### 1) Tractament estadístic en cada experiment.

S'ordenen per ordre creixent les respostes, es a dir els angles, de cada BR. Es suprimeixen els valors extrems (màxim i mínim) i es calcula la mitjana ( $\bar{x}_i$ , equació 5) i la desviació estàndard ( $\sigma_i$ , equació 6).

equació 5: 
$$\overline{x}_i = \frac{\sum_{n=1}^{N_i} x_n}{N_i}$$
; equació 6:  $\sigma_i = \sqrt{\frac{\sum_{n=1}^{N_i} (x_n - \overline{x}_i)^2}{N_i - 1}}$ 

on  $x_n$  és la resposta individual, i  $N_i$  el nombre total de plantes assajades.

Per normalitzar les dades cal restar el control. La resposta normalitzada (M - Mc) s'obté restant la mitjana del control (Mc) a la mitjana de les dades (M). La desviació estàndard associada és  $\sigma(M - Mc)$  o  $\sigma \bar{x}_{ic}$  la qual es calcula segons la propietat de la variància que es defineix a continuació:

Si x i y són dos variables aleatòries independents, llavors es compleix l'equació 7.

equació 7: 
$$\sigma^2(ax-by) = a^2\sigma^2x + b^2\sigma^2y$$

Assumint que x correspon a la mitjana del compost  $(\bar{x}_i)$ , y a la mitjana del control  $(\bar{x}_c)$ , que la diferència correspon a l'equació 8 i sabent que les desviacions estàndard d'aquestes variables venen representades per les equacions 9a i 9b, la variància pren la forma de l'equació 10 i la desviació estàndard de l'equació 11.

equació 8: 
$$\overline{x}_{ic} = \overline{x}_i - \overline{x}_c$$

equació 9a: 
$$\sigma x = \sigma \overline{x}_i = \frac{\sigma_i}{\sqrt{N_i}}$$
; equació 9b:  $\sigma y = \sigma \overline{x}_c = \frac{\sigma_c}{\sqrt{N_c}}$   
equació 10:  $\sigma^2 \overline{x}_{ic} = \sigma^2 (\overline{x}_i - \overline{x}_c) = \sigma^2 \overline{x}_i - \sigma^2 \overline{x}_c = \frac{\sigma_i^2}{N_i} + \frac{\sigma_c^2}{N_c}$   
equació 11:  $\sigma \overline{x}_{ic} = \sqrt{\frac{\sigma_i^2}{N_i} + \frac{\sigma_c^2}{N_c}}$ 

Aquesta desviació estàndard normalitzada ( $\sigma \bar{x}_{ic}$ ) rep el nom de repetitibilitat (*REPET*) i dóna idea de la variabilitat de la resposta d'un determinat BR a una concentració concreta en un experiment concret.

#### 2) Tractament estadístic per cada repetició de l'experiment.

Es realitzen diferents experiments per cada BR a cada dosi. Per cadascun s'han obtingut els valors de la mitjana (M), desviació estàndard ( $\sigma$ ), mitjana normalitzada (M - Mc) i repetitibilitat (REPET). Sobre aquests valors es realitza un nou tractament estadístic. Es calcula la mitjana global (MITJA, equació 12) a la qual se l'associen dos desviacions estàndard: 1) la desviació estàndard corresponent a les mitjanes dels experiments, també anomenat reproductibilitat (REPROD, equació 13) i 2) la mitjana de les desviacions estàndard normalitzades dels experiments (REPET MITJ, equació 14).

equació 12: 
$$MITJA: \ \bar{x}_{ik} = \frac{\sum_{j=1}^{L_i} (M - Mc)}{L_i}$$
  
equació 13:  $REPROD = \sqrt{\frac{\sum_{i=1}^{L_i} ((M - Mc) - MITJA)^2}{L_i - 1}}$   
equació 14:  $REPET MITJ = \frac{\sum_{n=1}^{L_i} REPET}{L_i}$ 

on  $L_i$  és el nombre total d'experiments.

#### 2.4. Corba dosi-resposta.

Quan la resposta (expressada com a *MITJA*) a 1 µg/planta d'un BR és superior a 45º es considera que un BR és actiu i es realitzen assajos a diferents dosi per representar i construir la corba que relaciona el logaritme de la dosi amb la resposta. A les dades s'ajusta una funció sigmoïdal (*Figura 179*). Els paràmetres que la descriuen són: el valor màxim ( $y_{max}$ ), el valor mínim ( $y_{min}$ ), l'abscissa corresponent al punt intermedi entre el màxim i el mínim ( $x_{50}$ ) i el pendent (K).



Figura 179: Equació i exemple de corba dosi-resposta.

L'ajust es realitza amb el programa FigP,<sup>206</sup> el qual dóna els paràmetres  $y_{max}$ ,  $y_{min}$ , K, i  $x_{50}$ . Precisament el paràmetre  $x_{50}$  canviat de signe correspon al menys logaritme de la dosi a la qual la resposta és el 50% de la resposta màxima  $(-\log(dosi_{50\%}) = -x_{50})$ . Aquest és el valor d'activitat que cal prendre en els estudis de QSAR, ja que està relacionat amb l'energia lliure de Gibbs del procés que va des de que s'injecta el BR fins que es provoca l'obertura de la fulla. No obstant, a l'equip s'ha fet servir tradicionalment com a valor d'activitat el  $-\log(dosi_{45\%})$  enlloc del  $-\log(dosi_{50\%})$ . Aquest valor té l'origen en que per alguns BRs no es pot determinar el màxim de la corba, ja que aquest queda més enllà de la dosi màxima aplicable (*Figura 180*). Com per la majoria de BRs actius la *resposta* màxima està al voltant de 90° es considera que el 50% de la resposta màxima correspon a 45°. Per tant, s'associa 50% a 45° i es té un valor d'activitat que és independent de la qualitat de l'ajust.



*Figura 180*: Exemple de corba incompleta amb el màxim indefinit degut a que està per sobre de la dosi màxima aplicable, que en aquest cas és de 2  $\mu$ g/planta.

#### 3. Error experimental.

En tot l'apartat anterior no s'arriba a concretar ni l'error experimental de la resposta a un dosi concreta, ni l'error experimental de l'activitat ja sigui expressada com a  $-\log(dosi_{50\%})$  o com a  $-\log(dosi_{45^{2}})$ . El tema s'aborda a dos nivells: 1) error de la resposta a una dosi concreta i 2) error de la corba i del valor d'activitat.

#### 3.1. Error experimental de la resposta a una dosi concreta.

#### a) Error en la resposta d'un únic experiment.

Com ja s'ha comentat, els resultats de cada experiment s'ordenen per ordre creixent i es calcula la mitjana  $(\bar{x}_i)$  i la desviació estàndard de les dades  $(s_i)$ . Tant la mitjana com la desviació estàndard són més precisos quan major és el nombre de plantes assajades  $(n_i)$ . Per determinar el grau d'incertesa de la mitjana es pot calcular l'error estàndard de la mitjana  $(\bar{x}_i)$  equació 15). A partir d'ara, els resultats es presentaran a traves de la mitjana  $(\bar{x}_i)$  més/menys el seu error estàndard  $(s_{\bar{x}_i})$  i en el cas que es cregui oportú també es presentarà la desviació estàndard de les dades  $(s_i)$ .

equació 15: 
$$S_{\bar{x}_i} = \frac{S_i}{\sqrt{n_i}}$$

on  $s_i$  és la desviació estàndard de les dades i  $n_i$  el nombre total de plantes assajades.

El següent pas consisteix únicament en restar el control. Anteriorment aquesta etapa s'anomenava "normalització" de les dades.<sup>90,118,201-205</sup> No obstant, la terminologia no és l'adequada i a partir d'ara es parlarà de "rectificació" de les dades.

La resposta final, rectificada, d'un experiment ( $\overline{\alpha}_i$ ) és la diferència entre la mitjana de les dades ( $\overline{x}_i$ ) i la mitjana del control ( $\overline{x}_c$ ). Tots dos valors tenen associat un error estàndard ( $s_{\overline{x}_i}$  i  $s_{\overline{x}_c}$  respectivament). L'error estàndard de la resposta ( $s_{\overline{\alpha}_i}$ ) es calcula a partir de l'error estàndard de les dades i del control aplicant les regles de transmissió de l'error. Així doncs, la resposta final i el seu error s'expressen mitjançant l'equació 16. A l'error estàndard de la resposta ( $s_{\overline{\alpha}_i}$ ) se l'anomenava anteriorment *REPET*.

equació 16: 
$$\overline{\alpha}_i = \overline{x}_i - \overline{x}_c \quad \pm \sqrt{s_{\overline{x}_i}^2 + s_{\overline{x}_c}^2}$$

#### b) Error en la resposta de diversos experiments.

Per una dosi concreta es fan repeticions  $(n_j)$ . La resposta experimental resultant  $(\overline{\alpha}_{exp})$ s'obté com la mitjana de les respostes de cada experiment  $(\overline{\alpha}_i)$ . Cada experiment té associat un error estàndard  $(s_{\overline{\alpha}_i})$ . L'error estàndard de la resposta experimental resultant  $(s_{\overline{\alpha}_{exp}})$  es calcula a partir de l'error estàndard de cada experiment aplicant les regles de transmissió de l'error i correspon a l'error experimental de la resposta a una dosi. A partir d'ara, la resposta experimental resultant i el seu error s'expressen mitjançant l'equació 17.

equació 17: 
$$\overline{\alpha}_{exp} = \frac{\sum_{i=1}^{j} \overline{\alpha}_{i}}{n_{j}} \pm \frac{\sqrt{\sum_{i=1}^{j} s_{\overline{\alpha}_{i}}^{2}}}{n_{j}}$$

#### 3.2. Error experimental del valor d'activitat.

Coneixent la resposta a diferents dosis es pot ajustar la corba dosi-resposta, la qual segueix una funció sigmoïdal (*Figura 179*, pàgina 200). L'ajust es realitza amb el programa FigP,<sup>206</sup> el qual no només dóna els paràmetres que descriuen la corba ( $y_{max}$ ,  $y_{min}$ , K, i  $x_{50}$ ) sinó que també dóna els seus errors estàndards ( $s_{y_{max}}$ ,  $s_{y_{min}}$ ,  $s_{K}$ , i  $s_{x_{50}}$ ). Com s'ha

comentat anteriorment, el valor d'activitat  $-\log(\text{dosi}_{50\%})$  es correspon amb el paràmetre  $x_{50}$  canviat de signe. Per tant, l'error experimental del valor d'activitat es correspon directament amb el paràmetre  $s_{x_{50}}$  (equació 18).

equació 18: 
$$-log(dosi_{50\%}) = -x_{50} \pm s_{x_{50}}$$

Tradicionalment a l'equip s'ha fet servir com a valor d'activitat el  $-\log(dosi_{45^{\circ}})$  enlloc del  $-\log(dosi_{50\%})$ . Aquest es calcula mitjançant l'equació 19. L'error experimental  $(s_{-log(dosi_{45^{\circ}})})$  es calcula a partir dels errors estàndards dels paràmetres de la corba  $(s_{y_{max}}, s_{y_{min}}, s_{K}, i s_{x_{50}})$  i de l'error estàndard de la resposta a 45°  $(s_{45^{\circ}})$ , aplicant les regles de transmissió de l'error (equació 20). Malgrat es desconeix l'error estàndard de la resposta a 45°  $(s_{45^{\circ}})$ , una bona estimació és el valor de l'error estàndard experimental de la resposta experimental més propera a 45°  $(s_{\overline{\alpha}_{rn} \to 45^{\circ}})$ .

equació 19: 
$$-log(dosi_{45^{\circ}}) = -\left[x_{50\%} - \frac{1}{K}ln\left(\frac{y_{\max} - 45^{\circ}}{45^{\circ} - y_{\min}}\right)\right]$$

$$equació 20: \ s_{-log(dosi_{45^\circ})} = \sqrt{s_{x_{50}}^2 + \left[\frac{1}{K}ln\left(\frac{y_{\max} - 45^\circ}{45^\circ - y_{\min}}\right)\right]^2 \cdot \left(\frac{s_K}{K}\right)^2 + \frac{1}{K^2} \cdot \left[\frac{s_{y_{\max}}^2 + s_{45^\circ}^2}{\left(y_{\max} - 45^\circ\right)^2} + \frac{s_{y_{\min}}^2 + s_{45^\circ}^2}{\left(45^\circ - y_{\min}\right)^2}\right]^2}\right]$$

#### 4. Factors que influeixen en la corba dosi-resposta.

Un cop descrits els errors experimentals cal veure la influència que determinats factors poden tenir sobre la resposta a una dosi concreta i el seu efecte sobre la corba dosi-resposta i el valor d'activitat.

#### 4.1. Nombre de repeticions i experiments erronis.

Un punt crític és la discussió sobre el nombre de repeticions de cada bioassaig així com la supressió d'experiments erronis. Per il·lustrar aquesta discussió, en la *Taula 19* es mostra, a mode d'exemple, els resultats per la bl a  $10^{-5}$  µg/planta.

Experiment	Resposta ( $\overline{\alpha}_i$ )
COMB-58	26,0
COMB-60	72,3
COMB-62	20,6
COMB-65	62,6
COMB-69	51,4

**Taula 19:** Resultats de la bl a  $10^{-5} \mu g/planta$ .

En els dos primers assajos, les respostes són molt diferents entre si  $(26,0^{\circ})$  enfront de  $72,3^{\circ}$ ) i pot donar la sensació que hi ha un punt erroni. Per tant, es fa una tercera repetició. En aquesta es torna a obtenir una resposta baixa  $(20,6^{\circ})$  i pot donar la sensació que el segon punt  $(72,3^{\circ})$  és erroni. Per tant es fa una quarta repetició. En aquesta s'obté una resposta elevada  $(62,6^{\circ})$  i la distribució de respostes és dos a dos. Per tant, es fa una cinquena repetició. En aquesta s'obté una resposta intermèdia-alta  $(51,4^{\circ})$ .

Aquest raonament és molt detallista i es basa excessivament en el valor de la resposta de cada experiment. Els assajos s'interpreten independentment i es mostra una gran preocupació pel fet que pugui haver algun punt erroni. Una visió global aporta nova informació molt interessant. En la *Figura 181* es mostren les respostes individuals de cada experiment ( $x_i$ ), les respostes individuals de tots els assajos agrupades i la situació de la resposta experimental final ( $\overline{\alpha}_{exp}$ ) dins de la corba dosi-resposta. Per cada punt de la corba es representa la dispersió de les dades.



*Figura 181*: Resultats de la bl a  $10^{-5} \mu g$ /planta: Gràfics de les respostes individuals distribuïdes segons els bioassaig, agrupades i situació dins la corba dosi-resposta.

D'acord amb aquest nou punt de vista, la interpretació dels resultats és la següent: s'han realitzat cinc repeticions on s'han assajat un total de 51 plantes. Les respostes individuals van des dels 0° fins als 98°. Per comparació amb altres dosis s'observa que s'està en el pendent de la corba. Per tant, sembla raonable que petites variacions incontrolables entre experiments puguin fer que les respostes d'un assaig tendeixin a l'alça (cas de COMB-60 i COMB-65), a la baixa (cas de COMB-58 i COMB-62) o es distribueixin per tot el marge (cas de COMB-69).

Per aprofundir en la discussió és molt útil estudiar els resultats del control i de la homocastasterona (hcs) a 1 µg/planta. Tots dos s'han aplicat en tots els bioassajos i es disposa d'una gran quantitat de dades. A la *Figura 182* es mostren les respostes individuals dels 68 bioassajos,<sup>\*\*\*\*\*</sup> les mateixes respostes agrupades, i la seva distribució. En verd es mostren les respostes de la hcs a 1 µg/planta i en blau les del control.

- La hcs a 1 μg/planta segueix una distribució normal en la qual s'han assajat 680 plantes. Les respostes individuals abasten tot el marge de respostes possibles anant des de 10<sup>o</sup> fins a 180<sup>o</sup>. La resposta promig és 98,5<sup>o</sup> ±1,0 sent la dispersió de les dades de 26,5<sup>o</sup>.
- El control és més complex ja que no segueix una distribució normal. S'han assajat 737 plantes de les quals un 15% presenten una resposta igual a 0º. Ometent aquestes respostes, el control seguiria, amb 623 plantes assajades, una distribució assimètrica amb una resposta promig de 8,2 ±0,2 sent la dispersió de les dades de 3,9º.

<sup>\*\*\*\*\*</sup> Numèricament es tenen referenciats 73 experiments. No obstant, els experiments compresos entre COMB-26 i COMB-29, i l'experiment COMB-47 corresponen a un altre tipus de bioassaig, concretament al test d'inclinació de la làmina d'arròs amb planta tallada.<sup>90</sup>





*Figura 182*: Resultats del control (en blau) i la hcs a 1 µg/planta (en verd): Gràfics de les respostes individuals distribuïdes segons els bioassaig, agrupades i distribució.

En el cas de la hcs a 1 µg/planta es realitzen proves d'hipòtesi per comprovar amb un 95% de seguretat si la mitjana d'un experiment és igual a la mitjana de totes les dades. A la *Figura 183* es mostra les mitjanes de cada experiment amb el seu error estàndard, sobre mitjana de totes les dades (línea horitzontal granat). En vermell es marquen els punts on la mitjana de l'experiment no coincideix amb la mitjana global.



