

A mis padres

A mi hermana Julia

A Farmor

A Marc

“Se debe hacer todo tan sencillo como sea posible, pero no más sencillo”

Albert Einstein

Agradecimientos.

En primer lugar quiero agradecer al Dr. Teixidó, director de esta tesis, por darme la oportunidad de llevar a cabo este proyecto, y además de su ayuda y consejos tanto en el campo de la química computacional como en la síntesis.

Al IQS por la beca que me ha permitido realizar la tesis.

Al grupo del Dr. Esté, en especial a Imma, por la determinación de la actividad antiviral y el modo de acción de mis moléculas, y por explicarme cómo interpretar los resultados.

También quiero agradecer al Dr. Borrell su interés y colaboración, especialmente en las últimas etapas del proyecto.

Al Servicio de Espectroscopia, al Dr. Batllori por su ayuda en la interpretación de algún RMN complicado, a Nuria, y a Mari Carmen por su inestimable trabajo.

Sobre todo quisiera agradecer a todos los compañeros del Laboratorio de Síntesis y Diseño Molecular su ayuda y apoyo a lo largo de estos años. A Alberto, que empezó con la parte sintética de este proyecto y me ayudó durante los primeros meses. Obdulía y Violeta por su paciencia y gran ayuda en los temas computacionales. A Miriam por seguir donde yo lo dejé. A Rai por su ayuda y consejos. A Laura, Sonia, Anna, Maia y Gemma, con las que he compartido muchas comidas, cenas, risas y cotilleos. A mis compañeros del laboratorio "nuevo", Juan, Lyhen y Laia. A los TKIs, Inyaki y Xavi. A al resto de compañeros con los que he coincidido durante estos años, Adaya, Ángeles, Astrid, David, Gonzalo, Ignasi, Irene, Jordi, Laia, Llorenç, Manel, Mares, María, Marta B., Marta C., Montse, Núria, Ofir, Roger, Rubén y Xevi.

A mis amigos. En especial a Eli, con la que he compartido muchas vivencias desde hace muchos años, y a Carles. A Aroha, ya hace muchos años que nos conocemos, y a David. A Berta y Aimar, cuántas cosas han pasado desde Erlangen. A Anthony y Cristina. Y a todos los demás.

A mi familia española, a mi familia sueca, a mi familia finlandesa y a mi familia "política". A Anika, a José Luís y a María Jesús.

A Marc, por su confianza, por tener tanta paciencia conmigo y por todos los momentos que hemos vivido juntos a lo largo de estos años.

A mis padres, que me han apoyado y me han ayudado siempre, y a mi hermana Julia.

Abreviaciones.

3TC	lamivudina
ABC	abacavir
Ac	acetilo
ACN	acetonitrilo
ADME	absorción, distribución, metabolismo y excreción
AEO	análisis elemental orgánico
APV	amprenavir
ATV	atazanavir
AZT	azidotimidina
Boc	<i>tert</i> -butoxicarbonilo
BTAC	cloruro de benziltriethylamonio
CA	cápside
CC ₅₀	concentración citotóxica 50
CCF	cromatografía de capa fina
CCR	receptor de quimiocinas de la clase CC
CDC	<i>Centres for Disease Control and Prevention</i>
CoMFA	<i>comparative molecular field analysis</i>
CoMSIA	<i>comparative molecular similarity indices analysis</i>
CXCR	receptor de quimiocinas de la clase CXC
d	doblete
δ	desplazamiento químico
d4T	stavudina
DCE	dicloroetano
ddA	dideoxiadenosina
ddC	dideoxicitidina
ddl	dideoxinosina
Dep	dietoxifosforilo
DIBAH	hidruro de diisobutilaluminio
DLV	delavirdina
DMSO	dimetilsulfóxido
DNA	ácido desoxirribonucleico
DRV	darunavir
DS	sulfato de dextrano
EC ₅₀	concentración efectiva 50
EFV	efavirenz
ESI	ionización por electrospray
Et	etilo
<i>f</i>	flexión

FDA	<i>Food and Drug Administration</i>
FOS	fosamprenavir
FTC	emtricitabina
gdl	grados de libertad
gp	glicoproteína
GP	grupo protector
GPCR	receptor celular asociado a proteína G
GRID	<i>gay-related immune deficiency</i>
HAART	<i>highly active antiretroviral therapy</i>
HEPT	1-[(2-hidroxietoxi)metil]-6-(tiofenil)timina
HRC	<i>clusters</i> jerárquicos
HRMS	espectroscopía de masas de alta resolución
HTLV	virus T-linfotrópico humano
HTS	<i>high-throughput screening</i>
IC ₅₀	concentración de inhibición 50
IDV	indinavir
IE	impacto electrónico
IN	integrasa
iPr	isopropilo
IR	espectroscopía de infrarrojo
LAV	virus asociado a linfadenopatía
logP	coeficiente de partición octanol/agua
LOO	<i>leave one out</i>
LTR	<i>long terminal repeats</i>
LVP	lopinavir
m	multiplete
MA	matriz
Me	metilo
MS	espectroscopía de masas de baja resolución
MTT	bromuro de 3-[4,5-dimetiltiazol-2-il]-2,4-difeniltetrazolio
MW	microondas
v	número de onda
NBS	<i>N</i> -bromosuccinimida
NC	nucleocápside
NCI	<i>National Cancer Institute</i>
NFV	nelfinavir
NNRTI	inhibidor de transcriptasa inversa no análogo de nucleósido
NRTI	inhibidor de transcriptasa inversa análogo de nucleósido
NVP	nevirapina
PCA	análisis de componentes principales

PEOE	<i>partial equalization of orbital electronegativities</i>
Ph	fenilo
PI	inhibidor de proteasa
PLS	mínimos cuadrados parciales
<i>pol</i>	<i>polymerase gene</i>
ppm	partes por millón
PR	proteasa
q	cuarteto
QSAR	<i>quantitative structure-activity relationship</i>
QSPR	<i>quantitative structure-property relationship</i>
quint	quinteto
RMN	resonancia magnética nuclear
RNA	ácido ribonucleico
RT	transcriptasa inversa
RTV	ritonavir
s	singlete
sa	singlete ancho
SAR	<i>structure-activity relationship</i>
sc	señal compleja
SDF-1	<i>stromal cell-derived factor-1</i>
SQV	saquinavir
SU	superficie
<i>t</i>	tensión
t	triplete
TDF	tenofovir disoproxil fumarato
TFA	ácido trifluoroacético
THF	tetrahidrofurano
TM	transmembrana
TMS	tetrametilsilano
TPV	tipranavir
Troc	2,2,2-tricloroetoxicarbonilo
Ts	tosilo
TSPNa	(trimetilsilil)propionato sódico
VIH	virus de la inmunodeficiencia humana

Sumario.

El sida es una enfermedad causada por el virus de la inmunodeficiencia humana (VIH), que ataca a las células del sistema inmunitario. Actualmente se estima que viven unos 33 millones de personas infectadas por este virus.

La terapia antirretroviral actual consiste en combinaciones de dos familias de compuestos: los inhibidores de la transcriptasa inversa y los inhibidores de la proteasa, ambos dirigidos a enzimas específicos producidos por el VIH. Sin embargo, una de las principales dificultades en el tratamiento antirretroviral es la rápida evolución del virus, que le permite hacerse resistente a los fármacos utilizados. Por lo tanto, hay una continua necesidad de usar nuevos fármacos anti-VIH en combinación con los actuales, que deberían actuar preferentemente en nuevas dianas del ciclo de replicación del VIH.

El descubrimiento de los cofactores celulares implicados en la entrada del VIH y una mejor comprensión de las proteínas de la envoltura que controlan el proceso de fusión han renovado el interés en las primeras etapas de la replicación del VIH como dianas para una intervención terapéutica. La entrada del VIH consta de 4 etapas interrelacionadas: la unión del virus a la superficie celular, la unión del virus al receptor CD4, la interacción del complejo CD4-glicoproteína de la envoltura con los correceptores de entrada y la fusión virus-célula. Los principales correceptores utilizados por el VIH para su entrada en la célula son los receptores de quimiocinas CCR5 y CXCR4. En este ámbito se han aprobado recientemente dos nuevos fármacos, Fuzeon® (Hoffman-La Roche y Trimeris), que inhibe la fusión del virus con la célula, y Selzentry® (Pfizer), que es un antagonista de CCR5. Se han identificado agentes que bloquean CXCR4, entre los que destaca AMD3100. Sin embargo, su desarrollo se ha abandonado por una toxicidad cardíaca y falta de biodisponibilidad oral, ambas relacionadas con su elevada carga positiva a pH fisiológico. Por lo tanto se necesita disponer de nuevos inhibidores de la unión del VIH al correceptor CXCR4 que solventen estas desventajas.

El interés de la presente tesis es diseñar una quimioteca combinatoria de análogos al AMD3100 con menor basicidad, que incorporen en su estructura las principales características de este compuesto, un *linker* *p*-fenilbismetilénico y sistemas heterocíclicos. A partir de la estructura de los compuestos de la quimioteca diseñada, se desarrollan y validan las rutas sintéticas necesarias para su obtención. Además, se aplican técnicas computacionales de selección de diversidad para explorar el espacio químico representado por los compuestos de la quimioteca. A partir de los compuestos activos identificados en este proceso e inhibidores conocidos del correceptor CXCR4 se desarrollan modelos QSAR y farmacofóricos. Estos modelos se utilizan para realizar un cribado virtual del resto de compuestos de la quimioteca diseñada y aumentar así la probabilidad de seleccionar nuevos compuestos con actividad antiviral. Mediante este procedimiento, se han identificado nuevos inhibidores de CXCR4, uno de ellos con elevada actividad anti-VIH. Además, este compuesto sólo presenta cuatro átomos de nitrógeno en su estructura frente a los ocho del AMD3100.