*Figura 183*: Resultats de la hcs a 1 µg/planta: Resposta mitjana i el seu error estàndard segons bioassaig situats sobre la mitjana global.

S'observa com en 8 dels 68 experiments, les dos mitjanes no són iguals. A cadascun d'ells es fa una anàlisi detallada de les respostes. A mode d'exemple es mostra l'anàlisi fet sobre COMB-34. En l'assaig es van estudiar 2 BRs a 3 dosis diferents: hbl a  $10^{-3}$ ,  $5 \cdot 10^{-4}$  i  $10^{-4}$  µg/planta i S hbl a  $10^{-1}$ ,  $10^{-2}$  i  $10^{-3}$  µg/planta. Es representa la situació d'aquests

punts dins de la corba dosi-resposta considerant les respostes mitjanes de cada experiment (*Figura 184*). Al contrari que passa amb la hcs a 1  $\mu$ g/planta les 6 respostes restants no són ni superiors ni qualitativament diferents en relació amb els altres punts de les respectives corbes.



Figura 184: Respostes del bioassaig COMB-34 (en vermell).

El mateix passa amb la resta d'experiments conflictius, excepte el COMB-61. En aquest experiment la resposta de la hcs a 1 µg/planta és la més alta dels 68 bioassajos (142º) i es van estudiar 2 BRs a diferents dosis: bl a  $5 \cdot 10^{-6}$  i cs a 1,  $10^{-1}$ ,  $10^{-2}$ ,  $10^{-3}$  i  $5 \cdot 10^{-5}$  µg/planta (*Figura 185*). Igual que passa amb la hcs a 1 µg/planta les 6 respostes d'aquest assaig són les més altes en la seva dosi i en alguns casos són molt diferents al valor habitual.



Figura 185: Respostes del bioassaig COMB-61 (en vermell) per la bl i la cs.

Per tant, en la majoria d'assajos conflictius, el fet que la hcs a 1 µg/planta doni una resposta mitjana diferent a la mitjana global es deu a la pròpia dispersió del creixement de les plantes (factors intrínsecs). Només en un assaig (COMB-61) la presència d'algun factor extrínsec ha afectat al conjunt del bioassaig fent que les respostes siguin més elevades del que és habitual. Aquest factor només ha afectat a aquest experiment ja que en els dos assajos veïns (COMB-60 i COMB-62) la resposta de la hcs a 1 µg/planta presenta els valors habituals (*Figura 183*).

# 4.2. Reducció de punts experimentals.

#### a) Influència en un bioassaig.

En tot experiment es suprimeixen els valors extrems (màxim i mínim) assumint que són els que més error poden tenir. No obstant s'ha vist que a nivell d'un experiment, suprimir els extrems no sempre és eficaç. A mode d'exemple, en la *Figura 186* es mostren els resultats de la bl a 0,001  $\mu$ g/planta en l'experiment COMB-58. S'observa com el primer

punt sembla estar dins de la tendència mentre que els dos últims punts s'allunyen bastant. Si es suprimeixen els extrems s'elimina un punt que no és erroni, i dels dos punts erronis només se n'elimina un. S'observa com la mitjana suprimint els dos últims punts és molt més precisa que la mitjana amb tots els punts i la mitjana suprimint els punts extrems.



*Figura 186*: Resultats de la bl a 0,001 µg/planta en l'experiment COMB-58.

#### b) Influència en múltiples bioassajos.

A nivell general s'estudia l'efecte sobre la hcs a 1 µg/planta. A continuació es mostren la distribucions considerant les 680 respostes individuals (*Figura 187a*) i les 544 respostes individuals havent suprimint de cada experiment els valors màxim i mínim (*Figura 187b*). Com era d'esperar, en suprimir els punts extrems de cada experiment la distribució es fa més estreta. Destacar, que no hi ha diferències significatives en el valor promig de les dos distribucions (de 98,5 ±1,0 es passa a 99,1 ±0,8).



*Figura 187*: Gràfics de la distribució de les respostes de la hcs a 1 µg/planta.

# c) Nova proposta de reducció de punts experimentals.

La reducció de dades basada en la supressió del màxim i mínim no s'ha mostrat eficaç ja que el marge de respostes continua sent molt ampli, des de 18º fins a 163º. Per evitar aquesta situació es pot fer servir un tractament en el qual es suprimeixin les respostes més allunyades de la mitjana. El marge de respostes queda establert mitjançant el límit de confiança de la mitjana.

En el cas dels BRs aquesta reducció es realitza considerant tots els punts independentment de l'experiment, suprimint els valors d'angle que quedin fora dels marges establerts pel límit de confiança de la mitjana amb un grau de seguretat del 95% (L.C. 95%), i repetint el procés fins que tots els valors estan dins del límit de confiança. A continuació es mostren la distribucions considerant respostes individuals (*Figura 188a*) i respostes un cop suprimits els valors que queden fora del límit de confiança (*Figura 188b*).

Aplicant aquesta reducció de dades, s'han assajat 488 plantes les respostes de les quals van des de 76° fins a 120°. Com era d'esperar, la distribució encara es fa més estreta. Destacar que un cop més no hi ha diferències significatives en el valor promig de les dos distribucions (de 98,5 ±1,0 es passa a 98,0 ±0,5).



Figura 188: Gràfics de la distribució de les respostes de la hcs a 1 µg/planta.

Segons el bioassaig, hi ha una gran diversitat de casos (Figura 189):

- Experiments on s'han suprimit els dos punts extrems (cas de COMB-73).
- Experiments on s'han suprimit un o més punts només d'un extrem (cas de COMB46 per al mínim i COMB-48 per al màxim)
- Experiments on s'ha suprimit diversos punts dels extrems (cas de COMB-01)
- Experiments on no s'ha suprimit cap punt (cas de COMB-09).

Aquest tractament pot ajudar a identificar experiments erronis. Hi ha 6 experiments on la reducció de punts és superior al 50%. Tres d'ells es degut a una elevada dispersió de les dades (COMB-01, COMB-18 i COMB-62) mentre que els altres tres es degut a que les dades tendeixen a l'alça (COMB-24, COMB-54 i COMB-61). Especialment significatiu és el cas de COMB-61 on 6 dels 7 punts han estat suprimits.

Quantitativament, la reducció de punts experimentals comporta una disminució de l'error tan de la mitjana global i com de la mitjana de cada experiment. Per tant, quan es realitzen les proves d'hipòtesi per comprovar amb un 95% de seguretat si la mitjana d'un experiment és igual a la mitjana de totes les dades s'obtenen més experiments conflictius (*Figura 190*). Només pel fet de suprimir els punts extrems de cada experiment es passa de 8 a 22 assajos conflictius. Suprimint els punts que estan fora del L.C. 95% es passa de 8 a 11 assajos conflictius. No obstant, en aquest cas els experiments COMB-25 i COMB-34 deixen de ser conflictius.



*Figura 189*: Gràfics de les respostes individuals distribuïdes segons els bioassaig. En vermell es marquen els punts suprimits.



*Figura 190*: Resposta mitjana amb el seu error estàndard de la hcs a 1  $\mu$ g/planta segons bioassaig situats sobre la mitjana global.

Com s'ha realitzat anteriorment, s'analitzen els experiments conflictius representant les respostes mitjanes dels altres BRs estudiats dins de la seva corba dosi-resposta. Només a l'experiment COMB-61 els valors de tots els punts són superiors als habituals. En la resta els valors no són pas ni superiors ni qualitativament diferents en relació amb els altres punts de la corba.

#### d) Influència de la reducció de punts experimentals en la corba dosi-resposta.

Per estudiar l'efecte de la reducció de punts experimentals sobre la corba i el valor d'activitat es superposen les corbes dosi-resposta. A mode d'exemple, a la *Figura 191* es mostra la superposició en el cas de la bl.



*Figura 191*: Superposició de les corbes dosi-resposta considerant totes les dades (en verd), suprimint els punts extrems (en granat) i suprimint els punts que es surten del L.C. 95% (en blau).

Suprimir els punts extrems (màxim i mínim) de cada bioassaig no comporta una variació significativa de la corba. En conseqüència no hi ha diferències en el valor d'activitat (de  $3,84 \pm 0,09$  es passa a  $3,86 \pm 0,08$ ).

En canvi suprimir el punts que queden fora del L.C. 95% si que modifica significativament la corba. En concret el pendent de la zona central és més acusat. Això és degut a que la supressió de punts afecta de forma diferent segons la zona de la corba (*Figura 192*):

- En el màxim es suprimeixen punts que estan tan per sobre com per sota de la mitjana amb la qual cosa gairebé no es modifica el seu valor (*Figura 192a*).
- En la zona de transició entre el màxim i la zona central, la reducció de punts afecta principalment als punts amb valors baixos amb la qual cosa el valor de la mitjana augmenta significativament (*Figura 192b*).
- En la zona central, la dispersió de les dades és molt gran i no es suprimeix cap punt (*Figura 192c*).
- En la zona de transició entre el mínim i la zona central, la reducció de punts afecta principalment als punts amb valors elevats amb la qual cosa el valor de la mitjana disminueix significativament (*Figura 192d*).
- En el mínim, es suprimeix algun valor elevat que redueix lleugerament el valor de la mitjana (*Figura 192e*).

Igual que en el cas anterior, suprimir els punts que queden fora del L.C. 95% no modifica significativament el valor d'activitat (de 3,84  $\pm$ 0,09 es passa a 3,95  $\pm$ 0,06). No obstant, l'error disminueix com a conseqüència que el pendent en la zona central és més acusat.



*Figura 192*: Efecte de la supressió dels punts que es surten del L.C. 95% (en vermell) en diferents parts de la corba dosi-resposta.

Resumint, la reducció de punts experimentals comporta un augment del nombre d'experiments susceptibles de ser considerats erronis i no afecta al valor d'activitat. Per tant, es pot afirmar que aquest tractament no és necessari.

#### 4.3. Rectificació de les dades.

Sovint es tendeix a oblidar que les dades estan rectificades i s'associa la resposta final a un angle real. Això crea certa confusió quan per efecte de restar el control la resposta és negativa, ja que la tendència és a pensar en un angle negatiu el qual traslladat a la planta és fisiològicament impossible. Per estudiar l'efecte de la rectificació es superposen les corbes dosi-resposta amb les dades rectificades i sense rectificar.

A mode d'exemple, la *Figura 193* mostra el resultat de la superposició en el cas de la bl. S'observa que rectificar les dades comporta únicament un desplaçament cap avall de la corba. El valor d'activitat, expressat com  $-\log(\text{dosi}_{50\%})$ , es manté inalterat (de 3,84 ±0,09 es passa a 3,85 ±0,08).

Tradicionalment, l'equip ha fet servir com a valor d'activitat el  $-\log(\text{dosi}_{45^\circ})$ . Aquest valor porta implícit la rectificació de les dades. Si no es rectifiquen, l'equivalent a 45° són 55°. Entre els valors d'activitat expressats com a  $-\log(\text{dosi}_{45^\circ})$  o com a  $-\log(\text{dosi}_{55^\circ})$  tampoc hi ha diferències significatives (de 3,91 ±0,12 es passa a 4,02 ±0,13).

Resumint, la rectificació de les dades comporta un desplaçament cap avall de la corba que no afecta al valor d'activitat. Per tant, es pot afirmar que aquest tractament tampoc és necessari.



*Figura 193*: Superposició de les corbes dosi-activitat rectificant les dades (en vermell) i sense rectificar (en verd).

#### 4.4. Resposta experimental.

#### a) Mitjana de totes les respostes individuals vs. mitjana dels experiments.

Tradicionalment, la resposta experimental resultant ( $\overline{\alpha}_{exp}$ ) s'obté com la mitjana de les respostes de cada experiment ( $\overline{\alpha}_i$ ). No obstant, s'ha demostrat la gran utilitat de considerar totes les respostes individuals independentment del bioassaig que provenen. Per tant, una altre manera de calcular la resposta experimental és a través de la mitjana de totes les respostes individuals independentment del bioassaig.

A nivell general s'estudia l'efecte sobre la hcs a 1 µg/planta. A continuació es mostren la distribucions de les respostes de la hcs a 1 µg/planta considerant les 680 respostes individuals independentment de l'experiment (*Figura 194a*) i les 68 respostes mitjanes de cada experiment (*Figura 194b*). Un cop més, no hi ha diferències significatives en el valor promig de les dos distribucions (de 98,5 ±1,0 es passa a 98,9 ±1,5). No obstant, l'error augmenta en considerar les respostes mitjanes de cada experiment (de ±1,0 es passa a ±1,5).



Figura 194: Gràfics de la distribució de les respostes de la hcs a 1 µg/planta.

Per estudiar l'efecte del càlcul de la resposta experimental sobre la corba i el valor d'activitat es superposen les corbes dosi-resposta. A mode d'exemple, a la *Figura 195* es mostra la superposició en el cas de la bl. El càlcul de la resposta experimental no comporta una variació de la corba i no hi ha diferències significatives en el valor d'activitat (de 3,84  $\pm$ 0,09 es passa a 3,97  $\pm$ 0,13). No obstant, l'error torna a augmentar en considerar les respostes mitjanes de cada experiment (de  $\pm$ 0,09 es passa a  $\pm$ 0,13).



*Figura 195*: Superposició de les corbes dosi-resposta considerant els punts independents de l'assaig (en verd) i dependents de l'assaig (en vermell).

#### b) Mitjana dels experiments vs. experiments indviduals.

A llarg d'aquest capítol, també s'ha demostrat la utilitat de considerar el conjunt de respostes de cadascun dels experiments individualment, i no només la resposta mitjana dels experiments.

A mode d'exemple, a la *Figura 196* es mostra la superposició en el cas de la homobrassinolida (hbl). S'observen diferències entre les dos corbes, però aquestes només afecten als extrems i de forma simètrica (el màxim és 9<sup>o</sup> superior i el mínim 9<sup>o</sup> inferior). Aquesta diferència s'atribueix a un desequilibri en el nombre de repeticions dels punts segons la zona de la corba. Mentre en el màxim i el mínim els punts estan molt espaiats i s'han fet una o dos repeticions, en la zona central els punts estan poc espaiats i s'han realitzat entre 4 i 5 repeticions. Un cop més no hi ha cap diferència significativa en el valor d'activitat (de 3,29 ±0,17 es passa a 3,26 ±0,34). No obstant, l'error augmenta degut al desequilibri en el nombre de repeticions.



*Figura 196*: Superposició de les corbes dosi-resposta considerant la mitjana dels experiments (en granat) i els experiments individuals (en verd). S'ha considerat un experiment erroni (en vermell).

Resumint, la forma de calcular la resposta experimental no afecta al valor d'activitat.

#### 4.5. Factors extrínsecs. Corbes dosi-resposta de la brassinolida i la castasterona.

En el tema de relacions estructura-activitat s'ha comentat que durant la tesi l'equip ha revisat les activitats de la brassinolida i la castasterona. Les corbes dosi-resposta inicials corresponen al 1998,<sup>90</sup> mentre la revisió de la brassinolida (bl) correspon al 2002,<sup>205</sup> i la de la castasterona (cs) al 2004.<sup>207</sup>

Les primeres divergències en els resultats experimentals de la brassinolida es van detectar al 2000 mentre es posava a punt una modificació del bioassaig per estudiar tan l'efecte sinergètic com l'efecte inhibidor.

- L'efecte sinergètic dels BRs amb les auxines està àmpliament descrit en la bibliografia.<sup>54,71,208,209</sup> Aprofitant aquesta sinergia es va estudiar la possibilitat d'utilitzar-la per tal d'augmentar la sensibilitat del bioassaig. Un test sensibilitzat permetria determinar si els BRs amb una resposta baixa a 1 μg/planta són realment inactius o la seva resposta es veu limitada per la impossibilitat d'aplicar dosis superiors.
- L'efecte inhibidor dels BRs amb diversos compostos ha estat descrit, però no s'ha estudiat sistemàticament. Un test que pogués estudiar aquest efecte, permetria determinar si els anàlegs BRs amb resposta nul·la a 1 µg/planta són inactius "per se" o són inactius perquè presenten un efecte inhibidor.

Un dels punts d'aquests treballs va consistir en determinar la dosi idònia a la qual la brassinolida devia ser coaplicada, ja fóra amb una auxina o amb un potencial inhibidor. Es van assajar les concentracions de  $5 \cdot 10^{-1}$ ,  $10^{-3}$ ,  $10^{-4}$  i  $10^{-5}$  µg/planta (*Taula 20*). Les respostes obtingudes van resultar ser inferiors a les del 1998. De fet, es va observar una dràstica reducció a  $10^{-4}$  i a  $10^{-5}$  µg/planta. Com l'objectiu era l'estudi de l'efecte sinergètic i l'efecte inhibidor i no pas la resposta de la brassinolida, no va ser fins al 2002 que no es va realitzar novament la corba dosi-resposta.

Concentració	0,5 µg/planta	10⁻³ µg/planta	10 <sup>−4</sup> µg/planta	10 <sup>-5</sup> µg/planta
bl – 1998	-	94º	68 <u>°</u>	47º
bl – 2000	95º	79º	10º	7⁰

*Taula* **20**: Respostes de la bl a diverses concentracions en els estudis de sinergisme i inhibició del 2000. Comparació amb els resultats del 1998.

En la revisió de la corba de la brassinolida (*Figura 197a*) s'observa:

- La corba del 1998 presenta una amplada menor que la del 2002. La diferència entre el màxim i el mínim és de 87º al 1998 i de 105º al 2002.
- Les dos corbes comencen a decaure simultàniament entre els punts -2 i -3. No obstant, la corba del 1998 decau més suaument que la del 2002.
- En la corba del 1998 l'activitat expressada com a -log(dosi<sub>50%</sub>), és de 4,73. En la corba del 2002 és de 3,84. Això suposa una diferència de 0,98 unitats de logaritme.

Un cop revisada l'activitat de la brassinolida, resulta que l'activitat de la castasterona  $(-\log(dosi_{50\%})=4,01)$  és superior a la de la brassinolida  $(-\log(dosi_{50\%})=3,84)$ . Aquest fet és contradictori amb la realitat dels BRs naturals. La castasterona és el precursor biogenètic de la brassinolida i, per tant, és menys activa. El 2004 es va decidir revisar la corba de la castasterona.

En la revisió de la corba de la castasterona (*Figura 197b*) s'observa:

- La corba del 1998 presenta una amplada menor que la del 2002. La diferència entre el màxim i el mínim és de 93º al 1998 i de 100º al 2004.
- Les dos corbes comencen a decaure simultàniament entre els punts -1 i -2. No obstant, la corba del 1998 decau més suaument que la del 2004.

• En la corba del 1998 l'activitat expressada com a -log(dosi<sub>50%</sub>), és de 4,01. En la corba del 2002 és de 3,27. Això suposa una diferència de 0,74 unitats de logaritme.



*Figura 197*: Superposició de les corbes dosi-resposta de: *(a)* la brassinolida i *(b)* la castasterona. Les corbes del 1998 es representen en granat i les revisions del 2002 i el 2004 en verd.

Donada l'especial incidència que té la variació del valor d'activitat en els estudis de QSAR, es decideix revisar exhaustivament cadascun dels experiments per tal de trobar una explicació.

El fet més destacable de les corbes de la brassinolida i la castasterona del 1998 és que les dos decauen suaument. El seus pendents (*K*) són de 1,0 i 1,2 respectivament. Les corbes de la resta de BRs naturals decauen de forma més abrupte. Els pendents són superiors a 1,8, sent 2,1 el valor més freqüent.

Si s'afirma que els pendents de les corbes del 1998 són inusualment suaus, les activitats haurien de ser inusualment elevades.

L'equip sempre s'ha mostrat sorprès de la discontinuïtat en els salts d'activitats que es produeixen en variar el reste alquílic en C24 i la seva configuració.<sup>90,97</sup> Quan intervenen la brassinolida i la castasterona, els salts d'activitat són molt grans en comparació als salts on no intervenen (*Figura 198*):

- El salt entre la cadena tipus *homo* (reste 24*S*-etil) i la cadena tipus *bl* (reste 24*S*-metil) és més del triple que el salt entre la cadena tipus *epi* (reste 24*R*-metil) i la cadena tipus *homo* (reste 24*S*-etil).
- La diferència entre lactona i cetona augmenta progressivament en passar de la cadena tipus *epi* (reste 24*R*-metil) a la cadena tipus *homo* (reste 24*S*-etil) i a la cadena tipus *bl* (reste 24*S*-metil). No obstant, l'augment en passar de la cadena tipus *homo* a la cadena tipus *bl* és gairebé el doble que per passar de la cadena tipus *epi* a la cadena tipus *homo*.



*Figura 198*: Salts d'activitat del 1998 que es produeixen en variar el reste alquílic en C24 i la seva configuració i la seva relació entre lactones i cetones.
El fet que les diferències estructurals siguin escasses (un 5% de la molècula), combinat amb les discontinuïtats dels salts d'activitat van fer que els primers models QSAR no fossin capaços d'explicar la brassinolida. Posteriorment, en refinar els models es va ser capaç d'explicar tan la brassinolida com la castasterona. No obstant, la correlació tendia a estar molt influenciada per aquests dos compostos.<sup>95,210</sup>

Per tant, tan la discontinuïtat en els salts d'activitat com la dificultat d'establir una correlació estructura-activitat evidencien que les activitats del 1998 de la brassinolida i la castasterona són inusualment elevades.

A les revisions del 2002/2004, els pendents de la brassinolida i la castasterona adquireixen els valors habituals, sent de 2,1 i 1,9 respectivament. Sempre i quan les corbes del 1998 i del 2002/2004 comencin a decaure simultàniament, com és el cas (*Figura 197*), un pendent major comporta una activitat menor. La variació experimental del valor de  $X_{50\%}$  associada al canvi del pendent, coincideix força amb la variació teòrica. Per la brassinolida ( $K_{1998}$ =1,0,  $K_{2002}$ =2,1) la variació teòrica de  $X_{50\%}$  és de 1,05 unitats i experimentalment és de 0,98. Per la castasterona ( $K_{1998}$ =1,2,  $K_{2002}$ =1,9), la variació teòrica és de 0,80 unitats i experimentalment és de 0,74.

Els salts d'activitat entre cadenes laterals tipus són del mateix ordre de magnitud (*Figura* 199). L'augment en la diferència d'activitats entre lactona i cetona és progressiu i es deu a que el salt d'activitat en passar d'una cadena lateral a l'altre és 0,10 unitats superior en les lactones que en les cetones.



*Figura 199*: Salts d'activitat del 2002/2004 que es produeixen en variar el reste alquílic en C24 i la seva configuració i la seva relació entre lactones i cetones.

Per tant, els valors d'activitat de la brassinolida del 2002 i la castasterona del 2004 estan en consonància amb els valors d'activitat de la resta de BRs i són els que, actualment, l'equip accepta.

Per entendre el pendent inusualment suau de la brassinolida i la castasterona del 1998 es visualitzen les corbes dosi-resposta representant els resultats de cada experiment enlloc de la seva mitjana. A la *Figura 200* es mostren el resultat per la brassinolida del 1998 i el 2002. Els experiments que es van considerar erronis s'indiquen en un color més suau.

S'observa com en la part intermèdia-baixa de la corba del 1998 (*Figura 200a*, punts inferiors a –4) hi ha una gran dispersió dels resultats que contrasta amb les de la resta de la corba. Per exemple, a  $5 \cdot 10^{-6} \mu g/p$ lanta les respostes van des de 14º fins a 85º ( $\Delta$ =71º) sent el valor promig de 44º. En canvi, en la corba del 2002, les respostes en punts inferiors a –4 (*Figura 200b*) són menors que les de 1998, però la seva dispersió es bastant homogènia amb la de la resta de la corba. Posant com a exemple la mateixa concentració, a  $5 \cdot 10^{-6} \mu g/p$ lanta les respostes van des de 15º fins a 49º ( $\Delta$ =34º) sent el valor promig de 27º.



*Figura 200*: Corbes dosi-resposta de la brassinolida *(a)* del 1998 i *(b)* del 2002 sobre les respostes individuals de cada experiment, indicant en un color més suau les que es van considerar errònies.

En la castasterona (*Figura 201*), la comparació de la dispersió de respostes al llarg de la corba del 1998 es veu molt afectada per l'assaig COMB-61 (*Figura 185*, pàgina 205) i pel fet que es van fer menys repeticions de cada punt. No obstant, s'observa una tendència similar a de la brassinolida. Per exemple, a  $5 \cdot 10^{-5} \mu g/planta$  les respostes del 1998 van des de 36º fins a 68º ( $\Delta$ =32º) sent el valor promig de 52º, mentre que al 2004 van des de 10º a 17º ( $\Delta$ =7º) sent el valor promig de 14º.



*Figura 201*: Corbes dosi-resposta de la castasterona *(a)* del 1998 i *(b)* del 2004 sobre les respostes individuals de cada experiment, indicant en un color més suau les que es van considerar errònies.

Resumint, a la zona intermèdia-baixa de les corbes del 1998 de la brassinolida i la castasterona s'observa una dispersió inusual de les dades, que provoca que la respostes promig siguin inusualment elevades, que provoca que el pendent sigui inusualment suau, que provoca que l'activitat sigui inusualment elevada.

A continuació es fa un repàs dels diferents factors extrínsecs que poden haver influït de cara a fer que les dispersions de les respostes siguin diferents segons l'any i en conseqüència el valor promig, el pendent i l'activitat.

### • Origen dels compostos.

Exceptuant la brassinolida i la castasterona, la resta de BRs han estat sintetitzats a l'equip.<sup>80,81,83,98,99,106,211</sup> Ambdós han estat adquirits a CIDtech Reseach Inc. (1200 Franklin Blvd., Cambridge, Ontario, Canada) l'any corresponent. Per tant, la bl del 1998 i la del 2002 pertanyen a diferents stocks, i el mateix passa amb la cs del 1998 i la del 2004. A més a més, la seva puresa no ha pogut ser avaluada donada la petita quantitat adquirida per raons econòmiques. Al 1999, 10 mg de brassinolida van costar 275\$.

Per avaluar la influència de la presència de impureses sobre la corba dosi-resposta, es consideren dos casos segons si la impuresa afecta o no a la planta.

Si la impuresa no té cap efecte en la planta, la disminució de la puresa equival a una disminució de la dosi aplicada en el bioassaig, la corba en tot el seu conjunt es desplaça cap dosis inferiors i el valor d'activitat disminueix. Fixant-se només en el valor d'activitat, podria pensar-se que els compostos del 2002/2004 presenten alguna impuresa mentre els del 1998 són purs. Aquesta afirmació presenta tres contradiccions. La primera és que les corbes del 2002/2004 són més coherents amb les de la resta de BRs naturals que no pas les del 1998. La segona és que, per exemple, si en un cas extrem la puresa fóra del 50%, la reducció de l'activitat només seria de 0,3 unitats. En canvi, la reducció observada per la brassinolida és de 0,89 unitats i per la castasterona de 0,74 unitats. La tercera és que les corbes del 1998 i del 2002/2004 haurien de ser paral·leles i no ho son.

Si la impuresa afecta a la planta, és lògic pensar que els seu efecte serà menor que el de la brassinolida i la castasterona, ja que aquests són els dos compostos més actius. Podria donar-se el cas que la brassinolida estigues impurificada amb castasterona, ja que és el seu precursor sintètic. Aquest seria el cas on l'efecte seria màxim. Estudiant com varien les corbes en funció de la proporció dels dos compostos, s'observa que si la proporció de brassinolida és superior al 50%, l'activitat resultant no varia significativament, ja que la reducció de l'activitat associada a la disminució de la concentració es veu compensada per l'efecte additiu de la castasterona. Quan la proporció de brassinolida és inferior al 50% l'activitat comença a disminuir apreciablement i a mesura que la seva proporció disminueix, tan la corba com l'activitat resultants s'acosten a la de la castasterona. Per tant, havent vist el pitjor dels casos, es pot afirmar que una impuresa hauria de tenir una activitat considerable i estar present en quantitats elevades per observar la seva influència. Es a dir, la impuresa hauria de deixar de ser una impuresa. De no ser així, la corba del BR emmascararà la resposta de la impuresa.

Per tant, d'acord amb el que s'ha explicat, ni l'origen dels compostos ni la seva puresa no justifiquen per si sols ni la diferència d'activitats entre els anys, ni la suposada sobreestimació de les respostes en la part inferior de la corbes del 1998.

### • Origen i/o envelliment de les llavors.

Per evitar l'envelliment de les llavors d'arròs, cada any s'adquireixen noves llavors de la varietat *Bahia* a la Cambra Arrossera del Delta de l'Ebre tot just desprès de la collita. La qualitat de l'arròs pot variar d'un any a l'altre, ja que depèn de factors tan diversos com la climatologia, la qualitat de l'aigua del riu, l'ús de pesticides, etcètera.

Podria pensar-se que les llavors del 1998 han estat més sensibles a l'acció dels BRs que les llavors del 2002 i el 2004. Aquesta major sensibilitat seria més acusada quan menor fóra la concentració de BR i podria explicar la suposada sobreestimació de les respostes en la part inferior de la corbes del 1998. No obstant, aquest efecte s'hauria hagut de detectar en les respostes intermèdies-baixa d'altres BRs estudiats el 1998 donant pendents suaus, i no ha estat així, afectant només a la brassinolida i la castasterona.<sup>90</sup> A més a més, si aquest

efecte fóra realment important, també s'hauria hagut de detectar en la variació de les respostes d'un bioassaig a l'altre en fer el canvi de les llavors i no ha estat així.<sup>212</sup>

Si be, l'origen de les llavors podria explicar la suposada sobreestimació de les respostes en la part inferior de les corbes del 1998, sembla incoherent que només s'hagués observat en la brassinolida i la castasterona i que l'efecte hagués estat tan important com per provocar una variació de l'activitat de 0,98 i 0,74 unitats de logaritme, respectivament.

### • Aplicació de la dosi correcte.

Els estudis per posar a punt el protocol del bioassaig van demostrar que l'efecte dels errors en la pesada, dilució, pressa de mostra, injecció i/o variació de la concentració per evaporació del dissolvent afecten poc a la resposta experimental.<sup>90</sup>

### • Experimentador.

Cadascun dels estudis l'ha realitzat un experimentador diferent. No obstant, des de que un nou experimentador s'inicia en el protocol del bioassaig fins que els seus resultats començen a considerar-se vàlids, hi ha un procés d'aprenentatge que dura entre un i dos mesos. Exceptuant casos puntuals, en els treballs realitzats en paral·lel per diversos experimentadors no s'han detectat variacions significatives de les respostes.<sup>212</sup>

### • Factors indeterminats.

Per intentar veure si hi ha algun factor indeterminat que pugui explicar la suposada sobreestimació de les respostes en la part inferior de la corbes del 1998, s'estudien els bioassajos en que es van realitzar les corbes de la brassinolida i la castasterona.

Primer s'estudia si hi ha alguna tendència en la seqüència de les respostes a una dosi concreta. Per exemple, la seqüència de les respostes de la brassinolida a  $10^{-5}$  µg/planta consta de cinc repeticions (pàgines 202–203), obtenint-se consecutivament una resposta baixa (26°), una elevada (72°), una altre baixa (21°), una altre elevada (63°) i finalment una intermèdia-alta (51°). Igual que en aquest exemple, a la resta de dosis no s'observa cap tendència en la seqüència de les respostes.

Després s'estudia les respostes segons el bioassaig. S'ha detectat un assaig erroni (COMB-61), un assaig on les respostes de la zona intermèdia baixa tendeixen a estar sobreestimades (COMB-60), i unes poques respostes aïllades que dins del bioassaig no segueixen la tendència sigmoïdal. La supressió d'aquestes respostes explicaria força bé la sobreestimació de les respostes de la castasterona (*Figura 202b*). No obstant, per la brassinolida encara queda dispersió per explicar (*Figura 202a*).



*Figura 202*: Revisió de les corbes de *(a)* la brassinolida i *(b)* la castasterona indicant en vermell els punts que es consideren clarament erronis i en groc els punts de l'assaig COMB-60. La corba original del 1998 es marca en gris i la de la revisió en verd.

### • <u>Conclusions.</u>

El valor d'activitat del 2002 de la brassinolida i el del 2004 de la castasterona són els que s'accepten com valors correctes, ja que tan les corbes com els propis valors d'activitat són més coherents amb els de la resta de BRs.

L'activitat inusualment elevada del 1998 es deu a que el pendent de les corbes és inusualment suau ja que les respostes promig en la zona intermèdia-baixa són inusualment elevades alhora que presenten una dispersió molt gran.

No s'ha trobat cap factor que per si sol justifiqui l'elevada dispersió de les respostes de la brassinolida i la castasterona en la zona intermèdia-baixa.

### 5. Discussió de resultats.

### 5.1. Error experimental i QSAR.

Recordar que en el models QSAR els errors es donen mitjançant la desviació estàndard de l'error de càlcul (SDEC) i la desviació estàndard de l'error de predicció (SDEP). De forma anàloga es pot calcular la desviació estàndard de l'error experimental (SDEE, equació 21). En cap cas ni l'error de càlcul (SDEC) ni l'error de predicció (SDEP) poden ser inferiors a l'error experimental (SDEE). De ser-ho implicaria que el model és capaç de calcular i/o predir l'error experimental. En aquests casos es diu que el model ha estat sobreajustat (overfit).

equació 21: 
$$SDEE = \sqrt{\frac{\sum_{i}^{j} error_{i}^{2}}{n_{j}}}$$

### 5.2. Error experimental de la resposta a 1 µg/planta.

La *Taula 21* mostra la resposta i l'error experimental a  $1 \mu g/planta$  dels BRs que inicialment formen part del model a  $1 \mu g/planta$ . Els BRs estan organitzats en ordre decreixent segons la resposta. Aquesta s'ha calculat sense fer cap reducció de les dades, sense rectificar-les i calculant la mitjana de totes les respostes individuals.