Sumari.

La sida és una malaltia causada pel virus de la immunodeficiència humana (VIH), que ataca a les cèl·lules del sistema immunitari. Actualment s'estima que viuen uns 33 milions de persones infectades per aquest virus.

La teràpia antiretroviral actual consisteix en combinacions de dues famílies de compostos: inhibidors de la transcriptasa inversa i inhibidors de la proteasa, ambdós dirigits a enzims específics produïts pel VIH. Una de les principals dificultats en el tractament antiretroviral és la ràpida evolució del virus, que li permet fer-se resistent als fàrmacs que s'utilitzen. Per tant, hi ha una necessitat contínua d'utilitzar nous fàrmacs anti-VIH en combinació amb els actuals, i haurien d'actuar preferentment en noves dianes del cicle de replicació del VIH.

El descobriment dels cofactors cel·lulars implicats en l'entrada del VIH i una millor comprensió de les proteïnes de l'envolta que controlen el procés de fusió han renovat l'interès en les primeres etapes de la replicació del VIH com a dianes per a una intervenció terapèutica. L'entrada del VIH consta de 4 etapes interrelacionades: la unió del virus a la superfície cel·lular, la unió del virus al receptor CD4, la interacció del complex CD4-glicoproteïna de l'envolta amb el coreceptors d'entrada i la fusió virus-cèl·lula. Els principals coreceptors utilitzats pel VIH per a la seva entrada a la cèl·lula són els receptors de quimiocines CCR5 i CXCR4. En aquest àmbit s'han aprovat recentment dos nous fàrmacs, Fuzeon® (Hoffman-La Roche i Trimeris), que inhibeix la fusió del virus amb la cèl·lula, i Selzentry® (Pfizer), que és un antagonista de CCR5. S'han identificat agents que bloquegen CXCR4, entre els que destaca AMD3100. Tot i això, el seu desenvolupament s'ha abandonat per una toxicitat cardíaca i falta de biodisponibilitat oral, ambdues relacionades amb la seva elevada càrrega positiva a pH fisiològic. Per tant, es necessita disposar de nous inhibidors de la unió del VIH al coreceptor CXCR4 que solucionin aquests desavantatges.

L'interès d'aquesta tesi és dissenyar una quimioteca combinatòria d'anàlegs al AMD3100 amb menor basicitat, que incorporin a la seva estructura les principals característiques d'aquest compost, un *linker p*-fenilenbismetilènic i sistemes heterocíclics. A partir de l'estructura dels compostos de la quimioteca dissenyada es desenvolupen i validen les rutes sintètiques necessàries per a la seva obtenció. A més, s'apliquen tècniques computacionals de selecció de diversitat per a explorar l'espai químic representat pels compostos de la quimioteca. A partir de les molècules actives identificades en aquest procés i inhibidors coneguts del coreceptor CXCR4 es desenvolupen models QSAR i farmacofòrics. Aquests models s'utilitzen per a realitzar un cribatge virtual de la resta de compostos de la quimioteca dissenyada i augmentar la probabilitat de seleccionar nous compostos amb activitat antiviral. Mitjançant aquest procediment s'han identificat nous inhibidors de CXCR4, un d'ells amb elevada activitat anti-VIH. A més, aquest compost només presenta quatre àtoms de nitrogen a la seva estructura, enfront dels vuit de l'AMD3100.

Summary.

AIDS is a disease of the immune system caused by the human immunodeficiency virus (HIV). Today, about 33 million people live with this virus.

Current antiretroviral therapies consist of combinations of two compound families: reverse transcriptase inhibitors and protease inhibitors, both directed to specific enzymes produced by the HIV. One of the major drawbacks in this antiretroviral treatment is drug resistance, produced by the rapid evolution of the virus. Hence there is a need to develop new anti-HIV drugs which should preferably act on new stages in HIV's life cycle.

The discovery of cellular cofactors involved in the entry of HIV to host cells and a better comprehension of envelope proteins controlling the fusion process have renewed the interest in early steps of HIV replication as targets for therapeutic intervention. HIV entry to the cell can be dissected into 4 interrelated steps: virus attachment to the cell surface, virus binding to the CD4 receptor, interaction of the CD4-envelope glycoprotein complex with entry coreceptors and virus-cell fusion. The main HIV entry coreceptors are chemokine receptors CCR5 and CXCR4. In this context, two new drugs have recently been approved, Fuzeon® (Hoffman-La Roche and Trimeris), that blocks fusion of HIV with the cell, and Selzentry® (Pfizer), a CCR5 antagonist. Agents that block CXCR4 have also been identified, with AMD3100 as lead compound. However, the development of this compound has been discontinued due to cardiac toxicity and lack of oral bioavailability related to its high positive charge at physiological pH. Therefore, there is a need to develop new CXCR4 coreceptor inhibitors that will solve these limitations.

The aim of this thesis is the design of a combinatorial library of AMD3100 analogues with lower basicity while preserving the main structural features of AMD3100, a *p*-phenylenic linker and heterocyclic systems. Synthetic routes for these compounds have been developed and validated. Moreover, computational diversity selection techniques have been applied to explore the chemical space represented by the library. Active compounds identified by this process and known CXCR4 inhibitors have been used in QSAR and pharmacophore modelling. Then, virtual screening of the rest of compounds in the library with these models has been applied to increase the probability of selecting new active compounds. This has led to the identification of new CXCR4 inhibitors, one of them showing high anti-HIV activity. Furthermore, the structure of this compound only includes four nitrogen atoms, in contrast to the eight in AMD3100.

Índice.

INTRODUCCIÓN	1
0.1. El sida.....	3
0.1.1. Historia del sida.....	3
0.1.2. La enfermedad.....	3
0.2. El VIH.....	6
0.2.1. Clasificación, estructura y genoma.....	6
0.2.2. Ciclo de vida del VIH.....	8
0.2.3. Etapa de entrada y fusión.....	10
0.2.4. Los correceptores CXCR4 y CCR5. Tropismo del VIH.....	11
0.2.5. Variabilidad genética.....	13
0.3. Terapia antirretroviral.....	15
0.3.1. Terapia antirretroviral actual.....	15
0.3.2. Fármacos en desarrollo.....	20
0.4. Los biciclamos.....	23
0.4.1. Estructura general de los biciclamos y actividad anti-VIH.....	23
0.4.2. Desarrollo de monociclamos.....	25
0.4.3. Modo de acción del AMD3100.....	27
0.5. Investigación sobre el sida en el IQS.....	28
0.6. Objetivos.....	32
Bibliografía.....	33
CAPÍTULO 1. ANÁLOGOS POLINITROGENADOS NO CICLÁMICOS DEL AMD3100	43
1.1. Diseño de una quimioteca de análogos polinitrogenados del AMD3100.....	45
1.2. Selección de compuestos de la quimioteca virtual con PRALINS.....	51
1.2.1. Descriptores moleculares.....	51
1.2.2. Análisis de componentes principales (PCA).....	52
1.2.3. Selección de compuestos con PRALINS.....	52
1.3. Predicción de propiedades ADME-Tox.....	57
1.3.1. Introducción.....	57
1.3.2. Métodos utilizados para la predicción de propiedades ADME-Tox.....	58
1.3.3. Resultados de la predicción de propiedades ADME-Tox.....	59
1.4. Metodología sintética.....	63
1.4.1. Aminación reductora.....	63
1.4.2. Condensación de aldehídos o cetonas con hidrazinas.....	66
1.4.3. Estrategia de obtención de los compuestos de la quimioteca diseñada.....	66
1.4.3.1. Estrategia de obtención de diaminas simétricas 32 {x,x}.....	66
1.4.3.2. Estrategia de obtención de dihidrazonas simétricas 33 {x,x}.....	67

1.4.3.3. Estrategia de obtención de diaminas no simétricas 32 {x,y}.....	68
1.4.3.4. Estrategia de obtención de dihidrazonas no simétricas 33 {x,y}.....	71
1.4.3.5. Estrategia de obtención de aminohidrazonas 37 {x,y}.	72
1.5. Evaluación de la actividad anti-VIH y del modo de acción de un fármaco.....	73
1.6. Resultados y discusión.	75
1.6.1. Síntesis y actividad anti-VIH.	75
1.6.2. Ampliación de la quimioteca inicial a partir del compuesto más activo 32 {8,8}...	81
1.6.3. Modo de acción.	85
Bibliografía.	88
CAPÍTULO 2. MONOCICLAMOS SUSTITUIDOS	93
2.1. Introducción a los monociclamos sustituidos.	95
2.2. Diseño y análisis retrosintético de monociclamos sustituidos.....	101
2.3. Protección N-1 del ciclamo.	104
2.3.1. Formación de complejos metálicos.....	104
2.3.2. Formación de carbamatos.	105
2.3.3. Protección con derivados de fósforo.....	106
2.3.4. Protección con amidas.....	107
2.3.5. Síntesis del tri- <i>tert</i> -butilcarbamato del ciclamo 66.	109
2.4. Vía sintética 1.	111
2.4.1. Etapa 1: síntesis de 1-[4-(bromometil)fenilmetil]-4,8,11-tris(<i>tert</i> -butoxicarbonil)- 1,4,8,11-tetraazaciclotetradecano (67).....	111
2.4.2. Etapa 2: síntesis de monociclamos sustituidos 87 a partir del derivado ciclámico bromado 76	112
2.5. Vía sintética 2.	114
2.5.1. Síntesis de 4-(bromometil)benzaldehído (74).....	114
2.5.2. Etapa 1: síntesis de 1-[4-(carboxaldehído)fenilmetil]-4,8,11-tris(<i>tert</i> - butoxicarbonil)-1,4,8,11-tetraazaciclotetradecano (90).....	117
2.5.3. Etapa 2: síntesis de monociclamos sustituidos con poliaminas 87 a partir del derivado ciclámico 90	118
2.5.4. Desprotección de los grupos Boc de los monociclamos sustituidos 87	121
2.6. Resultados y discusión.	122
Bibliografía.	127
CAPÍTULO 3. QSAR	131
3.1. QSAR (quantitative structure-activity relationship).....	133
3.2. Parámetros utilizados en el análisis de los modelos QSAR.....	136
3.3. Selección de <i>training set</i> y <i>test set</i>	139
3.4. Alineamiento.	140