BRs	1 μg/planta*	BRs	1 µg/planta		BRs	1 μg/planta*
bl	117,5 ±2,8	S_epibl	$80,5 \pm 5,0$		hcs_2b3b_5OH	39,6 ±4,1
hbl_cis	107,3 ±4,6	epics	80,3 ±8,8		S_epics_2n3b_5OH	$34,0\pm 3,7$
CS	106,1 ±2,1	hcs_60Hb	$79,9 \pm 3,7$		hcs_2n3Br_5OH	$29,6 \pm 4,7$
hcs	101,0 ±3,8	hcs_2b3b	78,5 ±1,8		S_epics_2n3b_11OH	29,3 ±6,0
epics_2n3b	101 ±N.D.	epics_2n3b_11OH	$77,2\pm 3,5$		epics_2n3b_23n	24 ±N.D.
hbl	99,0 ±4,6	hcs_2n3b	$75,6\pm 5,2$		S_hcs_2CO3n	$23,0\pm 3,2$
epibl	95,2 ±4,0	hcs_2n3n	$75,5 \pm 4,7$		S_hcs_22n	22 ±N.D.
hcs_2CO3n	93,2 ±2,6	epics_2n3n	$74,2 \pm 5,2$		S_epics_11OH	$20,4 \pm 5,0$
S_hcs_23n	93 ±N.D.	epibl_11OH	$72,5 \pm 4,9$		S_hcs_2n3Br	$15,9 \pm 3,2$
hcs_2n3a	$90,0\pm 3,4$	hcs_eter	70,1 ±4,6		S_epics_2b3b	10,3 ±2,7
epics_11OH	87,3 ±4,2	epics_2b3b	$67,6 \pm 5,7$		S_hcs_2n3b_5OH	9,9 ±1,4
S_hbl	85,8 ±5,2	S_epibl_11OH	66,1 ±4,4		S_hcs_2n3Br_5OH	8,8 ±1,5
hcs_2a3n	85,0 ±4,2	hcs_60Ha	61,1 ±7,0		S_epics_2n3n	8,3 ±1,7
hcs_cis	84,6 ±5,9	S_hcs_2a3n	$59,9 \pm 4,8$		epics_22n23n	7,8 ±1,8
S_hcs	82,0 ±3,2	S_hcs_5OH	$52,3 \pm 6,7$		hbl_2n3n_22n23n	$7,3\pm 0,9$
hcs_5OH	81,5 ±5,5	hcs_2n3n_22n23n	46,3 ±5,3		epics_2n3b_22n	7 ±N.D.
hcs_22n23n	81,4 ±5,7	hcs_2n3Br	45,8 ±6,0		S_hcs_eter	$5,0 \pm 1,5$
S enics	81.1 + 5.3	hbl 2h3h	$429 \pm 68$	]		

\* N.D. Dades no disponibles.

Taula 21: Resposta a 1 µg/planta i error experimental dels BRs presents en el model a 1 µg/planta.

L'error experimental (SDEE) de les respostes a 1  $\mu$ g/planta és de  $\pm$ 4,5º.

Els errors de càlcul (SDEC) i de predicció (SDEP) del model a 1  $\mu$ g/planta són de ±7,3° i ±9,6°, respectivament. Tots dos errors són superiors a l'error experimental, el que indica que el model no està sobreajustat. No obstant, la diferència és petita, la qual cosa inclina a pensar que el model està força ben ajustat i no té gaire marge de millora.

### 5.3. Error experimental de l'activitat (-log(dosi<sub>50%</sub>)).

La *Taula 22* mostra l'activitat expressada com –log(dosi<sub>50%</sub>) amb el seu error experimental dels BRs que formen part del model QSAR. Els BRs estan organitzats segons la cadena lateral tipus. L'activitat s'ha calculat sense fer cap reducció de les dades, sense rectificar-les i calculant la mitjana de totes les respostes individuals.

L'error experimental (SDEE) de l'activitat expressada com  $-\log(dosi_{50\%})$  és de  $\pm 0,13$ , i com a  $-\log(dosi_{55^{\circ}})$  de  $\pm 0,21$ .

El model *HIP* utilitza l'activitat expressada com  $-\log(dosi_{50\%})$ . Els errors de càlcul (SDEC) i de predicció (SDEP) són de  $\pm 0,24$  i  $\pm 0,52$ , respectivament. Tots dos errors són superiors a l'error experimental, el que indica que el model no està sobreajustat. No obstant, la diferència és suficientment gran com per afirmar que el model encara té estadísticament marge per millorar, sobretot en la predicció la qual presenta un error quatre cops superior a l'error experimental.

BRs	-log(dosi <sub>50%</sub> )	-log(dosi <sub>55</sub> )	BRs	-log(dosi <sub>50%</sub> )	-log(dosi <sub>55</sub> )
bl	$3,84 \pm 0,09$	$3,91 \pm 0,12$	hbl	$3,27 \pm 0,19$	$3,30 \pm 0,26$
CS	$3,27 \pm 0,09$	3,27 ±0,19	hbl_cis	$2,37 \pm 0,10$	$2,40 \pm 0,13$
epibl	$2,69 \pm 0,16$	$2,56 \pm 0,24$	hcs	$2,78 \pm 0,09$	$2,75 \pm 0,14$
epics	$2,28 \pm 0,04$	$2,19 \pm 0,10$	hcs_cis	$1,52 \pm 0,10$	1,41 ±0,14
epics_2b3b	$0,36 \pm 0,14$	0,31 ±0,18	hcs_2n3a	$2,06 \pm 0,10$	$1,86 \pm 0,17$
epics_2n3n	$0,51 \pm 0,06$	$0,47 \pm 0,13$	hcs_2n3b	$1,69 \pm 0,16$	$1,34 \pm 0,30$
S_hbl	$2,05 \pm 0,03$	$2,04 \pm 0,04$	hcs_60Ha	$1,74 \pm 0,22$	$1,43 \pm 0,38$
S_hcs	1,21 ±0,06	1,08 ±0,11	hcs_60Hb	$1,69 \pm 0,11$	1,55 ±0,16
S_epibl	2,15 ±0,05	$2,08 \pm 0,08$	hcs_50H	1,75 ±0,08	$1,56 \pm 0,13$
S epics	1,01 ±0,31	$0,84 \pm 0,50$	hcs eter	0,66 ±0,10	0,48 ±0,16

Taula 22: Activitat i error experimental dels BRs presents en els models QSAR.

### 6. Conclusions.

S'ha comprovat que els diversos tractaments estadístics realitzats per calcular l'activitat (reducció dels punts experimentals, rectificació de les dades i el càlcul de la resposta experimental) no afecten significativament al seu valor i, per tant, no són necessaris. Es a dir, s'està aplicant un tractament estadístic massa potent per al càlcul de l'activitat.

S'ha determinat l'error experimental de la *resposta* a 1  $\mu$ g/planta i de l'*activitat* expressada com a  $-\log(\text{dosi}_{50\%})$  i com a  $-\log(\text{dosi}_{55^2})$ . Comparant-los amb els errors dels models, es pot afirmar que el model a 1  $\mu$ g/planta està força ben ajustat i no té gaire marge de millora, i que el model QSAR pot millorar considerablement, sobretot en la predictibilitat.

Relacionant el possible marge de millora del model QSAR amb les conclusions del capítol 3, es pot afirmar que amb el conjunt de BRs actual aquesta millora no sigui possible. Això és degut a les interaccions estructurals presents que fan que la predictibilitat variï en funció de l'estructura. Per tant, es torna a fer evident que per millorar el model cal augmentar racionalment el nombre de compostos de forma que el conjunt de BRs del model estigui estructuralment equilibrat.

# **CONCLUSIONS**

# Capítol 1: Androstans.

- Els anàlegs androstànics presenten una elevada flexibilitat de la cadena lateral que dificulta l'estudi computacional.
- És difícil assignar una conformació activa *in silico* als anàlegs BRs androstànics amb el grup funcional protegit. Hi ha diverses conformacions possibles, les quals presenten uns paràmetres numèrics molt similars. Cada conformació té una forma diferent d'explicar la manera d'actuar dels androstans sent coherent en alguns aspectes i incoherent en altres.

### Capítol 2: Conformació activa dels BRs.

- La raó fonamental per a la qual s'han trobat tres de les quatre conformacions actives enlloc d'una sola es troba únicament en l'anàlisi conformacional dels BRs i no pas en els processos de selecció de la conformació activa. Ni l'anàlisi sistemàtic en forma d'arbre ni l'anàlisi aleatori amb SYBYL han estat exhaustius, la qual cosa ha provocat que la posterior selecció no parteixi d'una base de confórmers sòlida.
- La conformació activa in silico és la conformació activa anomenada HIP.
- Aquesta conformació explica tridimensionalment la complementarietat geomètrica i la complementarietat electrostàtica dels 225,235-BRs amb el receptor.
- Els 22*R*,23*R*-BRs s'uneixen amb el receptor a través de la conformació més estable, mentre que els 22*S*,23*S*-BRs ho fan a través d'una conformació d'elevada energia. Per això els 22*R*,23*R*-BRs són més eficients i en conseqüència més actius que els 22*S*,23*S*-BRs.

# Capítol 3: QSAR.

- A través dels models, s'han detectat cinc interaccions estructurals en el conjunt de BRs que formen el *"data set"* que els afecten en major o menor grau. Aquestes interaccions són:
  - 1. Totes les lactones presenten el diol  $2\alpha$ ,  $3\alpha$ .
  - 2. Tots els 22*S*,23*S*-BRs presenten el diol  $2\alpha$ , $3\alpha$ .
  - 3. El 50% dels BRs amb cadena lateral tipus *epi* representen el 66% de les cetones sense cap funcionalitat en  $\alpha$  en l'anell A.
  - 4. Tots els BRs amb modificacions estructurals en l'anell B presenten la cadena lateral tipus *homo*.
  - 5. Tots els BRs sense carbonil en l'anell B presenten el diol  $2\alpha$ ,  $3\alpha$ .
- La interacció que té més influència és la primera i obliga a tractar els anells A i B de forma conjunta enlloc de fer-ho de forma separada.
- El model a 1  $\mu$ g/planta explica l'*especificitat per substrat* dels BRs, explicant els requeriments estructurals que fan que un BR sigui actiu o inactiu.

- a) Els requisits estructurals per a que un BR sigui actiu (resposta elevada a 1 μg/planta) són:
  - 1. La presència simultània dels cinc grups funcionals polars degut a la seva contribució a la formació d'enllaços d'hidrogen. Per zones la seva influència és:
    - o Diol  $2\alpha$ ,  $3\alpha$ .
    - o Carbonil a l'anell B amb una major contribució del carbonil de la lactona.
    - Diol de la cadena lateral tan en configuració 22*R*,23*R* com 22*S*,23*S*.
  - 2. La presència d'hidrogen (absència de funcionalitat) en les posicions  $3\beta$  i  $5\alpha$  degut a la seva contribució estèrica.
  - 3. La cadena lateral dels 22*R*,23*R*-BRs quantitativament per la seva contribució estèrica en la zona compresa entre l'anell C i el metil C21, qualitativament perquè la conformació activa es troba en el mínim absolut d'energia o en un mínim molt proper.
- b) Els requisits estructurals per a que un BR sigui inactiu (resposta baixa a 1  $\mu$ g/planta) són:
  - 1. La presència d'hidrogen (absència de funcionalitat) en les posicions  $2\alpha$  i  $3\alpha$  degut a la seva contribució estèrica.
  - 2. L'absència de funcionalitat ja sigui amb configuració R o S (presència d'hidrogen) en C22 degut a la seva contribució estèrica.
  - 3. La cadena lateral dels 22*S*,23*S*-BRs quantitativament per la seva contribució estèrica en la zona compresa entre els anells B i D, qualitativament perquè la conformació activa es troba en el mínim d'elevada energia respecte del mínim absolut.
- c) Els requisits que són indiferents a l'activitat o inactivitat (resposta a 1  $\mu$ g/planta) són:
  - 1. La presència d'hidroxils en  $2\beta$  i/o  $3\beta$  pel que respecta a la seva contribució a la formació d'enllaços d'hidrogen excepte en la zona comuna amb els hidroxils  $2\alpha$  i  $3\alpha$ .
  - 2. La presència d'un hidroxil en  $5\alpha$  pel que respecta a la seva contribució a la formació d'enllaços d'hidrogen.
  - 3. La presència d'un hidroxil en 11α tan pel que respecta a la seva contribució a la formació d'enllaços d'hidrogen com per la seva contribució estèrica.
  - 4. Les cadenes laterals més enllà de la distinció entre 22*R*,23*R*-BRs i 22*S*,23*S*-BRs.
- El model a QSAR (Model *HIP*) explica l'*afinitat per substrat* dels BRs, explicant els requeriments estructurals que fan que un BR actiu sigui més o menys actiu (grau d'activitat).
  - a) Els requisits estructurals per a que un BR actiu presenti una elevada activitat  $(-\log(dosi_{50\%}))$  són:
    - 1. Les  $2\alpha$ , $3\alpha$ -lactones degut a la seva contribució a la formació d'enllaços d'hidrogen, tenint en compte que l'anell B s'explica a través de la cadena lateral.
    - 2. La presència d'hidrogen (absència de funcionalitat) en  $3\beta$  degut a la seva contribució estèrica.
    - 3. La cadena lateral dels 22*R*,23*R*-BRs quantitativament per la seva contribució estèrica en la zona compresa entre l'anell C i el metil C21, qualitativament perquè la conformació activa es troba en el mínim absolut d'energia o en un mínim molt proper.
    - 4. Presència d'un metil o etil amb configuració *S* (cadenes tipus *bl* i *homo*) degut a la seva contribució estèrica.

- b) Els requisits estructurals per a que un BR actiu presenti una baixa activitat  $(-\log(dosi_{50\%}))$  són:
  - 1. Les cetones sense grups funcionals en  $\alpha$  de l'anell A degut a la contribució estèrica dels hidrògens en  $2\alpha$  i  $3\alpha$  (absència de funcionalitats) i en menor grau a la contribució estèrica del carbonil de l'anell B i a la contribució a la formació d'enllaços d'hidrogen del carbonil de l'anell B, tenint en compte que s'explica a través de la cadena lateral.
  - 2. Els BRs amb unió A/B cis i diol  $2\beta$ , $3\beta$  degut a la contribució a la formació d'enllaços d'hidrogen de l'hidroxil  $2\beta$ .
  - 3. La cadena lateral dels 22*S*,23*S*-BRs quantitativament per la seva contribució estèrica en la zona compresa entre els anells B i D, qualitativament perquè la conformació activa es troba en el mínim d'elevada energia respecte del mínim absolut.
  - 4. Presència d'un metil amb configuració R (cadena tipus *epi*) degut a la seva contribució estèrica.
- c) Els requisits que són indiferents al grau d'activitat (-log(dosi<sub>50%</sub>)) són:
  - 1. La presència d'hidroxils en  $2\beta$  i/o  $3\beta$  pel que respecta a la seva contribució a la formació d'enllaços d'hidrogen excepte en la zona comuna amb l'hidroxil  $3\alpha$  i la zona intermèdia del diol.
  - 2. La presència d'un hidroxil en 5α pel que respecta a la seva contribució a la formació d'enllaços d'hidrogen, excepte en la zona comuna amb l'hidroxil 3α.
  - 3. Els BRs amb unió A/B cis degut a la contribució estèrica en la cara  $\alpha$ .
  - La presència d'un hidroxil en 6α o 6β pel que respecta a la seva contribució a la formació d'enllaços d'hidrogen, excepte en la zona comuna amb el carbonil en C6.

# Capítol 4: Activitat.

- S'està aplicant un tractament estadístic massa potent per al càlcul de l'activitat ja que ni reducció dels punts experimentals, ni rectificació de les dades, ni el càlcul de la resposta experimental afecten significativament al seu valor.
- S'ha determinat l'error experimental de la resposta a 1 μg/planta i de l'activitat expressada com a –log(dosi<sub>50%</sub>) i com a –log(dosi<sub>558</sub>). Comparant-los amb els errors dels models, es pot afirmar que el model a 1 μg/planta està força ben ajustat i no té gaire marge de millora, i que el model QSAR pot millorar considerablement, sobretot en la predictibilitat.

### Conclusions finals.

- Només amb els estudis computacionals basats en l'AAA, no és possible explicar de forma inequívoca l'activitat o inactivitat dels anàlegs BRs androstànics ja sigui amb el grup funcional lliure o protegit.
- Els models QSAR estan esgotats. No se'ls pot extreure més informació, degut als desequilibris estructurals del conjunt de BRs que formen part del *"data set"*.
- Com a plantejaments de futur, es proposen dos línees de recerca:
  - a) La síntesi racional de BRs per tal d'equilibrar estructuralment els models.
  - b) Abordar computacionalment l'estudi del receptor, el BRI1.<sup>213-219</sup>

# PART EXPERIMENTAL

# <u>1. Maquinari.</u>

# 1.1. Ordinadors Personals (PC).

S'han treballat progressivament amb:

- PC amb processador 486 a 33 MHz amb 32 Mb de RAM i 3 Gb de disc dur, sota un sistema operatiu Windows 95.
- PC amb processador Pentium a 66 MHz amb 32 Mb de RAM i 4 Gb de disc dur, sota un sistema operatiu Windows 98.
- PC HP Pavilion 8720 amb processador Intel Pentium III a 733 MHz amb 64 Mb de RAM i 15 Gb de disc dur, sota un sistema operatiu Windows 2000.
- PC HP omnibook xe4500 amb processador Intel Pentium 4 mobile a 1,60 GHz amb 256 Mb de RAM i 28 Gb de disc dur, sota un sistema operatiu Windows XP.

També s'ha utilitzat:

• PC HP Pavilion 7965 amb processador Intel Pentium 4 a 1,70 GHz amb 256 Mb de RAM i 60 Gb de disc dur, sota un sistema operatiu Red Hat Linux 8.0.