3.5. Selección de descriptores.....	144
3.6. Modelos QSAR: ecuación, validación y parámetros estadísticos.....	146
3.6.1. Modelos QSAR de 3 descriptores.....	146
3.6.1.1. Modelo 1.....	146
3.6.1.2. Modelo 2.....	147
3.6.1.3. Modelo 3.....	149
3.6.2. Modelos QSAR de 4 descriptores.....	151
3.6.2.1. Modelo 4.....	151
3.6.2.2. Modelo 5.....	152
3.6.2.3. Modelo 6.....	153
3.6.2.4. Modelo 7.....	155
3.6.2.5. Modelo 8.....	157
3.6.2.6. Modelo 9.....	158
3.6.2.7. Modelo 10.....	160
3.6.2.8. Modelo 11.....	161
3.6.2.9. Modelo 12.....	162
3.6.2.10. Modelo 13.....	164
3.6.3. Modelo QSAR de 5 descriptores.....	166
3.6.3.1. Modelo 14.....	166
3.6.4. Comparación de los modelos QSAR obtenidos.....	167
3.7. Árbol de clasificación.....	170
3.7.1. Introducción.....	170
3.7.2. Análisis de los árboles de clasificación.....	171
3.7.3. Selección de <i>training set</i> y <i>test set</i>	173
3.7.4. Modelos de clasificación.....	174
3.8. Predicciones.....	178
Bibliografía.....	181
CAPÍTULO 4. FARMACÓFOROS.....	185
4.1. Introducción a los farmacóforos.....	187
4.1.1. Cribado virtual.....	187
4.1.2. Desarrollo de farmacóforos con MOE.....	189
4.2. Parámetros de análisis de los farmacóforos.....	192
4.3. Base de datos.....	195
4.3.1. Base de datos de activos.....	195
4.3.2. Base de datos de inactivos.....	197
4.3.2.1. Inactivos seleccionados de la base de datos <i>Maybridge Screening</i> <i>Collection</i>	197
4.3.2.2. Inactivos reales.....	199
4.4. Alineamiento.....	200

4.5. Cálculo de conformaciones.	201
4.6. Modelos farmacofóricos: análisis retrospectivo.....	202
4.6.1. Modelo 1.....	202
4.6.2. Modelos 2 y 3.....	206
4.6.3. Modelos 4 y 5.....	209
4.6.4. Modelos 6, 7 y 8.....	212
4.6.5. Modelo 9.....	216
4.6.6. Comparación de los modelos farmacofóricos.....	218
4.6.7. Modelo 10: filtrado mediante comparación del <i>fingerprint</i> de MACCS.....	222
4.6.7.1. Introducción.....	222
4.6.7.2. Resultados.....	222
4.7. Modelos farmacofóricos: análisis prospectivo.....	225
Bibliografía.....	229
PARTE EXPERIMENTAL.....	233
CONCLUSIONES.....	323
ANEXOS.....	329
A.1. Descriptores moleculares.....	331
A.2. <i>Training set</i> utilizado en los QSAR.....	337
A.3. <i>Test set</i> utilizado en los QSAR.....	343
A.4. <i>Training set</i> utilizado en los árboles de clasificación.....	345
A.5. <i>Test set</i> utilizado en los árboles de clasificación.....	353
A.6. Base de datos de activos utilizada en los farmacóforos.....	355
A.7. Base de datos de inactivos reales utilizada en los farmacóforos.....	375
A.8. Publicaciones.....	379

INTRODUCCIÓN

0.1. El sida.

0.1.1. Historia del sida.

El sida es una enfermedad que se define por una disminución severa de células T causada por el virus de la inmunodeficiencia humana (VIH). Los primeros informes sobre la enfermedad aparecieron en 1981, aunque datos que se conocen actualmente sugieren que el sida ha existido durante varias décadas. En junio de 1981 los *Centres for Disease Control and Prevention* (CDC) norteamericanos informaron sobre varios casos de neumonía por *Pneumocystis carinii* y también de sarcoma de Kaposi, un tipo de cáncer de piel, en hombres homosexuales previamente sanos de California y Nueva York.^[1] Poco después se dieron cuenta de que la enfermedad también afectaba a otros grupos de la población, como usuarios de drogas intravenosas, haitianos y hemofílicos.^[2,3]

Inicialmente la enfermedad recibió el nombre *GRID* (*gay-related immune deficiency*), y no fue hasta septiembre de 1982 cuando recibió su nombre actual: sida, acrónimo de síndrome de la inmunodeficiencia adquirida. Uno de los síntomas comunes que padecían los portadores de esta nueva enfermedad era un rápido decrecimiento del número de células T CD4⁺, que les hacía vulnerables a infecciones oportunistas.

Este mismo grupo de células era la diana del primer retrovirus humano patógeno descrito, HTLV-I (virus linfotrópico humano tipo-I), aislado en 1980 por Robert Gallo y colaboradores,^[4] en el *National Cancer Institute* (NCI). Este grupo posteriormente aisló el HTLV-III.^[5-7] Luc Montagnier, Françoise Barré-Sinoussi y sus colaboradores en el Instituto Pasteur, identificaron un nuevo retrovirus humano, aparentemente diferente de HTLV-I y lo llamaron LAV (virus asociado a linfadenopatía).^[8] No fue hasta 1986 cuando el Comité Internacional de Taxonomía de Virus recomendó llamar a este patógeno con varios nombres, virus de la inmunodeficiencia humana (VIH), identificado como el agente etiológico del sida.^[9] En el año 2008 los científicos franceses Montagnier y Barré-Sinoussi recibieron el Premio Nobel de Medicina por el descubrimiento del virus VIH.^[10,11]

0.1.2. La enfermedad.

Las células diana del VIH son las que presentan el receptor CD4 en su superficie, son componentes críticos del sistema inmunitario, engloba a linfocitos T, monocitos, macrófagos y células dendríticas. Por tanto, la infección por VIH causa un deterioramiento progresivo del sistema inmunitario.

Después de la exposición al virus, el VIH entra en el torrente sanguíneo, y tiene lugar la infección. La infección primaria o aguda es la primera etapa de la infección causada por el virus. Generalmente dura una o dos semanas, que es cuando el virus se establece en el cuerpo del huésped. En esta etapa se pueden tener síntomas parecidos a los de la gripe, como

fiebre, dolores musculares y articulares y ganglios linfáticos inflamados. El virus llega a los nodos linfáticos y allí se replica activamente, liberando partículas víricas al torrente sanguíneo de manera que el paciente presenta una elevada carga viral; esta etapa suele durar unos dos meses, aunque el organismo tarda entre uno y tres meses en producir anticuerpos contra el VIH. Seguidamente se llega a la etapa asintomática o de latencia clínica. Durante este tiempo los infectados suelen encontrarse bien, la única manifestación de la infección es un test positivo de anticuerpos contra el virus. De todas maneras, durante este tiempo el virus se replica activamente y destruye de manera progresiva el sistema inmunitario de la persona infectada. Cuando el recuento de las células T CD4⁺ llega a ser inferior a 200 unidades/mL el paciente entra en la etapa sintomática de la enfermedad, caracterizada por las infecciones oportunistas (Figura 0.1).^[12-14]

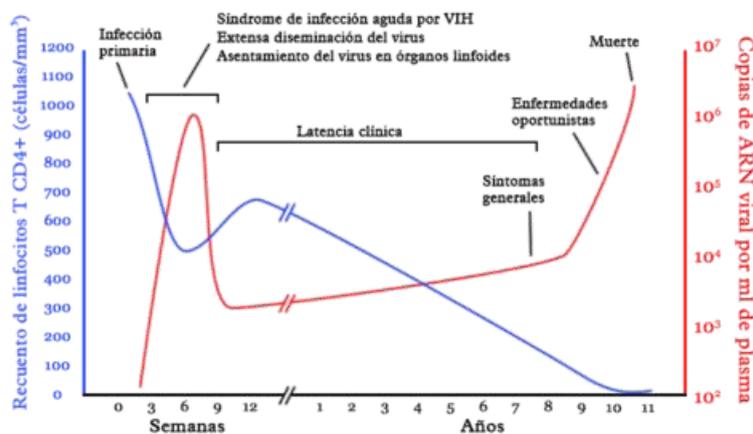


Figura 0.1. Curso de la infección por VIH en ausencia de una terapia antirretroviral. Las primeras semanas después de la infección primaria se caracterizan por una infección aguda, la diseminación del virus por los órganos linfoides y un aumento de la carga viral. Con el desarrollo de anticuerpos contra el virus se controla parcialmente la infección y se entra en una etapa de latencia o fase asintomática durante la cual se produce una disminución de las células T CD4⁺ hasta que llegan por debajo de 200 unidades/mL. En ese momento se inicia la fase sintomática, caracterizada por las infecciones oportunistas que provocan la muerte del paciente.^[15]

La infección se puede transmitir por vía sexual, transfusiones de sangre u otros productos sanguíneos, uso de drogas intravenosas y transmisión vertical de madre a hijo (vía la placenta).

La infección de VIH en humanos se considera una pandemia, una de las más destructivas de la historia. Se estima que en el 2007 vivían 33.2 millones de personas infectadas con VIH.

El porcentaje de personas infectadas por el virus se mantiene en el mismo nivel si se compara con años anteriores, se ha observado una reducción en la mortalidad (2.1 millones en 2007) y una reducción del número de nuevas infecciones (2.5 millones en 2007). El África subsahariana es la región más afectada (Figura 0.2).^[16]

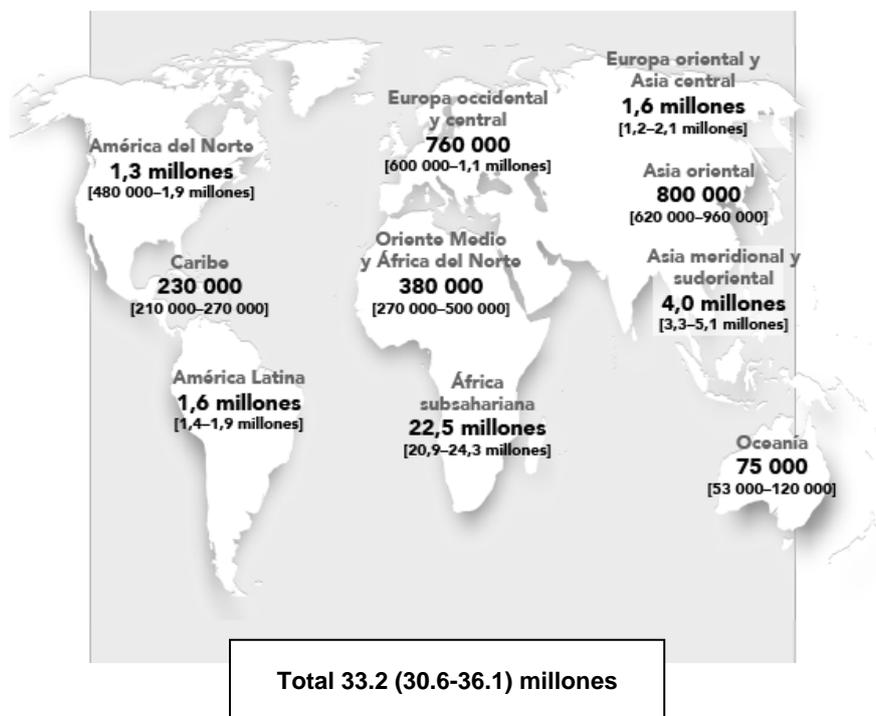


Figura 0.2. Número estimado de personas que vivían con el VIH en 2007 (OMS/UNAIDS). Se observa que la región más afectada es el África subsahariana, con 22.5 millones de infectados, cuando el total de infectados en el mundo es de 33.2 millones.^[16]

0.2. El VIH.

0.2.1. Clasificación, estructura y genoma.

El VIH pertenece al género *Lentivirus* de la familia *Retroviridae*. Los *Lentivirus* difieren de otros retrovirus por su largo periodo de latencia antes de la manifestación clínica de la enfermedad. Los retrovirus pueden estar en forma de viriones que contienen RNA de cadena sencilla (capaces de infectar una nueva célula) y como provirus de doble cadena de DNA (dentro de una célula huésped). El virión presenta una forma esférica de unos 110 nm de diámetro. El material genético del VIH-1 está formado por dos moléculas de RNA de cadena sencilla, situadas en el interior de un core proteínico cónico compuesto de las proteínas p24 de la cápside (CA). Esta cápside está rodeada por la proteína p6, que regula la liberación del virus de la célula. Este conjunto está envuelto por las proteínas p17 de la matriz (MA). El virión completo está envuelto por una bicapa lipídica, derivada de la membrana plasmática de la célula huésped durante el proceso de salida del virión por gemación. Las glicoproteínas de la superficie (SU) gp120 y transmembrana (TM) gp41 forman un complejo no covalente. Las enzimas virales, transcriptasa inversa (RT), integrasa (IN) y proteasa (PR) están empaquetadas junto con el RNA en el core (Figura 0.3).^[13,14,17]

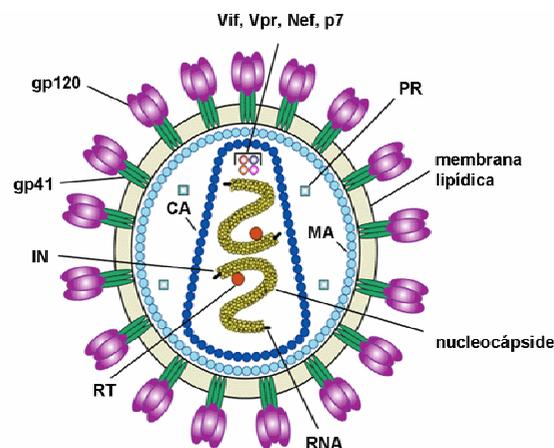


Figura 0.3. Estructura del virión (figura adaptada de www.niaid.nih.gov).

El genoma de la forma integrada del VIH-1 es de 9.8 kb (kilobases) aproximadamente y está formado por nueve genes (Figura 0.4).

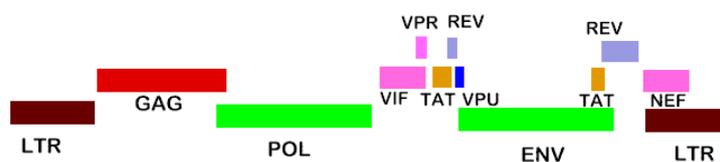


Figura 0.4. Estructura del genoma del VIH (figura de pathmicro.med.sc.edu).

El DNA proviral está flanqueado por unas secuencias repetidas, *Long Terminal Repeats* (LTR), que regulan la transcripción del virus. Como todos los retrovirus, el VIH consta de los tres genes, *gag* (*group antigen gene*), *pol* (*polymerase gene*) y *env* (*envelope gene*). El gen *gag* codifica las proteínas estructurales del core: p24 de la cápside viral, p6 y p7 de la nucleocápside y p17 de la matriz. El gen *env* codifica gp160 precursor de gp120 y gp41, proteínas incrustadas en la bicapa lipídica. El gen *pol* codifica varias enzimas, RT, IN y PR. Los genes *tat*, *rev*, *vif*, *vpr*, *vpu* y *nef*, codifican las proteínas reguladoras Tat, Rev, Nef, Vif, Vpr y Vpu respectivamente (Tabla 0.1).^[18,19]

Tabla 0.1. Genes y proteínas del VIH-1.

Gen	Proteína	Tamaño	Función	Localización
<i>gag</i>	Gag (MA)	p17	Estabilización de la partícula viral	Virión
			Interacción con Env	
			Transporte nuclear del core viral	
	Gag (CA)	p24	Core de la cápside viral	
Gag (NC)	p7	Nucleocápside	Virión	
		Estabilización del RNA		
	p6	Unión a Vpr		
<i>pol</i>	PR	p15	Hidrólisis de los precursores	Virión
	RT	p66/p51	Transcripción inversa	Virión
			Actividad RNasa H	
IN	p32	Integración del DNA proviral		
<i>env</i>	Env	gp120/gp41	Glicoproteínas virales externas Unión a CD4 y correceptores	Membrana viral

Gen	Proteína	Tamaño	Función	Localización
<i>tat</i>	Tat	p16/p14	Transactivador transcripcional	Núcleo celular
<i>rev</i>	Rev	p19	Transporte del RNA del núcleo al citoplasma	Núcleo y citoplasma celular
<i>vif</i>	Vif	p23	Promueve maduración viral e infectividad	Citoplasma, virión
<i>vpr</i>	Vpr	p10/p15	Transporte del complejo de preintegración Para el ciclo de las células infectadas en G2/M Implicada en apoptosis	Núcleo celular, virión
<i>vpu</i>	Vpu	p16	Aumenta la liberación de viriones	Proteína integral de membrana
<i>nef</i>	Nef	p27/p25	Regulación de CD4 negativa Influye en la activación celular Aumenta la infectividad de los viriones	Membrana plasmática, citoplasma celular, virión

0.2.2. Ciclo de vida del VIH.

El ciclo de vida del VIH se puede dividir en dos fases. La fase temprana comprende desde la etapa de unión del virus a la célula hasta la integración del DNA vírico en el genoma celular, y la fase tardía se inicia con la expresión de los genes virales y continúa hasta la salida de la partícula viral de la célula y su posterior maduración.

La fase temprana del ciclo de replicación del VIH se inicia con la entrada del virus en la célula por fusión de la envoltura viral con la membrana de celular. Esta etapa ocurre por medio de una unión específica de la glicoproteína gp120 de la envoltura del virus con el receptor CD4 celular. Esta unión induce una serie de cambios conformacionales que permiten la interacción con un correceptor celular (CXCR4 o CCR5) y finalmente la fusión de la envoltura viral y la membrana celular. Después de la fusión, la nucleocápside viral entra en el citoplasma celular y el RNA viral se copia a una doble cadena de cDNA mediado por la enzima RT. Esta enzima posee tres actividades importantes para la replicación del virus: la transcripción inversa, la actividad RNasa H, que hidroliza el RNA en híbridos de RNA/DNA durante la síntesis de DNA, y la actividad DNA polimerasa DNA-dependiente, para la síntesis de una segunda cadena de DNA proviral. La doble cadena de DNA proviral complejada con proteínas (complejo de preintegración) es transportada al núcleo celular, donde tiene lugar la integración en el genoma celular mediante la acción de la enzima viral IN. El provirus puede permanecer en forma latente durante largos periodos de tiempo.