# 1.2. Estacions de treball (Workstations).

S'han treballat progressivament amb:

- Workstation INDIGO2 de Silicon Graphics, amb un processador R4000/IP22 a 100 MHz amb 32 Mb de RAM i 4 Gb de disc dur, sota un sistema operatiu UNIX (IRIX 6.2).
- Workstation INDIGO2 de Silicon Graphics, amb un processador R10000/IP28 a 195 MHz amb 256 Mb de RAM i 4 Gb de disc dur, sota un sistema operatiu UNIX (inicialment IRIX 6.2, finalment IRIX 6.5).

També s'ha utilitzat:

• Workstation ORIGIN2 de Silicon Graphics, amb quatre processador R10000/IP27 a 180 MHz amb 512 Mb de RAM i dos disc durs de 4 Gb cadascun, sota un sistema operatiu UNIX (inicialment IRIX 6.2, finalment IRIX 6.5).

# 2. Programari.

S'han fet servir diversos programes generals de modelització molecular:

• *Hyperchem*:<sup>220</sup> Programa general de modelització molecular desenvolupat per Hypercube Inc. Funciona en ordinadors personals (PC) sota un entorn de Windows. Incorpora les eines necessàries per a la simulació química i el càlcul i la visualització de propietats moleculars.

- Insight II:<sup>221</sup> Programa general de modelització molecular i química computacional desenvolupat per Molecular Simulations Inc. (MSI) que el 2001 va passar a anomenar-se Accelrys Inc. Funciona en estacions de treball Silicon Graphics amb sistema operatiu UNIX. Funciona amb un sistema de mòduls que donen accés a les diverses eines de càlcul. El mòdul bàsic permet visualitzar, construir i manipular les molècules. Els altres mòduls utilitzats són:
  - a) *Discover*: Incorpora un ampli rang de camps de forces per dur a terme minimitzacions, simulacions de dinàmica molecular i anàlisi conformacional.
  - b) *DeCipher*: Permet l'anàlisi dels resultats de les simulacions de dinàmica molecular.
  - c) *Analysis*: Analitza les característiques estructurals tan de molècules com de propietats atòmiques. Permet animar les trajectòries de les dinàmiques moleculars, visualitzar camps d'interacció molecular, etcètera.
- *SYBYL*:<sup>222</sup> Programa general de modelització molecular desenvolupat per Tripos Inc. Funciona en estacions de treball Silicon Graphics amb sistema operatiu UNIX. Incorpora les eines necessàries per a la modelització molecular que permeten construir estructures, optimitzar-les, comparar-les i visualitzar-les.
- *MOE (Molecular Operating Enviroment)*:<sup>223</sup> Programa general de modelització molecular desenvolupat per Chemical Computing Group que funciona en diferents plataformes i sistemes operatius. En aquesta tesi s'ha treballat sota l'entorn Windows i Linux. Proporciona els instruments necessaris per dur a terme estudis de modelització molecular i química computacional.

S'han utilitzat programes específics per realitzar tasques concretes:

- *GRID*:<sup>224,225</sup> Programa desenvolupat per P. Goodford i comercialitzat per Molecular Discovery Ltd. Es va començar amb la versió 17 que només funcionava amb UNIX i s'ha acabat amb la versió beta de GRID 22 que funciona amb Windows. Proporciona els instruments per calcular multitud de camps d'interacció molecular. Les primeres versions funcionaven amb una interfície de menús (*Great*), posteriorment van incorporar una interfície gràfica (*Greater*) i actualment incorporen un visualitzador (*Gview*).
- *MIPSIM (Molecular Interaction Potential SIMilarity analysis)*:<sup>226</sup> Desenvolupat per el Grup de Recerca en Informàtica Biomèdica (GRIB) de l'Institut Municipal d'Investigació Mèdica (IMIM) de Barcelona adscrit a la Universitat Pompeu Fabra (UPF). Funciona tan en estacions de treball Silicon Graphics amb sistema operatiu UNIX, com en PC amb sistema operatiu Linux. És una eina computacional per a l'exploració automàtica de la similitud entre biomolècules basada en els camps d'interacció molecular.
- *GOLPE (Generating Optimal Linear PLS Estimations)*:<sup>227,228</sup> Desenvolupat i distribuït per MIA S.r.l., només pot ser distribuït als usuaris de SYBYL amb una llicència valida dels mòduls QSAR i Advanced QSAR ja que CoMFA és una metodologia patentada per Tripos Inc. Funciona en estacions de treball Silicon Graphics amb sistema operatiu UNIX. Es va començar amb la versió 4.0 i s'ha acabat amb la 4.5. És una eina quimiomètrica per a construir, validar i interpretar models de 3D-QSAR.

- *TSAR (Tools for Structure-Activity Relationships)*:<sup>229</sup> Fulla de càlcul amb capacitat estadística i funcions de base de dades. Desenvolupat per Oxford Molecular Group el qual l'any 2001 es va integrar a Accelrys Inc. Funciona en estacions de treball Silicon Graphics amb sistema operatiu UNIX. Incorpora les eines necessàries per dur a terme estudis QSAR clàssic. El mòdul *ASP* permet realitzar alineaments moleculars.
- *Excel*:<sup>230</sup> Fulla de càlcul desenvolupada per Microsoft. Funciona en ordinadors personals amb sistema operatiu Windows. Proporciona eines eficaces que permeten analitzar, compartir i administrar grans conjunts de dades.

# 3. Modelització Molecular.

### 3.1. Construcció dels compostos.

• *Construcció de les molècules*: S'ha fet indistintament amb Hyperchem, Insight II i MOE. Es construeixen a partir del BR més similar en la conformació activa d'interès, afegint i/o traient les funcionalitats necessàries. Posteriorment s'optimitza la geometria d'acord amb les opcions per defecte de cada programa.

# 3.2. Anàlisi Conformacional.

### Dinàmica molecular (DM):

- Simulació: Es realitza amb el mòdul Discover del programa Insight. Les condicions de la simulació són: simulació en aigua (dissolvent implícit, ε=80), pas de temps (2 fs), etapa de escalfament (5 ps), duració de la simulació (200 ps), camp de forces (CVFF), algorisme d'integració (Verlet-Leapfrog), conjunt canònic (NVT) i control de temperatura (escalat de velocitats a 300K o 900K).
- Anàlisi dades: Mitjançant els mòduls Analysis i DeCipher, s'exporten les dades de la simulació a un fitxer de text. Es verifica la compatibilitat entre programes pel que fa al punt o la coma decimal i s'importa a *Microsoft Excel*. Mitjançant una macro creada a tal efecte, s'avalua la freqüència dels diferents valors per cada angle de torsió.
- Construcció dels confórmers: Es realitza amb el programa SYBYL mitjançant una macro creada a tal efecte. Mitjançant la combinació dels valors més freqüents de cada angle de torsió es genera una base de dades amb tots els conformers. La mateixa macro optimitza l'energia dels conformers i crea un fitxer de text amb els resultats de l'optimització. Les condicions de l'optimització són: Camp de forces (MM94FF o TriposFF), cargues (Gasteiger-Marsili), algorisme d'optimització (gradient conjugat) i criteri de convergència (gradient < 0.001).</li>

### Dinàmica molecular gestionada a través d'un procés de fusió simulada (DM-SA) :

- *Simulació*: Es realitza amb el mòdul *Discover* del programa *Insight*. Les condicions de la simulació són:
  - Condicions de la fusió simulada: 500 iteracions per cada angle de torsió flexible, simulació en aigua (dissolvent implícit,  $\varepsilon$ =80) i camp de forces (CVFF).

- Condicions de la dinàmica molecular: esquelet esteroide rígid, pas de temps (1 fs), etapa de escalfament de 298K a 900K (500 fs), etapa de la simulació a 900K (1.000 fs), etapa de refredament de 900K a 298K (100 fs), algorisme d'integració (Verlet), conjunt canònic (NVT) i control de temperatura (escalat de velocitats).
- Condicions de l'optimització de l'energia: esquelet esteroide flexible, nombre màxim d'iteracions (300) i algorisme d'integració en tres passos amb criteri de convergència (gradient): a) Steepest Descent (gradient < 1000), b) Gradient Conjugat (gradient < 10) i c) Newton (bfgs, gradient < 0,001).</li>
- Anàlisi dades: Mitjançant els mòduls Analysis i DeCipher, s'exporten les dades de la simulació a un fitxer de text. Es verifica la compatibilitat entre programes pel que fa al punt o la coma decimal i s'importa a *Microsoft Excel*. Mitjançant la mateixa macro d'abans, s'avalua la freqüència dels diferents valors per cada angle de torsió.
- *Confórmers*: Mitjançant una macro d'*Insight* creada a tal efecte, s'extreuen del fitxer de la simulació els conformers d'interès.

# 3.3. Alineament molecular.

- Alineament molecular a traves del Potencial Molecular Electrostàtic (MEP): Es realitza mitjançant el mòdul ASP del programa TSAR. És maximitza l'índex de similitud electrostàtic de Carbó entre la molècula d'estudi i la conformació activa de la brassinolida. L'algorisme d'optimització és SIMPLEX.<sup>231</sup>
- Alineament molecular a traves dels anells C i D de l'esteroide: Es realitza mitjançant una aplicació pròpia de SYBYL. Es minimitza la distància entre els 9 àtoms que formen els anells C i D de l'esquelet esteroide de la molècula d'estudi i de la conformació activa de la brassinolida.
- Alineament molecular a traves dels camps moleculars de GRID: Es realitza mitjançant el mòdul COMP de programa MIPSIM el qual necessita del programa GRID. Les condicions d'alineament són: Prealineament mitjançant els anells C i D<sup>+++++</sup> i utilització de la sonda *oh*.<sup>+++++</sup>

### 3.4. 3D-QSAR.

• *GRID*: Es calcula l'energia d'interacció amb les sondes d'aigua (oh2) i metil (C3) en aquest ordre. La mida de la caixa es defineix de forma que s'estengui 5 Å des de l'estructura dels compostos. S'utilitza un pas de malla de 0,5 Å i s'aplica un tall d'energia de +5 kcal/mol.

<sup>&</sup>lt;sup>+++++</sup> En la mesura del que ha estat possible, també s'ha inclòs l'anell A i/o la cadena lateral total o parcialment.

<sup>&</sup>lt;sup>+++++</sup> S'utilitza la sonda de oh (hidroxil) enlloc de oh2 (aigua) per recomanació dels propis programadors de MIPSIM.

- GOLPE: S'importen els camps moleculars de GRID.
  - Pretractament de les dades: Es realitza prèviament al càlcul dels models i consisteix en: a) es suprimeixen els valors positius per considerar només les interaccions favorables de forma que la interpretació dels coeficients és directe, b) s'aplica un tall d'energia de 0,1 kcal/mol per suprimir les variables amb més soroll, c) s'aplica un tall en la desviació estàndard de 0,1 kcal/mol per suprimir les variables amb menys informació, d) es suprimeixen les variables a 2, 3 i 4 nivells, les quals poden distorsionar el model i d) es realitza un escalat per bloc de les sondes per donar-les el mateix pes.
  - PCA: Es realitza el càlcul de 10 components principals, però només es tenen en compte aquells que es poden interpretar des d'un punt de vista estructural. Els components s'interpreten a partir dels camps de coeficients, ajudant-se en cas que sigui necessari dels gràfics de variables i dels camps de les contribucions de determinats compostos. Aquesta interpretació es corrobora amb el gràfic d'objectes, colorant els compostos segons els pertoqui.
  - PLS: Es realitza el càlcul de 5 variables latents, però només es tenen en compte aquelles on es pot interpretar la relació entre l'estructura dels BRs i l'activitat. Aquesta relació s'interpreta a partir dels gràfics de correlació interna i dels camps de les variables, dels pesos i dels coeficients. Els models es validen aplicant tècniques de validació interna. Concretament s'utilitza el mètode de deixar un a fora (Leave One Out (L.O.O.)).
  - Reducció de variables:<sup>228</sup> S'aplica als models amb bona capacitat predictiva  $(q^2_{(LV1, LOO)} > 0,3)$ . Es realitza una reducció de les variables de l'espai dels pesos, aplicant un disseny de D-òptim. Aquesta reducció s'aplica en diverses etapes fins que el nombre de variables està al voltant de les 5.000–6.000. En cada etapa es redueix el nombre de variables un 50%.
  - Agrupament de variables: Es realitza mitjançant l'algorisme d'agrupament Smart Region Detection (SDR)<sup>232-234</sup> i s'aplica un cop feta la reducció de variables. Les condicions de l'agrupament són: nombre de llavors (10% de les variables actives), distància crítica (5 Å) i distància d'unificació de grups (10 Å).
  - Selecció de variables: S'aplica un cop fet l'agrupament de variables. Mitjançant un disseny factorial fraccionat (FFD)<sup>228,234</sup> es realitzen una sèrie de models reduïts i s'avalua la influència de les variables sobre l'activitat. Les condicions de la selecció de variables son: Aplicació sobre la primera variable latent, utilització dels grups de variables, en cada model es recalculen els pesos, els models es validen internament amb el mètode L.O.O., la relació entre el nombre d'experiments i de variables és de 5 i en el model final es mantenen les variables incertes.
- *Predicció de l'activitat*: Es construeixen els compostos amb l'alineament i la conformació activa que li pertoca segons els model. Es calculen els camps moleculars de GRID en les mateixes condicions que les del model (sondes, mida de caixa i pas de malla). Aquests s'importen a GOLPE i es realitza el següent pretractament de les dades: a) es suprimeixen els valors positius i b) es realitza un escalat per bloc de les sondes assignant-los el mateix pes que tenen en el model. Finalment, s'utilitza l'aplicació corresponent per predir l'activitat.

Intersecció dels mapes de GRID: Es calcula l'energia d'interacció amb la sonda d'aigua (oh2) dels dos compostos a comparar. Les condicions de la caixa no poden excedir el nombre de 65.536 variables (nombre màxim de files permeses per Microsoft Excel). Per aquest motiu la mida de la caixa ha d'incloure només les regions d'interès i s'utilitza un pas de malla de 1 Å o 0,5 Å. S'aplica un tall d'energia de +5 kcal/mol i es demana que el fitxer amb les energies estigui en format ASCII. Es verifica la compatibilitat entre programes pel que fa al punt o la coma decimal i s'importen les energies a Microsoft Excel. Mitjançant una plantilla creada a tal efecte, es calcula el mapa d'energies corresponents a les zones comuns, es a dir, la intersecció. Es crea un fitxer de GRID amb les energies de la interacció i es torna a verificar la compatibilitat entre programes pel que fa al punt o la coma decimal. Finalment, per poder visualitzar aquest fitxer cal canviar el format de "text - Windows" a "text - UNIX".

# 4. Estadística.

# 4.1. Proves d'hipòtesi.<sup>235</sup>

Proves d'hipòtesi sobre les mitjanes de dos distribucions normals de variàncies desconegudes i diferents. (-2, -2), ) **T** T

Hipòtesi

Final Figures nullia:  

$$H_{0}: \mu_{1} = \mu_{2} \quad (\sigma_{1}^{*} \neq \sigma_{1}^{*} \quad desconegudes)$$
Estadístic de prova:  

$$t_{0} = \frac{\overline{x}_{1} - \overline{x}_{2}}{\sqrt{\frac{s_{1}^{2}}{n_{1}} + \frac{s_{2}^{2}}{n_{2}}}, \quad v = \frac{\left(\frac{s_{1}^{2}}{n_{1}} + \frac{s_{2}^{2}}{n_{2}}\right)^{2}}{\frac{\left(s_{1}^{2}/n_{1}\right)^{2}}{n_{1} + 1} + \frac{\left(s_{2}^{2}/n_{2}\right)^{2}}{n_{2} + 1}}$$

Hipòtesi alternativa i criteri de rebuig:

• 
$$H_1: \mu_1 \neq \mu_2$$
,  $|t_0| > t_{\frac{\alpha}{2}, \nu}$ 

S'ha fet servir per determinar amb un 95% de seguretat, si la mitjana de la resposta de la hcs a 1 μg/planta d'un bioassaig és igual a la mitjana de totes les dades.

Proves d'hipòtesi sobre les mitjanes de dos distribucions normals de variàncies conegudes.

Hipòtesi nul·la:  
Estadístic de prova:  

$$H_{0}: \mu_{1} = \mu_{2} \quad \left(\sigma_{1}^{2} i \sigma_{1}^{2} \quad conegudes\right)$$

$$Z_{0} = \frac{\overline{x_{1}} - \overline{x_{2}}}{\sqrt{\frac{s_{1}^{2}}{n_{1}} + \frac{s_{2}^{2}}{n_{2}}}}$$

Hipòtesi alternativa i criteri de rebuig:

• 
$$H_1: \mu_1 \neq \mu_2$$
,  $|z_0| > z_{\frac{\alpha}{2}}$ 

S'ha fet servir per determinar amb un 95% de seguretat, si l'activitat d'un compost és igual a la d'un altre.

2

# **BIBLIOGRAFIA**

- 1. Johri, M. M.; Mitra, D. Action of plant hormones. Curr. Sci. 2001, 80 (2), 199-205.
- 2. Gray, W. M. Hormonal regulation of plant growth and development. *PLoS Biology* **2004**, *2* (9), 1270-1273.
- 3. Seitz, H. U. Plant hormones control growth and development. *Praxis der Naturwissenschaften, Biologie in der Schule* **2005,** *54* (6), 1-5.
- 4. Dharmasiri, N.; Estelle, M. Auxin signaling and regulated protein degradation. *Trends in Plant Science* **2004**, *9* (6), 302-308.
- 5. Gomi, K.; Matsuoka, M. Gibberellin signalling pathway. *Current Opinion in Plant Biology* **2003**, 6 (5), 489-493.
- 6. Guo, H. W.; Ecker, J. R. The ethylene signaling pathway: new insights. *Current Opinion in Plant Biology* **2004**, *7* (1), 40-49.
- 7. Himmelbach, A.; Yang, Y.; Grill, E. Relay and control of abscisic acid signaling. *Current Opinion in Plant Biology* **2003**, 6 (5), 470-479.
- 8. Kakimoto, T. Perception and signal transduction of cytokinins. *Annual Review of Plant Biology* **2003**, *54*, 605-627.
- 9. Turner, J. G.; Ellis, C.; Devoto, A. The jasmonate signal pathway. *Plant Cell* **2002**, *14*, 153-164.
- 10. Wang, Z. Y.; He, J. X. Brassinosteroid signal transduction choices of signals and receptors. *Trends in Plant Science* **2004**, *9* (2), 91-96.
- 11. Napier, R. Plant hormone binding sites. Annals of Botany 2004, 93 (3), 227-233.
- 12. Neljubov, D. N. Uber die horizontale Nutation der Stengel von *Pisum sativum* und einiger anderen. *Pflanzen. Beiheftezum Botanischen Zentralblatt* **1901**, (10), 128-139.
- 13. Haberlandt, G. Zur Physiologie der Zellteilung. Sitzber. K. Preuss. Akad. Wiss. 1913, 319.
- 14. Went, F. W. On growth-accelerating substances in the coleoptile of *Avena sativa*. Amsterdam, **1926**; pp 10-19.
- 15. Went, F. W. Wuchsstoff und wachstum. Rec. Trav. Bot. Neerl. 1928, 25, 1-116.
- 16. Yabuta, T. Biochemistry of the 'bakanae' fungus of rice. Agr. Hort. 1935, 10, 17-22.
- 17. Heacock, R. A.; Wain, R. L.; Wightman, F. Plant-growth-regulating substances. XII. Polycyclic acids. *Annals of Applied Biology* **1958**, *46*, 352-365.
- 18. Ohkuma, K.; Smith, O. E.; Lyon, J. L.; Addicott, F. T. Abscisin 2, An Abscission-Accelerating Substance from Young Cotton Fruit. *Science* **1963**, *142* (359), 1592.
- 19. Bagni, N.; Serafini Fracassini, D. The role of polyamines as growth factors in higher plants and their mechanism of action. **1974**; pp 1205-1217.
- Grove, M. D.; Spencer, G. F.; Rohwedder, W. K.; Mandava, N.; Worley, J. F.; Warthen, J. D., Jr.; Steffens, G. L.; Flippen-Anderson, J. L.; Cook, J. C., Jr. Brassinolide, a plant growth-promoting steroid isolated from *Brassica napus* pollen. *Nature (London, United Kingdom)* 1979, 281 (5728), 216-217.
- 21. Parthier, B. Jasmonates Hormonal Regulators Or Stress Factors in Leaf Senescence. *Journal of Plant Growth Regulation* **1990**, *9* (1), 57-63.
- 22. Heftmann, E. Functions of Sterols in Plants. Lipids 1971, 6 (2), 128.
- 23. Heftmann, E. Functions of Steroids in Plants. *Phytochemistry* **1975**, *14* (4), 891-901.
- 24. Geuns, J. M. C. Steroid-Hormones and Plant-Growth and Development. *Phytochemistry* **1978**, *17* (1), 1-14.
- 25. Mitchell, J. W.; Whitehead, M. R. Responses of vegetative parts of plants following application of extract of pollen from *Zea mays*. *Botanical gazette* **1941**, *102*, 770-790.
- 26. Mandava, N.; Mitchell, J. W. New plant hormones. Chemical and biological investigations. *Indian Agriculturist* **1971**, *15* (1-2), 19-31.
- 27. Mitchell, J. W.; Mandava, N.; Worley, J. F.; Plimmer, J. R.; Smith, M. V. Brassins--a new family of plant hormones from rape pollen. *Nature (London, United Kingdom)* **1970**, 225 (5237), 1065-1066.

- 28. Mandava, N.; Mitchell, J. W. Structural elucidation of brassins. *Chemistry & Industry* (London, United Kingdom) **1972**, (23), 930-931.
- Wada, K.; Marumo, S.; Ikekawa, N.; Morisaki, M.; Mori, K. Brassinolide and Homobrassinolide Promotion of Lamina Inclination of Rice Seedlings. *Plant Cell Physiol*. **1981**, 22 (2), 323-325.
- 30. Yokota, T.; Arima, M.; Takahashi, N. Castasterone, A New Phytosterol with Plant-Hormone Potency, from Chestnut Insect Gall. *Tetrahedron Lett.* **1982**, *23* (12), 1275-1278.
- 31. Yokota, T.; Ogino, Y.; Takahashi, N.; Saimoto, H.; Fujioka, S.; Sakurai, A. Brassinolide Is Biosynthesized from Castasterone in *Catharanthus-Roseus* Crown Gall Cells. *Agricultural and Biological Chemistry* **1990**, *54* (4), 1107-1108.
- 32. Ikekawa, N.; Takatsuto, S.; Kitsuwa, T.; Saito, H.; Morishita, T.; Abe, H. Analysis of Natural Brassinosteroids by Gas-Chromatography and Gas-Chromatography Mass-Spectrometry. *J. Chromatogr.* **1984**, 290, 289-302.
- 33. Fujioka, S. Natural occurrence of brassinosteroids in the plant kingdom. In *Brassinosteroids: Steroidal Plant Hormones.*, Sakurai, A., Yokota, T., clouse, S. D., Eds.; Springer-Verlag, Tokyo, **1999**; pp 21-45.
- 34. Fujioka, S.; Sakurai, A. Biosynthesis and metabolism of brassinosteroids. *Physiol. Plant.* **1997**, 100 (3), 710-715.
- 35. Sakurai, A.; Fujioka, S. Studies on biosynthesis of brassinosteroids. *Bioscience Biotechnology* and *Biochemistry* **1997**, *61* (5), 757-762.
- 36. Fujioka, S.; Li, J.; Choi, Y. H.; Seto, H.; Takatsuto, S.; Noguchi, T.; Watanabe, T.; Kuriyama, H.; Yokota, T.; Chory, J.; Sakurai, A. The *Arabidopsis* deetiolated2 mutant is blocked early in brassinosteroid biosynthesis. *The Plant cell* **1997**, *9* (11), 1951-1962.
- 37. Choe, S.; Dilkes, B. P.; Fujioka, S.; Takatsuto, S.; Sakurai, A.; Feldmann, K. A. The DWF4 gene of Arabidopsis encodes a cytochrome P450 that mediates multiple 22α-hydroxylation steps in brassinosteroid biosynthesis. *Plant Cell* **1998**, *10* (2), 231-243.
- Sakurai, A. Biosynthesis. In Brassinosteroids: Steroidal Plant Hormones., Sakurai, A., Yokota, T., Clouse, S. D., Eds.; Springer-Verlag, Tokyo, 1999; pp 91-111.
- 39. Sasse, J. Physiological Actions of Brassinosteroids. In *Brassinosteroids: Steroidal Plant Hormones.*, Sakurai, A., Yokota, T., Clouse, S. D., Eds.; Springer-Verlag, Tokyo, **1999**; pp 21-45.
- 40. Braun, P.; Wild, A. The influence of brassinosteroid on growth and parameters of photosynthesis of wheat and mustard plants. *J. Plant Physiol.* **1984**, *116* (3), 189-196.
- 41. Haubrick, L. L.; Assmann, S. M. Brassinosteroids and plant function: some clues, more puzzles. *Plant, Cell Environ.* **2006**, 29 (3), 446-457.
- 42. Khripach, V.; Zhabinskii, V.; de Groot, A. Twenty Years of Brassinosteroids: Steroidal Plant Hormones Warrant Better Crops for the XXI Century. *Annals of Botany (London)* **2000**, *86* (3), 441-447.
- 43. Sasse, J. Physiological Actions of Brassinosteroids: An Update. *Journal of Plant Growth Regulation* **2003**, *22* (4), 276-288.
- 44. Clouse, S. D.; Sasse, J. M. Brassinosteroids: Essential regulators of plant growth and development. *Annual Review of Plant Physiology and Plant Molecular Biology* **1998**, 49, 427-451.
- 45. Hamada, K. Brassinolide in crop cultivation. *FFTC Book Series* **1986**, *34* (Plant Growth Regul. Agric.), 188-197.
- 46. Mandava, N. B. Plant growth-promoting brassinosteroids. *Annual Review of Plant Physiology* and Plant Molecular Biology **1988**, 39, 23-52.
- 47. Fujioka, S.; Sakurai, A. Brassinosteroids. Nat. Prod. Rep. 1997, 14 (1), 1-10.
- 48. Anuradha, S.; Rao, S. S. R. Effect of brassinosteroids on salinity stress induced inhibition of seed germination and seedling growth of rice (*Oryza sativa L*.). *Plant Growth Regulation* **2001**, *33* (2), 151-153.
- 49. Sasse, J.; Smith, R.; Hudson, I. Effect of 24-epibrassinolide on germination of seeds of *Eucalyptus camaldulensis* in saline conditions. *Proceedings of the Plant Growth Regulator Society of America* **1995**, *22nd*, 136-141.

- 50. Dhaubhadel, S.; Chaudharyl, S.; Dobinson, K. F.; Krishna, P. Treatment with 24epibrassinolide, a brassinosteroid, increases the basic thermotolerance of *Brassica napus* and tomato seedlings. *Plant Mol. Biol.* **1999**, 40 (2), 333-342.
- 51. Wilen, R. W.; Sacco, M.; Gusta, L. V.; Krishna, P. Effects of 24-epibrassinolide on freezing and thermotolerance of bromegrass (*Bromus inermis*) cell cultures. *Physiol. Plant.* **1995**, *95* (2), 195-202.
- 52. Sairam, R. K. Effects of homobrassinolide application on plant metabolism and grain yield under irrigated and moisture-stress conditions of two wheat varieties. *Plant Growth Regulation* **1994**, *14* (2), 173-181.
- 53. Schilling, G.; Schiller, C.; Otto, S. Influence of brassinosteroids on organ relations and enzyme activities of sugar-beet plants. *ACS Symp. Ser.* **1991**, *474* (Brassinosteroids), 208-219.
- 54. Bao, F.; Shen, J.; Brady, S. R.; Muday, G. K.; Asami, T.; Yang, Z. Brassinosteroids interact with auxin to promote lateral root development in *Arabidopsis*. *Plant Physiol*. **2004**, *134* (4), 1624-1631.
- 55. Handschack, M.; Adam, G.; Vorbrodt, H. M.; Marquardt, V.; Beinhauer, K. Stone fruit quality improvement by synergistic mixtures of auxins with brassinosteroids. *Patent*, 4039017, Germany, **1992**.
- 56. Yopp, J. H.; Mandava, N. B.; Thompson, M. J.; Sasse, J. M. Activity of brassinosteroid in selected bioassays in combination with chemicals known to synergize or retard responses to auxin and gibberellin. *Proceedings Plant Growth Regulation Society of America* **1981**, *8th*, 138-145.
- 57. Hetru, C.; Roussel, J. P.; Mori, K.; Nakatani, Y. Antiecdysteroid Activity of Brassinosteroids. *Comptes Rendus de l Academie des Sciences Serie li* **1986**, *302* (7), 417-420.
- 58. Richter, K.; Adam, G. Neurodepressing effect of brassinosteroids in the cockroach *Periplaneta americana*. *Naturwissenschaften* **1991**, *78* (3), 138-139.
- 59. Gruber, C. M.; Halbeisen, W. A. A Study on the Comparative Toxic Effects of Citric Acid and Its Sodium Salts. *J. Pharmacol. Exp. Ther.* **1948**, *94* (1), 65-67.
- 60. Boyd, E. M.; Shanas, M. N. Acute Oral Toxicity of Sodium Chloride. *Archives Internationales de Pharmacodynamie et de Therapie* **1963**, 144, 86.
- 61. Abe, H. Advances in brassinosteroid research and prospects for its agricultural application. *Japan Pesticide Information* **1989**, *55*, 10-14.
- 62. Kamuro, Y.; Takatsuto, S. Practical Application of Brassinosteroids in Agricultural Fields. In *Brassinosteroids: Steroidal Plant Hormones.*, Sakurai, A., Yokota, T., Clouse, S. D., Eds.; Springer-Verlag, Tokyo, **1999**; pp 223-241.
- 63. Khripach, V. A.; Zhabinskii, V. N.; Malevannaya, N. N. Recent advances in brassinosteroids study and application. *Proceedings of the Plant Growth Regulator Society of America* **1997**, 24th, 101-106.
- 64. Kamuro, Y.; Takatsuto, S. Capability for and Problems of Practical Uses of Brassinosteroids. *ACS Symp. Ser.* **1991**, *474*, 292-297.
- 65. Kamuro, Y.; Takatsuto, S.; Watanabe, T.; Noguchi, T.; Kuriyama, H.; Suganuma, H. Practical effects of brassinosteroid compound [TS303]. *Proceedings of the Plant Growth Regulator Society of America* **1997**, *24th*, 111-116.
- 66. Kohout, L.; Strnad, M.; Kaminek, M. Types of brassinosteroids and their bioassays. *ACS Symp. Ser.* **1991**, *474* (Brassinosteroids), 56-73.
- 67. Sasse, J. M. Recent progress in brassinosteroid research. *Physiol. Plant.* **1997,** *100* (3), 696-701.
- 68. Maeda, E. Rate of lamina inclination in excised rice leaves. *Physiol. Plant.* **1965**, *18* (3), 813-827.
- 69. Mitchel, J. W.; Livingstone, G. A. Methods of studying plant hormones and growthregulating substances. In *Agricultural Handbook No. 336*, Agric Res Ser USDA, **1968**.
- 70. Wada, K.; Marumo, S.; Abe, H.; Morishita, T.; Nakamura, K.; Uchiyama, M.; Mori, K. A rice lamina inclination test a microquantitative bioassay for brassinosteroids. *Agricultural and Biological Chemistry* **1984**, *48* (3), 719-726.
- 71. Takeno, K.; Pharis, R. P. Brassinosteroid-Induced Bending of the Leaf Lamina of Dwarf Rice Seedlings An Auxin-Mediated Phenomenon. *Plant Cell Physiol.* **1982**, *23* (7), 1275-1281.