La fase tardía del ciclo de replicación del virus se inicia con la transcripción del DNA proviral dando lugar a diferentes transcritos virales. Estos transcritos sufren una serie de eventos postranscripcionales y se pueden agrupar en 3 clases según su tamaño: sRNA de 9 kb para la síntesis de Gag y Gag-Pol, así como material genético para los nuevos viriones; sRNA

de 4 kb que codifica para Env, así como las proteínas reguladoras (Vif, Vpu y Vpr); y por último, RNAs de 2 kb que codifican para las proteínas reguladoras Rev, Tat y Nef. Env se procesa por la proteasa de la célula infectada dando lugar a gp120 y gp41. Las nuevas partículas virales se ensamblan en la membrana plasmática para formar el virión. Cada virión consta de Gag, Gag-Pol, genoma viral y Vpr. Después, la membrana plasmática comienza a formar una bicapa lipídica alrededor del *core* viral que contiene gp41 y gp120. Las nuevas partículas virales se liberan al medio extracelular y por último tiene lugar la etapa de maduración del virión, que es cuando adquiere su capacidad infecciosa. Durante la maduración, la proteasa viral hidroliza Gag y Gag-Pol en proteínas más pequeñas (Figura 0.5).^[17,20-22]

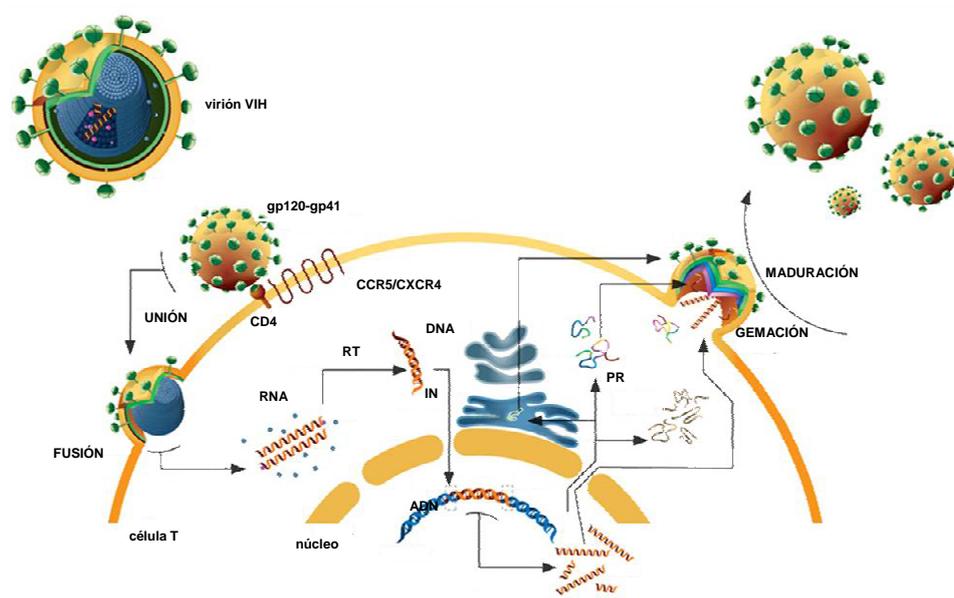


Figura 0.5. Ciclo de vida del VIH. Las etapas del ciclo replicativo del VIH son: unión del VIH a la célula, fusión, liberación de la nucleocápside viral en el citoplasma celular, transcripción inversa, transporte del DNA viral al núcleo celular, integración del DNA viral en el genoma del huésped, transcripción viral, síntesis de proteínas virales, ensamblaje de la nueva partícula viral, liberación al medio por gemación y maduración (figura adaptada de www.tibotec.com).

El virus se puede transmitir mediante infección de células sanas por partículas virales libres, aunque la vía más efectiva es por contactos entre células infectadas y sanas a través de sinapsis virológicas.^[23-27]

0.2.3. Etapa de entrada y fusión.

El VIH entra en la célula por interacción de las glicoproteína gp120 de la envoltura del virus con el receptor celular CD4 y los correceptores CXCR4 y CCR5 (receptores de quimiocinas). Después de la unión a los receptores celulares se produce la fusión de las membranas viral y celular permitiendo la liberación de la nucleocápside viral en el citoplasma celular (Figura 0.6).

En 1984 se identificó el receptor CD4 como el que usaba el VIH para entrar en las células.^[19,28,29] Este receptor es una glicoproteína transmembrana expresada principalmente en linfocitos T y otras células del sistema inmunológico e interviene en la transducción de señal que lleva a la activación de células T en presencia de un antígeno.^[30]

El proceso se inicia con la unión de la proteína viral gp120, que está en forma de trímero en la envoltura viral, con el receptor celular CD4. La glicoproteína gp120 está formada por un dominio interno y uno externo, consta de cinco regiones conservadas (C1-C5) y cinco variables (V1-V5). Una vez se ha unido a CD4, el virus necesita interactuar con un correceptor de quimiocinas para entrar en la célula. Los principales correceptores implicados en este proceso son CXCR4 y CCR5.

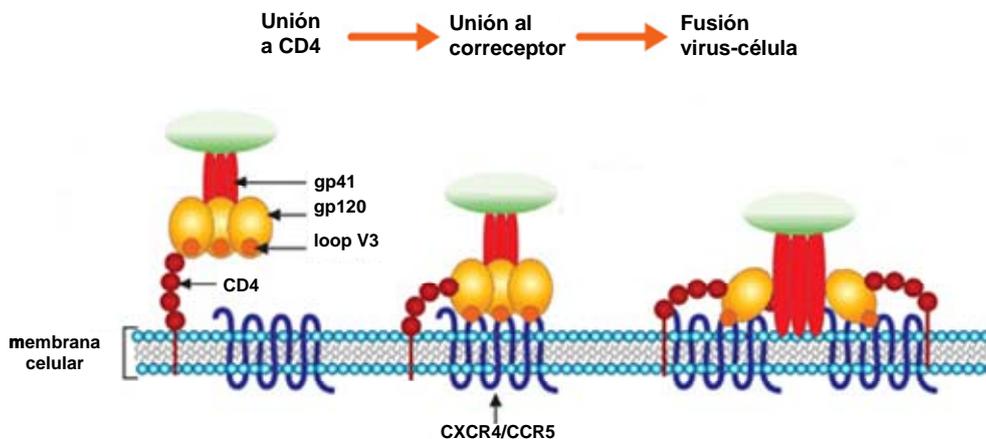


Figura 0.6. Entrada y fusión del VIH con la célula. La primera etapa es la unión del virus al receptor celular CD4, seguida de la unión al correceptor celular CXCR4 o CCR5 y por último la fusión de las membranas del virus y de la célula que permite liberar la nucleocápside viral en el citoplasma celular.^[31]

La formación del complejo CD4-gp120 induce una serie de cambios conformacionales en la estructura de la envoltura viral que exponen su sitio de unión con el correceptor CXCR4 o CCR5. CXCR4 y CCR5 pertenecen a la familia de receptores acoplados a proteínas G (GPCR) de 7 dominios transmembrana. El complejo CD4-gp120 se une al correceptor a través de la región V3 (Figura 0.6), cuya secuencia de aminoácidos determina el uso de CXCR4 y/o CCR5.

Esta unión causa otra serie de cambios conformacionales de las proteínas de la envoltura viral en que se reestructura la gp41 acercando dos regiones de hélice alfa, uno del extremo N-terminal (dominio HR1) y otro del extremo C-terminal (dominio HR2). Esta interacción de los trímeros de las dos hélices alfa provoca el paso por una estructura tipo horquilla, que termina con el acercamiento de las membranas viral y celular. El contacto entre ambas membranas crea el poro de fusión a través del cual se libera la nucleocápside viral en el citoplasma celular (Figura 0.7).^[32-39]

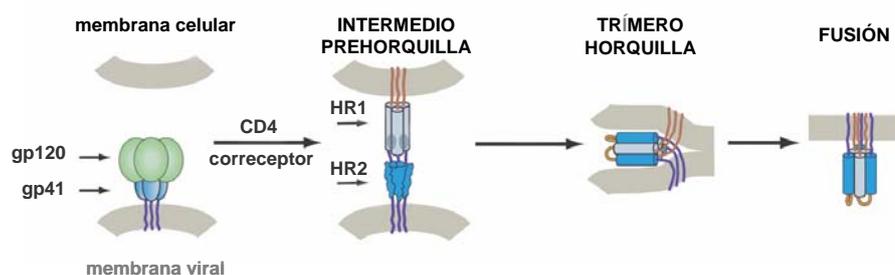


Figura 0.7. Fusión de las membranas viral y celular.^[33]

0.2.4. Los correceptores CXCR4 y CCR5. Tropismo del VIH.

Los receptores de quimiocinas CXCR4^[40-42] y CCR5^[43-46] se identificaron en 1996 como correceptores de la entrada del VIH en la célula. Además, el correceptor CXCR4 está involucrado en procesos de metástasis en varios tipos de cáncer.^[47]

Las quimiocinas son citocinas quimiotácticas que activan y dirigen la migración de leucocitos. Se caracterizan por la conservación de cuatro residuos de cisteína que se unen formando dos puentes disulfuro. Se pueden agrupar en dos familias, las quimiocinas CXC y las CC, dependiendo de si las dos primeras cisteínas están o no separadas por otro aminoácido. Las quimiocinas CXC activan los neutrófilos, mientras que las quimiocinas CC activan principalmente a monocitos, linfocitos, basófilos y eosinófilos y tienen un papel importante en la proliferación de células T y alergias.^[48]

El ligando natural de CXCR4 es la quimiocina SDF-1, mientras que CCR5 es el receptor de las quimiocinas RANTES, MIP-1 α y MIP-1 β .^[49] Se ha descrito una mutación en CCR5, conocida como CCR5- Δ 32, que prácticamente confiere resistencia completa a la infección por VIH-1, corroborando el papel de dicha proteína como correceptor de la entrada del virus (Figura 0.8).^[50]

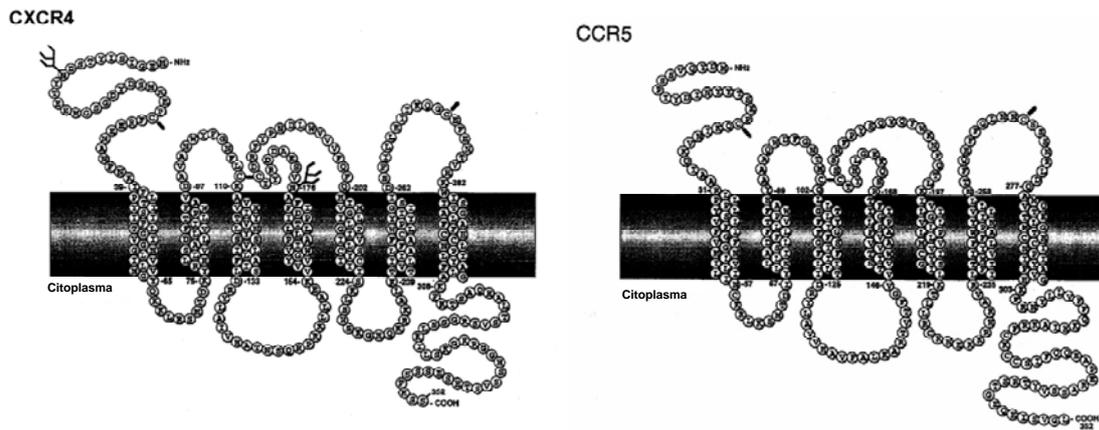


Figura 0.8. Representación esquemática de los correceptores CXCR4 y CCR5. Se aprecian los siete dominios transmembrana, así como los loops extra e intracelulares.^[51]

La capacidad del VIH-1 de infectar diferentes tipos de células se conoce como el tropismo celular (Figura 0.9). Las cepas del VIH-1 se pueden clasificar en cepas T-trópicas, M-trópicas o D-trópicas, según infecten a macrófagos, linfocitos T o ambos tipos de células respectivamente. También se pueden clasificar según el correceptor que utilicen para entrar en la célula: X4 para las cepas que utilizan el correceptor CXCR4, R5 para las cepas que utilizan el correceptor CCR5, y X4R5 para aquellas cepas que pueden utilizar ambos correceptores. Otra clasificación que del tropismo del VIH se basa en la capacidad de inducir la formación sincitios (SI), o fusión de una célula infectada con otras células adyacentes que no lo están, y la capacidad de no inducir la formación de sincitios (NSI).^[52,53] Como se ha indicado anteriormente, la secuencia de aminoácidos del dominio V3 de gp120 determina el uso de CXCR4 o CCR5.^[32,54,55]

Los virus que utilizan el correceptor CCR5 son los que se suelen transmitir de un individuo a otro y son los predominantes en la fase asintomática de la enfermedad, mientras que en la fase tardía de la enfermedad son mayoritarios los virus del fenotipo X4 y R5X4.^[34,54]

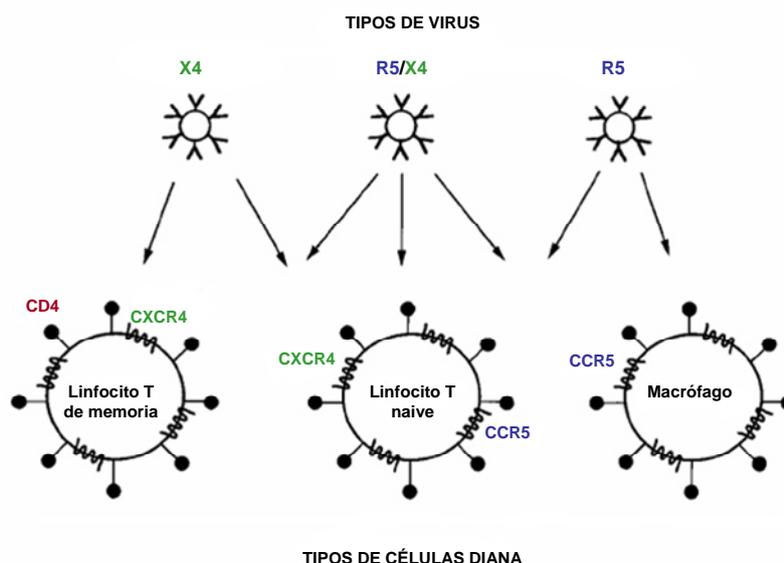


Figura 0.9. Tropismo celular del VIH-1.^[15]

0.2.5. Variabilidad genética.

El VIH es un virus muy variable debido a su elevada tasa de mutación^[56] junto con su rápida replicación, cada día se generan unos 10^{10} viriones.^[57] Se puede clasificar en dos tipos: VIH-1, causante de la mayoría de las infecciones, y VIH-2 que es minoritario y mayoritariamente se encuentra en el oeste de África. El VIH-1, a su vez, se clasifica en diferentes grupos: M (*major*), O (*outlier*) y N (*non-M/non-O*) (Figura 0.10).

Un estudio filogenético del VIH-1 ha localizado su origen en el retrovirus SIV_{cpz} (*simian immunodeficiency virus*), que también se ha encontrado en una colonia de chimpancés del sur de Camerún.^[58] Se cree que pasó a los humanos a través de unos chimpancés capturados por unos cazadores, y tras muchas mutaciones se transformó en un virus humano. La secuencia genética del virus de una biopsia de un nódulo linfático tomada en 1960 de una mujer adulta en Kinshasa se comparó con la de una muestra de plasma de 1959 de un pueblo cercano. El resultado demostró que la diversificación del VIH-1 ocurrió en África centro-occidental antes de que se conociera la pandemia del sida. Se estima la fecha de este antepasado común entre 1902 y 1921. Transmisiones independientes entre 1902 y 1941 generaron los tres grupos mayoritarios del VIH-1: M, O y N.^[59,60] Recientemente se han encontrado indicios de que el grupo O se pudo haber originado en gorilas^[61], aunque las cepas de los grupos M y N se encuentran en chimpancés.

Más del 90% de las infecciones corresponden al grupo M, y se divide en diferentes subtipos o clades: A, B, C, D, F, G, H, J y K.^[62-64] El genoma viral puede diferir un 15-20% entre

clades. En América, Europa y Australia la clade más común es la B, mientras que la C predomina en el sur de África.

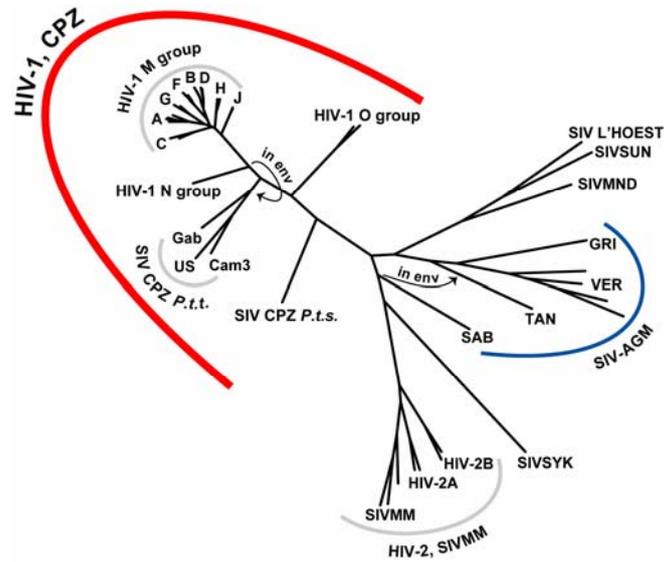


Figura 0.10. Relaciones filogenéticas de los lentivirus de primates (figura adaptada de www.hiv.lanl.gov).

0.3. Terapia antirretroviral.

0.3.1. Terapia antirretroviral actual.

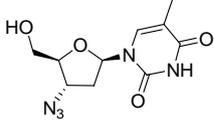
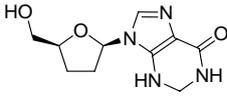
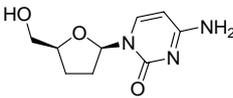
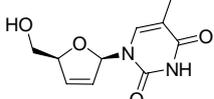
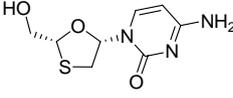
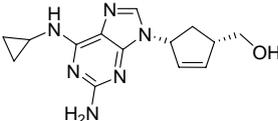
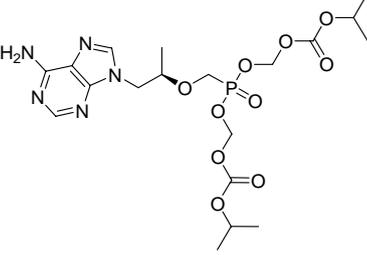
La elevada tasa de replicación y mutación del VIH provoca la rápida generación de cepas resistentes a fármacos, por tanto es recomendable una terapia con varios fármacos y dirigida a varias dianas. La terapia antirretroviral actual se basa en combinaciones de tres clases de fármacos: inhibidores de la transcriptasa inversa análogos de nucleósidos (NRTIs, *Nucleoside Reverse Transcriptase Inhibitors*), no nucleósidos (NNRTIs, *Non-Nucleoside Reverse Transcriptase Inhibitors*) e inhibidores de la proteasa (PIs, *Protease Inhibitors*). La combinación de estos fármacos se conoce como terapia antirretroviral altamente activa (HAART, *Highly Active Antiretroviral Therapy*). Esta terapia ha convertido el sida en una enfermedad crónica.^[65-68]

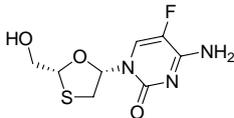
Inhibidores de la transcriptasa inversa análogos de nucleósidos.