- 72. Wada, K.; Marumo, S. Synthesis and plant growth-promoting activity of brassinolide analogs. *Agricultural and Biological Chemistry* **1981,** *45* (11), 2579-2585.
- 73. Thompson, M. J.; Meudt, W. J.; Mandava, N. B.; Dutky, S. R.; Lusby, W. R.; Spaulding, D. W. Synthesis of brassinosteroids and relationship of structure to plant growth-promoting effects. *Steroids* **1982**, *39* (1), 89-105.
- 74. Cerana, R.; Lado, P.; Anastasia, M.; Ciuffreda, P.; Allevi, P. Regulating effects of brassinosteroids and of sterols on growth and proton secretion in maize roots. *Zeitschrift fuer Pflanzenphysiologie* **1984**, *114* (3), 221-225.
- 75. Takatsuto, S.; Yazawa, N.; Ikekawa, N.; Takematsu, T.; Takeuchi, Y.; Koguchi, M. Structureactivity relationship of brassinosteroids. *Phytochemistry (Elsevier)* **1983**, *22* (11), 2437-2441.
- 76. Takatsuto, S.; Yazawa, N.; Ikekawa, N.; Morishita, T.; Abe, H. Synthesis of (24*R*)-28homobrassinolide analogs and structure-activity relationships of brassinosteroids in the ricelamina inclination test. *Phytochemistry (Elsevier)* **1983**, *22* (6), 1393-1397.
- 77. Takatsuto, S.; Yazawa, N.; Ikekawa, N. Synthesis and biological activity of brassinolide analogs, 26,27-bisnorbrassinolide and its 6-oxo analog. *Phytochemistry (Elsevier)* **1984**, *23* (3), 525-528.
- 78. Yokota, T.; Mori, K. Molecular structure and biological activity of brassinolide and related brassinosteroids. *Mol. Struct. Biol. Act. Steroids* **1992**, 317-340.
- 79. Capdevila, E.; Molist, M.; Vilaplana-Polo, M.; Brosa, C. 20 years of research on brassinosteroids at the Steroids laboratory of IQS. *Afinidad* **2007**, accepted.
- 80. Ferrer, J. C.; Lalueza, R.; Saavedra, O.; Brosa, C. Short Step Synthesis of (22E,24R)-5-Alpha-Ergosta-2,22-Dien-6-One, A Key Intermediate for the Preparation of 24-Epibrassinolide. *Tetrahedron Lett.* **1990**, *31* (27), 3941-3942.
- 81. Brosa, C.; Puig, R.; Comas, X.; Fernandez, C. New synthetic strategy for the synthesis of 24epibrassinolide. *Steroids* **1996**, *61* (9), 540-543.
- 82. Takatsuto, S.; Ikekawa, N. Synthesis of (22R,23R)-28-Homobrassinolide. *Chemical & Pharmaceutical Bulletin* **1982**, *30* (11), 4181-4185.
- 83. Brosa, C.; Peracaula, R.; Puig, R.; Ventura, M. Use of dihydroquinidine 9-O-(9'-phenanthryl) ether in osmium-catalyzed asymmetric dihydroxylation in the synthesis of brassinosteroids. *Tetrahedron Lett.* **1992**, *33* (46), 7057-7060.
- 84. Soca, L. Brassinoesteroides: Síntesi, avaluació de l'activitat i aplicacions a camp. *Tesi doctoral*, Institut Químic de Sarrià C.E.T.S., Barcelona, **1998**.
- 85. Molist, M. Brassinoesteroides: Síntesi d'anàlegs androstànics encaminats en aprofundir en la interacció brassinoesteroide-receptor. *Tesi doctoral*, IQS CETS, Barcelona, **2007**.
- 86. Soca, L.; Brosa, C. Intent de síntesi de (22*E*,24*R*) 5α-ergosta-1,22-dien-6β-ol. Determinació de l'activitat d'anàlegs brassinoesteroides mitjançant els tests d'inclinació de la làmina d'arrós i cultius de cèl·lules d'api. *Comunicació personal*, Institut Químic de Sarrià C.E.T.S., Barcelona, **1992**.
- 87. Puig, R. Brassinoesteroides: Síntesi, bioactivitat i modelització molecular. *Tesi doctoral*, Institut Químic de Sarrià C.E.T.S., Barcelona, **1993**.
- 88. Masó, M.; Brosa, C. Síntesi d'anàlegs brassinoesteroides amb funcionalitat 6-oxa i posada a punt del test d'inclinació de la làmina d'arròs amb planta tallada. *Comunicació personal*, Institut Químic de Sarrià C.E.T.S., Barcelona, **1998**.
- 89. Terricabras, E.; Brosa, C. Adaptación del test de la lámina de arroz a variedades autóctonas: Síntesis y determinación de la actividad de brasinoesteroides. *Comunicació personal*, Institut Químic de Sarrià C.E.T.S., Barcelona, **1994**.
- 90. Terricabras, E. Brassinoesteroides: Síntesi d'anàlegs, nou test d'inclinació de la làmina d'arròs i aplicació a camp. *Tesi doctoral*, Institut Químic de Sarrià C.E.T.S., Barcelona, **1999**.
- 91. Miró, X.; Brosa, C. Brassinoesteroides: síntesi de nous anàlegs amb funcions epòxid, evolució de la resposta en el bioassaig i aplicació de la 24-epibrassinolida sobre cultius de gira-sol. *Comunicació personal*, Institut Químic de Sarrià C.E.T.S., Barcelona, **1995**.
- 92. Kohout, L.; Velgova, H.; Strnad, M.; Kamínek, M. Brassinosteroids with Androstane and Pregnane Skeleton. *Collect. Czech. Chem. Commun.* **1987**, *52*, 476.
- 93. Kohout, L. Steroids. Part CCCXLVIII. Synthesis of brassino steroids with a five carbon atom ester functionality in position 17. *Collect. Czech. Chem. Commun.* **1989**, *54* (12), 3348-3359.

- 94. Kohout, L.; Kasal, A. On Steroids .374. 7A-Keto-B-Homotestosterone and 7A-Keto-B-Homoepitestosterone. *Collect. Czech. Chem. Commun.* **1994**, 59 (3), 649-657.
- 95. Zamora, I. Diseño y aplicación de metodologias para interpretar la influencia de la estructura en moléculas bioactivas mediante técnicas computacionales. Apliación sobre brasinoesteroides y aproximación a la síntesis de nuevos análogos. *Tesi doctoral*, Institut Químic de Sarrià C.E.T.S., Barcelona, **1998**.
- 96. Brosa, C.; Zamora, I.; Terricabras, E.; Kohout, L. The effect of electrostatic properties and ability to form hydrogen-bonds on the activity of brassinosteroid side-chain analogs. *Collect. Czech. Chem. Commun.* **1998**, *63* (10), 1635-1645.
- 97. Brosa, C. Structure-Activity Relationship. In *Brassinosteroids: Steroidal Plant Hormones.*, Sakurai, A., Yokota, T., Clouse, S. D., Eds.; Springer-Verlag, Tokyo, **1999**; pp 191-222.
- 98. Brosa, C.; Nusimovich, S.; Peracaula, R. Synthesis of new brassinosteroids with potential activity as antiecdysteroids. *Steroids* **1994**, *59* (8), 463-467.
- 99. Brosa, C.; Capdevila, J. M.; Zamora, I. Brassinosteroids: a new way to define the structural requirements. *Tetrahedron* **1996**, *52* (7), 2435-2448.
- 100. Li, J.; Nagpal, P.; Vitart, V.; McMorris, T. C.; Chory, J. A role for brassinosteroids in lightdependent development of *Arabidopsis*. *Science (Washington, D. C. )* **1996,** *272* (5260), 398-401.
- 101. Li, J. M.; Chory, J. A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. *Cell* **1997**, *90* (5), 929-938.
- Clouse, S. D.; Feldmann, K. A. Molecular Genetics of Brassinosteroids Action. In Brassinosteroids: Steroidal Plant Hormones., Sakurai, A., Yokota, T., Clouse, S. D., Eds.; Springer-Verlag, Tokyo, **1999**; pp 163-189.
- 103. Marshall, G. R.; Barry, C. D.; Bosshard, H. E.; Dammkoehler, R. A.; Dunn, D. A. The conformational parameter in drug design: the active analog approach. *ACS Symp. Ser.* **1979**, *112* (Comput.-Assisted Drug Des.), 205-226.
- 104. Beusen, D. D.; Marshall, G. R. Pharmacophore Definition Using the Active Analog Approach. In *Pharmacophore Perception, Development and Use in Drug Design,* International University Line, La Jolla, CA, USA, **2000**; pp 21-45.
- 105. Klebe, G.; Mietzner, T.; Weber, F. Different Approaches Toward An Automatic Structural Alignment of Drug Molecules Applications to Sterol Mimics, Thrombin and Thermolysin Inhibitors. *Journal of Computer-Aided Molecular Design* **1994**, *8* (6), 751-778.
- 106. Brosa, C.; Soca, L.; Terricabras, E.; Zamora, I. Brassinosteroids: looking for a practical solution. *Proc Plant Growth Regul Soc Am* **1996**, *23*, 21-26.
- 107. Brosa, C.; Zamora, I.; Teixidó, J. A structure-activity relationships for brassinosteroids. Lausanne, **1996**.
- 108. Free, S. M.; Wilson, J. W. Mathematical Contribution to Structure-Activity Studies. J. Med. Chem. **1964**, *7* (4), 395.
- 109. Kubinyi, H.; Kehrhahn, O. H. Quantitative Structure-Activity-Relationships .3. Comparison of Different Free-Wilson Models. J. Med. Chem. **1976**, *19* (8), 1040-1049.
- 110. Kubinyi, H. Quantitative Structure-Activity-Relationships .2. Mixed Approach, Based on Hansch and Free-Wilson Analysis. J. Med. Chem. **1976**, *19* (5), 587-600.
- 111. Kubinyi, H.; Kehrhahn, O. H. Quantitative Structure-Activity-Relationships .1. Modified Free-Wilson Approach. J. Med. Chem. **1976**, *19* (5), 578-586.
- 112. Brosa, C. Biological effects of brassinosteroids. *Biochemistry and Function of Sterols* **1997**, 201-220.
- Carbo, R.; Leyda, L.; Arnau, M. How similar is a molecule to another? An electron density measure of similarity between two molecular structures. *Int. J. Quantum Chem.* **1980**, *17* (6), 1185-1189.
- 114. Julià, G.; Brosa, C. Modelització d'anàlegs brassinoesteroides fluorats en cadena i aproximació a la seva síntesi. *Comunicació personal*, Institut Químic de Sarrià C.E.T.S., Barcelona, **1998**.
- 115. Amorós, M.; Brosa, C. Estudi de noves metodologies per obtenir anàlegs brassinoesteroides amb funcionalitat amino i fluor e l'anell A. *Comunicació personal*, Institut Químic de Sarrià C.E.T.S., Barcelona, **2000**.

- 116. David, A.; Brosa, C. Tentative de fluoration en position 3α de brassinoestéroïdes. *Comunicació personal*, Institut Químic de Sarrià C.E.T.S., Barcelona, **2000**.
- 117. Amorós, M. Brassinoesteroides: estudis d'estructura-funció mitjançant l'obtenció d'anàlegs amb modificacions estructurals. *Tesi doctoral*, Institut Químic de Sarrià C.E.T.S., Barcelona, **2005**.
- Puigdengolas, S.; Brosa, C. Brassinoesteroides: Síntesi d'anàlegs amb funcionalitat C-2 i avaluació de l'activitat. *Comunicació personal*, Institut Químic de Sarrià C.E.T.S., Barcelona, 2005.
- 119. Davi, C.; Brosa, C. Aproximació a la síntesi d'anàlegs brassinoesteroides amb funcionalitat amino a C3: síntesi d'azidoderivats. *Comunicació personal*, Institut Químic de Sarrià C.E.T.S., Barcelona, **2001**.
- Molist, M.; Brosa, C. Síntesi d'anàlegs brassinoesteroides amb una sola funcionalitat amina i azida en l'anell A. *Comunicació personal*, Institut Químic de Sarrià C.E.T.S., Barcelona, 2003.
- 121. Brosa, C.; Amoros, M.; Molist, M.; Hernandez, X. New brassinosteroid analogs having nitrogenated functionalities at C-3 to provide more information about the brassinosteroid-receptor interaction. *Tetrahedron* **2004**, *60* (38), 8529-8534.
- 122. Kim, S. K.; Mizuno, K.; Hatori, M.; Marumo, S. A Brassinolide-Inhibitor Km-01, Its Isolation and Structure Elucidation from A Fungus Drechslera-Avenae. *Tetrahedron Lett.* **1994**, *35* (11), 1731-1734.
- 123. Cruciani, G.; Goodford, P. J. A Search for Specificity in Dna-Drug Interactions. *Journal of Molecular Graphics* **1994**, *12* (2), 116-129.
- 124. Favre, H. A.; Giles, R. M.; Hellwich, K. H.; McNaught, A. D.; Moss, G. P.; Powell, W. H. Natural products and related compounds (IUPAC Recommendations 1999) (vol 71, pg 587, 1999). *Pure Appl. Chem.* **2004**, *7*6 (6), 1283-1292.
- 125. Moss, G. P. The Nomenclature of Steroids Recommendations 1989. Eur. J. Biochem. 1989, 186 (3), 429-458.
- 126. Cutler, H. G.; Yokota, T.; Adam, G. *Brassinosteroids: chemistry, bioactivity and applications;* ACS Symposium Series: Washington DC, **1991**.
- 127. Khripach, V. A.; Zhabinskii, V. N.; De Groot, A. E. *Brassinosteroids: A new class of plant hormones;* Academic Press: EUA, **1998**.
- 128. Sakurai, A.; Yokota, T.; Clouse, S. D. *Brassinosteroids*. *Steroidal plant hormones;* Springer: Tokyo, **1999**.
- 129. Zullo, M. A. T.; Kohout, L. Semisystematic nomenclature of brassinosteroids. XX Conference on Isoprenoids, Liberec, Czech Republic, **2003**.
- 130. Zullo, M. A. T.; Kohout, L. Semisystematic nomenclature of brassinosteroids. *Plant Growth Regulation* **2004**, *42* (1), 15-28.
- VanDeWaterbeemd, H.; Carter, R. E.; Grassy, G.; Kubinyi, H.; Martin, Y. C.; Tute, M. S.; Willett, P.; Haasnoot, C. A. G.; Kier, L. B.; Muller, K.; Rose, S. V.; Weber, J.; Wibley, K. S.; Wold, S.; Boyd, D. B.; Clark, D. E.; deHaen, C.; Heindel, N. D.; Kratochvil, P.; Kutscher, B.; Lewis, R. A.; Mabilia, M.; Metanomski, W. V.; Polymeropoulos, E. E.; Tollenaere, J. P.; Turnbull, M. D.; vanderLinden, W. E.; VanLenten, E. J. Glossary of terms used in computational drug design. *Pure Appl. Chem.* **1997**, *69* (5), 1137-1152.
- 132. Muller, P. Glossary of Terms Used in Physical Organic-Chemistry. *Pure Appl. Chem.* **1994**, 66 (5), 1077-1184.
- 133. Richet, M. C. Noté sur le rapport entre la toxicité et les propiétés physiques des corps. *Comt. Rend. Soc. Biol.* **1893,** *45*, 775-776.
- 134. Ehrlich, P. Address in pathology on chemotherapeutics: Scientific principles, methods and results. *Lancet II* **1913**, 445-451.
- 135. Lipnick, R. L. Overton, Charles, Ernest Narcosis Studies and A Contribution to General Pharmacology. *Trends Pharmacol. Sci.* **1986**, *7* (5), 161-164.
- 136. Lipnick, R. L. Meyer, Hans, Horst and the Lipoid Theory of Narcosis. *Trends Pharmacol. Sci.* **1989**, *10* (7), 265-269.
- 137. Hansch, C.; Fujita, T. Rho-Sigma-Pi Analysis . Method for Correlation of Biological Activity + Chemical Structure. J. Am. Chem. Soc. **1964**, *86* (8), 1616.

- 138. Ramsden, C. A. The rational design, mechanistic study & therapeutic application of chemical compounds. In *Quantitative Drug Design, Vol 4 of Comprehensive Medicinal Chemistry,* Hansch, C., Sammes, P. G., Taylor, J. B., Eds.; Pergamon Press, Oxford, **1990**.
- 139. Kubinyi, H. QSAR. Hansch analysis and related approaches. In *Methods and Principles in Medicinal Chemistry, Vol 1,* Mannhold, R., Krogsgaard, L., Timmerman, H., Eds.; VCH, Weinheim, **1993**.
- 140. Kubinyi, H. The Third Dimension in QSAR: An Introduction. In *3D QSAR in Drug Design*, Kubinyi, H., Ed.; ESCOM Science Publishers B. V., Leiden, **1993**; pp 3-10.
- 141. Minkin, V. I. Glossary of terms used in theoretical organic chemistry (IUPAC recommendations 1999). *Pure Appl. Chem.* **1999**, *71* (10), 1919-1981.
- 142. Dauberosguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J.; Genest, M.; Hagler, A. T. Structure and Energetics of Ligand-Binding to Proteins *Escherichia-Coli* Dihydrofolate Reductase Trimethoprim, A Drug-Receptor System. *Proteins-Structure Function and Genetics* **1988**, *4* (1), 31-47.
- 143. Gaedt, K.; Holtje, H. D. Consistent valence force-field parameterization of bond lengths and angles with quantum chemical ab initio methods applied to some heterocyclic dopamine D-3-receptor agonists. *J. Comput. Chem.* **1998**, *19* (8), 935-946.
- 144. Clark, M.; Cramer, R. D.; Vanopdenbosch, N. Validation of the General-Purpose Tripos 5.2 Force-Field. J. Comput. Chem. **1989**, *10* (8), 982-1012.
- 145. Weiner, P. K.; Kollman, P. A. Amber Assisted Model-Building with Energy Refinement A General Program for Modeling Molecules and Their Interactions. *J. Comput. Chem.* **1981**, *2* (3), 287-303.
- 146. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A New Force-Field for Molecular Mechanical Simulation of Nucleic-Acids and Proteins. *J. Am. Chem. Soc.* **1984**, *106* (3), 765-784.
- 147. Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An All Atom Force-Field for Simulations of Proteins and Nucleic-Acids. *J. Comput. Chem.* **1986**, *7* (2), 230-252.
- 148. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. Amber, A Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comput. Phys. Commun.* **1995**, *91* (1-3), 1-41.
- 149. Halgren, T. A. The Merck Molecular-Force Field Form, Scope, Parameterization and Performance. *Abstracts of Papers of the American Chemical Society* **1992**, 204, 38-COMP.
- 150. Halgren, T. A. Merck molecular force field .1. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17* (5-6), 490-519.
- 151. Halgren, T. A. Merck molecular force field .2. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **1996**, *17* (5-6), 520-552.
- 152. Halgren, T. A. Merck molecular force field .3. Molecular geometries and vibrational frequencies for MMFF94. *J. Comput. Chem.* **1996**, *17* (5-6), 553-586.
- 153. Halgren, T. A.; Nachbar, R. B. Merck molecular force field .4. Conformational energies and geometries for MMFF94. *J. Comput. Chem.* **1996**, *17* (5-6), 587-615.
- 154. Halgren, T. A. Merck molecular force field .5. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. *J. Comput. Chem.* **1996**, *17* (5-6), 616-641.
- 155. Halgren, T. A.; Bush, B. L. The Merck molecular force field (MMFF94). Extension and application. *Abstracts of Papers of the American Chemical Society* **1996**, *212*, 2-COMP.
- 156. Cruciani, G. Molecular Interaction Fields; Cruciani, G: Weinheim, 2006; Vol. 27.
- Goodford, P. J. A Computational-Procedure for Determining Energetically Favorable Binding-Sites on Biologically Important Macromolecules. J. Med. Chem. 1985, 28 (7), 849-857.
- 158. Boobbyer, D. N. A.; Goodford, P. J.; Mcwhinnie, P. M.; Wade, R. C. New Hydrogen-Bond Potentials for Use in Determining Energetically Favorable Binding-Sites on Molecules of Known Structure. *J. Med. Chem.* **1989**, *32* (5), 1083-1094.
- 159. Wade, R. C.; Clark, K. J.; Goodford, P. J. Further Development of Hydrogen-Bond Functions for Use in Determining Energetically Favorable Binding-Sites on Molecules of Known

Structure .1. Ligand Probe Groups with the Ability to Form 2 Hydrogen-Bonds. J. Med. Chem. **1993**, 36 (1), 140-147.