Los NRTIs actúan sobre la enzima vírica RT e impiden el desarrollo de DNA viral a partir del RNA. Requieren tres pasos de fosforilación en la célula huésped para pasar a la forma activa que se incorpora en el DNA viral. Actúan en el sitio de unión del sustrato de la RT.

El primer NRTI desarrollado fue la zidovudina, un análogo de timidina también conocido como AZT y aprobado por la FDA en 1987. Inicialmente se desarrolló como anticancerígeno y posteriormente se descubrió su actividad anti-VIH.^[69] El siguiente NRTI aprobado por la FDA fue didanosina, también conocido como ddl, en 1991. Es un profármaco de dideoxiadenosina (ddA). Zalcitabina (ddC) es un análogo de pirimidina, aprobado como antiviral en 1992. Dos años más tarde, en 1994, se aprobó la stavudina (d4T), un análogo de timidina. El año siguiente se aprobó lamivudina (3TC), un análogo de citidina, generalmente para el uso en combinación con AZT. Un análogo de guanosina, abacavir (ABC) fue aprobado en 1998. El último en aprobarse fue emtricitabina (FTC), un análogo de citidina, en 2003. El resto de NRTI aprobados por la FDA son combinación de estos fármacos. En este grupo también se incluye un análogo de nucleótido, tenofovir (tenofovir disoproxil fumarato, TDF), aprobado en 2001, esta molécula necesita dos pasos de fosforilación en vez de los tres del resto de compuestos del grupo (Tabla 0.2).^[67,70]

Tabla 0.2. Fármacos NRTIs aprobados por la FDA para el tratamiento de la infección por VIH.

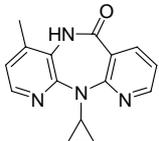
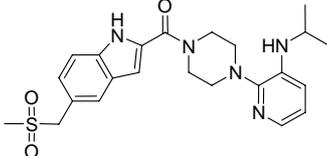
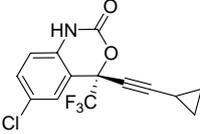
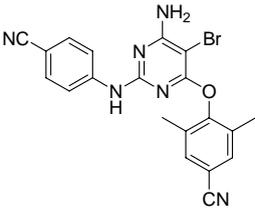
Nombre	Estructura	Fabricante	Fecha aprobación
Retrovir® zidovudina azidotimidina AZT ZDV		GlaxoSmithKline	19.03.87
Videx® didanosina dideoxinosina ddl		Bristol Myers-Squibb	09.10.91
Hivid® zalcitabina dideoxicitidina ddC		Hoffmann-La Roche	19.06.92
Zerit® stavudina d4T		Bristol Myers-Squibb	24.06.94
Epivir® lamivudina 3TC		GlaxoSmithKline	17.11.95
Combivir®	lamivudina + zidovudina	GlaxoSmithKline	27.09.97
Ziagen® abacavir succinato ABC		GlaxoSmithKline	17.12.98
Videx EC® didanosina cubierta ddl EC	didanosina cubierta	Bristol Myers-Squibb	31.10.00
Trizivir®	abacavir + zidovudina + lamivudina	GlaxoSmithKline	14.11.00
Viread® tenofovir disoproxil fumarato TDF		Gilead Sciences	26.10.01

Nombre	Estructura	Fabricante	Fecha aprobación
Emtriva® emtricitabina FTC		Gilead Sciences	02.07.03
Epzicom®	abacavir + lamivudina	GlaxoSmithKline	02.08.04
Truvada®	tenofovir + emtricitabina	Gilead Sciences	02.08.04

Inhibidores de la transcriptasa inversa no análogos de nucleósidos.

Los NNRTIs interactúan con la RT en un sitio alostérico (de no unión del sustrato) de la enzima viral (Tabla 0.3).^[67,70]

Tabla 0.3. Fármacos NNRTIs aprobados por la FDA para el tratamiento de la infección por VIH.

Nombre	Estructura	Fabricante	Fecha aprobación
Viramune® nevirapina NVP		Boehringer Ingelheim	21.06.96
Rescriptor® delavirdina DLV		Pfizer	04.04.97
Sustiva® efavirenz EFV		Bristol Myers-Squibb	17.09.98
Intelence® etravirina		Tibotec Therapeutics	18.01.08

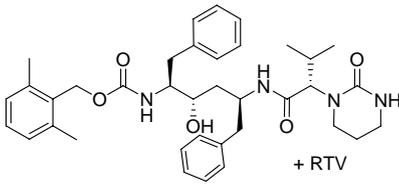
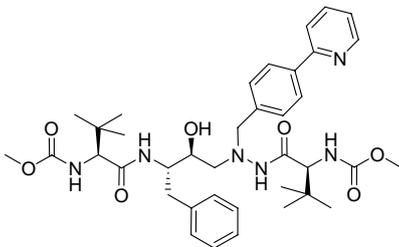
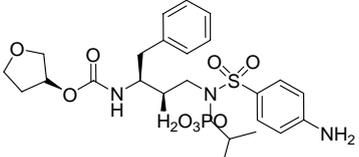
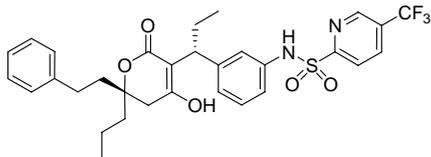
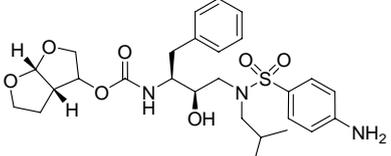
El 12 de julio de 2006 la FDA aprobó el fármaco Atripla® que es combinación de efavirenz (NNRTI), emtricitabina (NRTI) y tenofovir disoproxil fumarato (NRTI), comercializado por Bristol-Myers Squibb y Gilead Sciences.

Inhibidores de la proteasa.

La PR fragmenta los polipéptidos Gag y Gag-Pol, esenciales para la maduración del virión. Los inhibidores de la proteasa (PIs, *Protease Inhibitors*) actúan inhibiendo esta enzima, de manera que se obtienen partículas víricas inmaduras y no infecciosas (Tabla 0.4).^[67,70]

Tabla 0.4. Fármacos PIs aprobados por la FDA para el tratamiento de la infección por VIH.

Nombre	Estructura	Fabricante	Fecha aprobación
Invirase® saquinavir mesilato SQV		Hoffmann-La Roche	06.12.95
Norvir® ritonavir RTV		Abbott Laboratories	01.03.96
Crixivan® indinavir IDV		Merck	13.03.96
Viracept® nelfinavir mesilato NFV		Agouron Pharmaceuticals	14.03.97
Agenerase® amprenavir APV		GlaxoSmithKline	15.04.99

Nombre	Estructura	Fabricante	Fecha aprobación
Kaletra® lopinavir + ritonavir LPV + RTV		Abbott Laboratories	15.09.00
Reyataz® atazanavir sulfato ATV		Bristol-Myers Squibb	20.06.03
Lexiva® fosamprenavir calcio FOS-APV		GlaxoSmithKline	20.10.03
Aptivus® tipranavir TPV		Boehringer Ingelheim	22.06.05
Prezista® darunavir DRV		Tibotec, Inc.	23.06.06

Fármacos que actúan sobre otras dianas.

El 13 de marzo de 2003 la FDA aprobó el uso de un nuevo fármaco, Fuzeon® (enfuvirtide, T20), comercializado por Hoffmann-La Roche y Trimeris. Es un péptido lineal sintético de 36 L-aminoácidos naturales (Figura 0.11) que se une al dominio HR1 de la proteína viral gp41 impidiendo el cambio conformacional necesario para la fusión de las membranas viral y celular.^[71-75]

En agosto de 2007, Pfizer sacó al mercado Selzentry® (maraviroc), un antagonista del correceptor celular CCR5 (Figura 0.11). Evita que la variante T-trópica del virus, la más común en las etapas iniciales de infección por VIH, no pueda unirse al correceptor e infectar la célula.^[76,77]

En octubre del mismo año, Merck lanzó el inhibidor de la integrasa Isentress® (raltegravir), que previene la inserción del DNA viral en el DNA de la célula huésped, impidiendo la formación del provirus de VIH y por tanto la replicación viral (Figura 0.11).^[78]

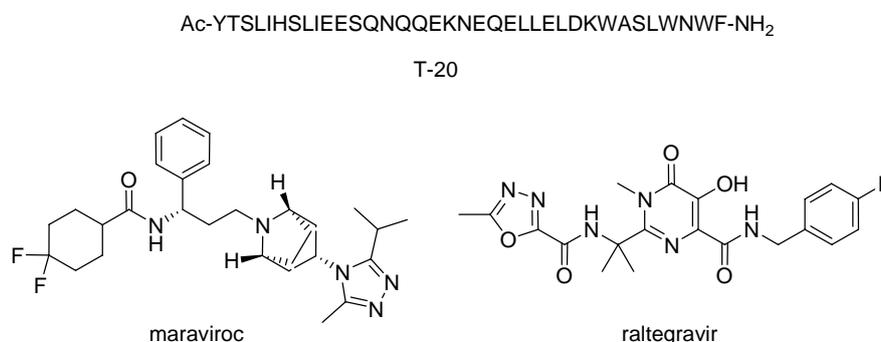


Figura 0.11. Estructuras del inhibidor de fusión T-20, del antagonista del correceptor CCR5 maraviroc y del inhibidor de la integrasa raltegravir.