- Wade, R. C.; Goodford, P. J. Further Development of Hydrogen-Bond Functions for Use in Determining Energetically Favorable Binding-Sites on Molecules of Known Structure .2. Ligand Probe Groups with the Ability to Form More Than 2 Hydrogen-Bonds. *J. Med. Chem.* **1993**, *3*6 (1), 148-156.
- 161. Goodford, P. J. The Basic Principles of GRID. In *Molecular Interaction Fields*, Cruciani, G., Ed.; Wiley-VCH, Weinheim, **2006**; pp 3-26.
- Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative Molecular-Field Analysis (Comfa) .1. Effect of Shape on Binding of Steroids to Carrier Proteins. J. Am. Chem. Soc. 1988, 110 (18), 5959-5967.
- 163. Dean, P. M. Molecular Similarity. In *3D QSAR in Drug Design*, Kubinyi, H., Ed.; ESCOM Science Publishers B. V., Leiden, **1993**; pp 150-172.
- 164. Kubinyi, H. A General View on Similarity and QSAR Studies. In *Computer-assisted lead finding and optimization*. *Current tools for medicinal chemistry*., VanDeWaterbeemd, H., Testa, B., Folkers, G., Eds.; Verlag Helvetica Chimia Acta, Basel, **1997**; pp 7-28.
- 165. Bender, A.; Glen, R. C. Molecular similarity: a key technique in molecular informatics. Organic & Biomolecular Chemistry **2004**, *2* (22), 3204-3218.
- 166. Hodgkin, E. E.; Richards, W. G. Molecular Similarity Based on Electrostatic Potential and Electric-Field. *Int. J. Quantum Chem.* **1987**, 105-110.
- 167. Academic Press dictionary of science and technology; Academic Press Inc.: San Diego, 1992.
- 168. Pastor, M.; Cruciani, G. A Novel Strategy for Improving Ligand Selectivity in Receptor-Based Drug Design. J. Med. Chem. **1995**, 38 (23), 4637-4647.
- Wold, S.; Johansson, E.; Cocchi, M. PLS Partial Least Squares Projections to Latent Structures. In *3D QSAR in Drug Design*, Kubinyi, H., Ed.; ESCOM Science Publishers B. V., Leiden, **1993**; pp 150-172.
- 170. Agnar Höskuldsson PLS regression methods. Journal of Chemometrics **1988**, 2 (3), 211-228.
- 171. Geladi, P.; Kowalski, B. R. Partial Least-Squares Regression A Tutorial. *Anal. Chim. Acta* **1986**, *185*, 1-17.
- 172. Wold, S. Validation of Qsars. *Quantitative Structure-Activity Relationships* **1991**, *10* (3), 191-193.
- 173. Gohlke, H.; Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angewandte Chemie-International Edition* **2002**, *41* (15), 2645-2676.
- 174. Voet, D.; Voet, J. G. Introducción a los enzimas. In *Bioquímica*, Editorial Omega SA, Barcelona, **1992**; pp 341-355.
- 175. Kubinyi, H. Biological Data. The activity of group contribution. In QSAR: Hansch Analysis and Related Approaches, Mannhold, R., Krogsgaard, L., Timmerman, H., Eds.; VCH Publishers, New York, **1983**.
- 176. Yokota, T.; Higuchi, K.; Kosaka, Y. Transport and Metabolism of Brassinosteroids in Rice. In *Progress in Plant Growth Regulation*, Karssen, C. M., van Loon, L. C., Vreugdenhil, D., Eds.; Kluwer, Dordrecht, **1992**; pp 298-305.
- 177. Suzuki, H.; Kim, S. K.; Takahashi, N.; Yokota, T. Metabolism of castasterone and brassinolide in mung bean explant. *Phytochemistry* **1993**, *33* (6), 1361-1367.
- Brosa, C.; Puig, E.; Vilaplana, M. Molecular modeling of a brassinosteroid analog having an adrostane ester side chain. XVIII Conference on Isoprenoids, Prahatice, Czech Republic, 1999.
- Brosa, C.; Kohout, L.; Vilaplana, M.; Piqué, M. New brassinosteroid derivates having an amine in the androstane ester side chain. XIX Conference on Isoprenoids, Jurata, Poland, 2001.
- Brosa, C.; Kohout, L.; Piqué, M.; Vilaplana, M. New brassinosteroid derivates having an hydroxil in the androstane ester side chain. XIX Conference on Isoprenoids, Jurata, Poland, 2001.
- 181. Hayashi, S.; Hohjoj, T.; Shida, A.; Ikekawa, N. Preparation of 23-phenylbrassinosteroids as plant growth regulators. *Patent*, EP 282,984, **1987**.

- 182. Huang, L. F.; Zhou, W. S. Novel method for construction of the side-chain of 23arylbrassinosteroids via Heck arilation and asymetric dihydroxilation as key steps. *J. Chem. Soc.*, *Perkin Trans.* 1 **1994**, 3579-3585.
- 183. Puig, E.; Brosa, C. Síntesi i avaluació de nous anàlegs brassinoesteroides androstànics esterificats amb l'acid α-hidroxivalèric. *Comunicació personal*, **2000**.
- 184. Piqué, M.; Brosa, C. Comunicació personal, 2001.
- 185. Tobella, L.; Brosa, C. Comunicació personal, 2003.
- 186. Strnad, M.; Kohout, L. A simple brassinolide analogue 2α, 3α-dihydroxy-17β-(3-methylbutyryloxy)-7-oxa-B-homo-5α-androstan-6-one which induces bean second internode splitting. *Plant Growth Regulation* **2003**, 40 (1), 39-47.
- Sisa, M.; Hnilickova, J.; Swaczynova, J.; Kohout, L. Syntheses of new androstane brassinosteroids with 17β-ester groups-butyrates, heptafluorobutyrates and laurates. *Steroids* 2005, 70 (11), 755-762.
- 188. Chodounska, H.; Pouzar, V.; Budesinsky, M.; Slavikova, B.; Kohout, L. Synthesis of 3methyl-3-hydroxy-6-oxo-androstane derivatives. *Steroids* **2004**, *69* (10), 605-612.
- 189. Brosa, C.; Tobella, L.; Kohout, L.; Rayó, P.; Vilaplana, M. The effect on activity of a nitrogenated functionality in androstane brassinosteroid analogs. XX Conference on Isoprenoids, Liberec, Czech Republic, **2003**.
- 190. Wilson, S. R.; Cui, W.; Moskowitz, J. W.; Schmidt, K. E. Applications of Simulated Annealing to the Conformational-Analysis of Flexible Molecules. *J. Comput. Chem.* **1991**, *12* (3), 342-349.
- 191. Wilson, S. R.; Cui, W.; Guarnieri, F. Conformational analysis of flexible molecules. *Data Handling in Science and Technology* **1995**, *15*, 351-367.
- 192. Bohachevsky, I. O.; Johnson, M. E.; Stein, M. L. Simulated annealing and generalizations. *Data Handling in Science and Technology* **1995**, *15*, 3-24.
- 193. Smellie, A.; Kahn, S. D.; Teig, S. L. Analysis of Conformational Coverage .1. Validation and Estimation of Coverage. J. Chem. Inf. Comput. Sci. **1995**, 35 (2), 285-294.
- 194. Smellie, A.; Kahn, S. D.; Teig, S. L. Analysis of Conformational Coverage .2. Application of Conformational Models. *J. Chem. Inf. Comput. Sci.* **1995**, *35* (2), 295-304.
- 195. Kutschabsky, L.; Adam, G.; Vorbrodt, H. M. Molecular and Crystal-Structure of (22S,23S)-Homobrassinolide. *Zeitschrift fur Chemie* **1990**, *30* (4), 136-137.
- 196. Vilaplana-Polo, M.; Zamora, I.; Brosa, C. Brassinosteroids: Trying to find a better way to select the active conformation to establish a QSAR. XX Conference on Isoprenoids, Liberec, Czech Republic, **2003**.
- 197. Leach, A. R. A Survey of Methods for Searching the Conformational Space of Small an Medium-Sized Molecules. In *Reviews of Computational Chemistry II*, Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers Inc., New York, **1991**; p 42.
- 198. Zamora, I. Comunicació personal, Sant Cugat del Vallés, 2003.
- 199. Ramirez, J. A.; Centurion, O. M. T.; Gros, E. G.; Galagovsky, L. R. Synthesis and bioactivity evaluation of brassinosteroid analogs. *Steroids* **2000**, *65* (6), 329-337.
- 200. Vilaplana-Polo, M.; Brosa, C. Study on activity data in QSAR models. Application on Brassinosteroids field. XXI Conference on Isoprenoids, Bialowieza, Poland, **2005**.
- 201. Modolell, A.; Brosa, C. Síntesi d'anàlegs brassinoesteroides amb modificacions a l'anell A i B. Augment de la sensibilitat del test d'inclinació de la làmina d'arròs amb planta sencera. *Comunicació personal,* Institut Químic de Sarrià C.E.T.S., Barcelona, **1998**.
- 202. Puig, E.; Brosa, C. Síntesi i avaluació de nous anàlegs brassinoesteroides androstànics esterificats amb l'acid α-hidroxivalèric. *Comunicació personal*, Institut Químic de Sarrià C.E.T.S., Barcelona, **2000**.
- 203. Vázquez, E.; Brosa, C. Aproximación a la síntesis de análogos brasinoesteroides fluorados en el anillo A, obtención de un nuevo análogo activo y estudios de sinergismo e inhibición. *Comunicació personal*, Institut Químic de Sarrià C.E.T.S., Barcelona, **2000**.
- 204. Piqué, M.; Brosa, C. Síntesi de l'anàleg  $2\alpha$ , $3\alpha$ -dihidroxi-17 $\beta$ -(2-acetoxi-3-metilbutiroxi)-6,7seco- $5\alpha$ -androstà-6,7-lactona: una dada a favor de la importància de la presència del grup hidroxil en el C<sub>23</sub> de la brassinolida. *Comunicació personal*, Institut Químic de Sarrià C.E.T.S., Barcelona, **2001**.

- 205. Tobella, L.; Brosa, C. Síntesi de derivats brassinoesteroides androstànics que contenen un grup hidroxil o amino en la cadena lateral. *Comunicació personal*, Institut Químic de Sarrià C.E.T.S., Barcelona, **2003**.
- 206. Fig P 2.2. for Windows, Biosoft, 49 Bateman Street, Cambridge, CB2 1LR, U.K. 1993.
- 207. Espelta, S.; Brosa, C. Comunicació personal, Barcelona, 2004.
- 208. Meudt, W. J.; Thompson, M. J. Investigations on the mechanism of the brassinosteroid response. II. A modulation of auxin action. *Proceedings Plant Growth Regulation Society of America* **1983**, *10th*, 306-311.
- 209. Yu, W.; Zhao, P.; Wang, Q. Interaction between brassinosteroid and auxin. *Xibao Shengwuxue Zazhi* **2005**, *27* (2), 165-167.
- 210. Vilaplana, M.; Brosa, C. Tècniques computacionals en brassinoesteroides: Estudi de la influència de l'alineament en la consecució d'un QSAR i disseny de nous anàlegs. *Comunicació personal*, Institut Químic de Sarrià C.E.T.S., Barcelona, **2001**.
- 211. Brosa, C.; Soca, L.; Terricabras, E.; Ferrer, J. C.; Alsina, A. New synthetic brassinosteroids: a 5α-hydroxy-6-ketone analog with strong plant growth promoting activity. *Tetrahedron* **1998**, 54 (40), 12337-12348.
- 212. Brosa, C. Comunicació personal, Barcelona, 2007.
- He, Z.; Wang, Z. Y.; Li, J.; Zhu, Q.; Lamb, C.; Ronald, P.; Chory, J. Perception of brassinosteroids by the extracellular domain of the receptor kinase BRI1. *Science* (*Washington, D. C.*) 2000, 288 (5475), 2360-2363.
- 214. Li, J.; Wen, J. Q.; Lease, K. A.; Doke, J. T.; Tax, F. E.; Walker, J. C. BAK1, an Arabidopsis LRR receptor-like protein kinase, interacts with BRI1 and modulates brassinosteroid signaling. *Cell* **2002**, *110* (2), 213-222.
- 215. Kinoshita, T.; Cano-Delgado, A.; Seto, H.; Hiranuma, S.; Fujioka, S.; Yoshida, S.; Chory, J. Binding of brassinosteroids to the extracellular domain of plant receptor kinase BRI1. *Nature* (London, United Kingdom) **2005**, 433 (7022), 167-171.
- 216. Seto, H.; Kinoshita, T. Binding mechanism of brassinosteroid, plant growth hormone, and receptor BRI 1. *BRAIN Techno News* **2005**, *109*, 17-22.
- 217. Wang, X.; Li, X.; Meisenhelder, J.; Hunter, T.; Yoshida, S.; Asami, T.; Chory, J. Autoregulation and homodimerization are involved in the activation of the plant steroid receptor BRI1. *Developmental Cell* **2005**, *8* (6), 855-865.
- 218. Wang, Z. Y.; Wang, Q. M.; Chong, K.; Wang, F. R.; Wang, L.; Bai, M. Y.; Jia, C. G. The brassinosteroid signal transduction pathway. *Cell Research* **2006**, *1*6 (5), 427-434.
- 219. Li, J. M.; Jin, H. Regulation of brassinosteroid signaling. *Trends in Plant Science* **2007**, *12* (1), 37-41.
- 220. Hyperchem, Autodesk Inc., Avenue des Champs-Montants 14B, CH-2074 Marin, Switzerland. **1992**.
- 221. Insight 97.0, Biosym/MSI, 9685 Scranton Road, San Diego, United States. 1997.
- 222. SYBYL 6.0, Tripos Assoc. Inc., St. Louis, Missouri, United States. 1993.
- 223. Molecular Operating Enviroment (MOE), Chemical Computing Group, Kaiser-Wilhelm-Ring 11, 50672 Köln, Germany. **2002**.
- 224. Wade, R. C.; Clark, K. J.; Goodford, P. J. Further Development of Hydrogen-Bond Functions for Use in Determining Energetically Favorable Binding-Sites on Molecules of Known Structure .1. Ligand Probe Groups with the Ability to Form 2 Hydrogen-Bonds. *Journal of Medicinal Chemistry* **1993**, *36* (1), 140-147.
- 225. Wade, R. C.; Goodford, P. J. Further Development of Hydrogen-Bond Functions for Use in Determining Energetically Favorable Binding-Sites on Molecules of Known Structure .2. Ligand Probe Groups with the Ability to Form More Than 2 Hydrogen-Bonds. *Journal of Medicinal Chemistry* **1993**, *36* (1), 148-156.
- 226. *MIPSIM 1.1,* Cáceres, M.; Villa, J.; Lozano, J.; Sanz, F. Grup de Recerca en Informatica Medica, Institut Municipal d'Investigació Mèdica (IMIM), Doctor Aiguader 80, 08003 Barcelona, Spain. **1999**.
- 227. Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. Generating Optimal Linear Pls Estimations (Golpe) - An Advanced Chemometric Tool for Handling 3D-Qsar Problems. *Quantitative Structure-Activity Relationships* **1993**, *12* (1), 9-20.
- 228. GOLPE 4.5., Multivariate Infometric Analysis Srl., Viale dei Castagni 16, Perugia, Italy. 1999.

- 229. Tools for Structure-Activity Relationship (T.S.A.R.), Oxford Molecular, Magdalen Centre, Oxford, U.K. **1997**.
- 230. *Microsoft Excel,* Microsoft, **2007**.
- 231. Spendley, W.; Hext, G. R.; Himsworth, F. R. Sequential Application of Simplex Designs in Optimisation and Evolutionary Operation. *Technometrics* **1962**, *4* (4), 441.
- 232. Pastor, M.; Cruciani, G.; Clementi, S. Smart region definition: A new way to improve the predictive ability and interpretability of three-dimensional quantitative structure-activity relationships. *J. Med. Chem.* **1997**, *40* (10), 1455-1464.
- 233. Cruciani, C.; Pastor, M.; Clementi, S. Region Selection in 3D-QSAR. In *Computer-assisted lead finding and optimization*. *Current tools for medicinal chemistry.*, VanDeWaterbeemd, H., Testa, B., Folkers, G., Eds.; Verlag Helvetica Chimia Acta, Basel, **1997**; pp 379-396.
- 234. Cruciani, G.; Clementi, S.; Pastor, M. GOLPE-guided region selection. *Perspectives in Drug Discovery and Design* **1998**, *12*, 71-86.
- 235. Prueba de hipótesis. In *Probabilidad y estadística aplicadas a la ingenieria*, Montgomery, D. C., Runger, G. C., Eds.; McGraw-Hill, Mexico D.F., **1996**; p 370.