0.3.2. Fármacos en desarrollo.

La aparición de virus resistentes a los fármacos de la terapia antirretroviral actual (NRTIs, NNRTIs y PIs) es un factor importante en el fracaso del tratamiento. Esta situación ha provocado la necesidad de encontrar fármacos que actúen sobre nuevas dianas terapéuticas o compuestos nuevos capaces de suprimir las cepas de virus resistentes a los inhibidores de RT y PR actuales (Tabla 0.5).

Tabla 0.5. Estructura y diana terapéutica de los fármacos en desarrollo para la terapia antirretroviral en septiembre del 2009.

Nombre	Diana terapéutica	Fase de desarrollo	Farmacéutica
TNX-355	CD4	II	Tanox
PRO140	CCR5	II	Progenics Pharmaceuticals
Vicriviroc	CCR5	II/III	Schering-Plough Corp
AMD070	CXCR4	II	Genzyme Corp.
Elvitegravir	IN	III	Gilead Sciences
Rilpivirina	NNRTI	III	Tibotec
Apricitabina	NRTI	III	Avexa
Elvucitabina	NRTI	II/III	Achillion Pharmaceuticals

Nombre	Diana terapéutica	Fase de desarrollo	Farmacéutica
Fosalvudina	NRTI	II	Heidelberg Pharma
KP-1461	NRTI	II	Koronis Pharmaceuticals
Racivir	NRTI	II/III	Pharmasset

Hay cuatro fármacos en fases clínicas que actúan sobre la etapa de entrada y fusión del virus y la célula. TNX-355 (también llamado ibalizumab) es un anticuerpo monoclonal que se une a CD4 en el dominio implicado en el cambio conformacional necesario para la entrada del virus en la célula.^[79,80] Otro anticuerpo monoclonal es PRO140, que se une a CCR5 en un dominio extracelular, inhibe la unión del virus pero no altera la actividad natural de este receptor.^[81,82] En cuanto a moléculas orgánicas pequeñas, está en desarrollo vicriviroc (también conocido como SCH-D), que actúa sobre CCR5. Este fármaco fue descubierto en un estudio de estructura-actividad (SAR, *structure-activity relationship*) que se llevó a cabo para mejorar las propiedades farmacocinéticas de SCH-C así como para reducir su toxicidad cardíaca (Figura 0.12).^[83,84]

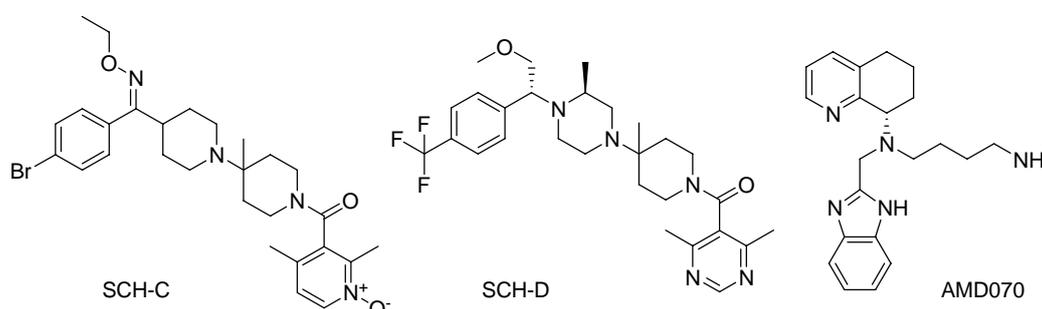


Figura 0.12. Estructuras de vicriviroc (SCH-D), inhibidor de la unión del VIH al receptor celular CCR5, su predecesor SCH-C y AMD070, inhibidor de la unión al receptor CXCR4.

Otra molécula pequeña que ha entrado en fase clínica II de desarrollo es AMD070, un antagonista específico del receptor celular CXCR4 (Figura 0.12).^[85,86]

Otras dianas terapéuticas son las enzimas virales IN y RT. Solamente hay un fármaco en desarrollo que actúa sobre la IN, elvitegravir cuyo modo de acción es por bloqueo de la reacción de transferencia de cadena de DNA viral al genoma del huésped mediada por la IN.^[87] En cambio, para la RT hay seis compuestos en fases de desarrollo clínico (Figura 0.13). Uno de ellos pertenece a la clase de NNRTIs, rilpivirina, de la familia de las diarilpirimidinas.^[88,89] Los otros cinco son de la clase de NRTIs, como apricitabina, un análogo de deoxicitidina.^[90-92] Elvucitabina es un análogo de beta-L-(-) nucleósido, que se desarrolló para mejorar la actividad

de lamivudina, un fármaco ya aprobado por la FDA para el tratamiento antirretroviral.^[93,94] Fosalvudina es un NRTI compuesto del monofosfato de 3-fluoro-2,3-dideoximidina unido a un portador lipídico.^[95] KP-1641 actúa por mutación viral selectiva o letal durante la replicación viral.^[96] Racivir es una mezcla racémica 50:50 de los β -enantiómeros (-) y (+) de emtricitabina, que es el (-)- β -enantiómero y está aprobado por la FDA para el tratamiento de la infección por VIH.^[97]

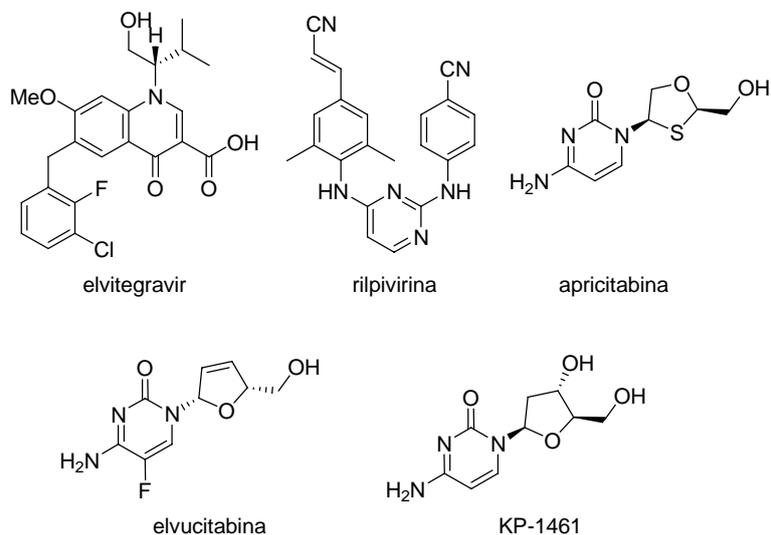


Figura 0.13. Estructuras de los fármacos en desarrollo que actúan sobre las enzimas IN y RT.

0.4. Los biclamos.

0.4.1. Estructura general de los biclamos y actividad anti-VIH.

Los biclamos constituyen una familia de compuestos cuya estructura general está formada por dos monoclamos (1,4,8,11-tetraazaciclotetradecano) unidos por un *linker* aromático o alifático. El biclamo JM1657 (Figura 0.14) se descubrió como una impureza en la preparación a gran escala del macrociclo poliaminado 1,4,8,11-tetraazaciclotetradecano (ciclamo, $IC_{50} = 399 \mu\text{M}$, $CC_{50} = 1248 \mu\text{M}$, en células MT-4 infectadas por VIH-1 (III_B)), englobado en un proyecto de preparación de complejos metálicos de compuestos orgánicos, formados por los cuatro átomos de nitrógeno del anillo de ciclamo, como nuevos compuestos anti-VIH.

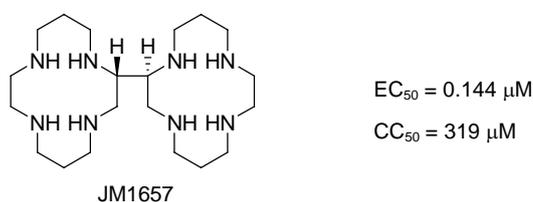
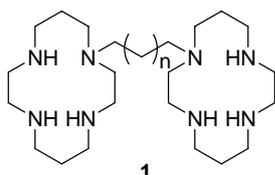


Figura 0.14. Estructura, actividad y citotoxicidad de JM1657 evaluadas en células MT-4 infectadas por VIH-1 (III_B).^[98]

En Johnson Matthey identificaron una serie de biclamos (**1**) unidos por un *linker* de tipo alifático (etileno, trimetileno, hexametileno) como inhibidores potentes y selectivos relacionados con un evento temprano de la replicación de los virus VIH-1 y VIH-2 (Tabla 0.6). Mediante ensayos de tiempo de adición determinaron que esta familia de compuestos actúa en una etapa del ciclo de vida del virus posterior a la adsorción del virus y anterior a la transcripción inversa.^[98] Además, estos compuestos no inhiben la formación de sincitios (fusión de una célula infectada con otras células adyacentes que no lo están) a las concentraciones en que son activos frente a la replicación del VIH, por tanto, difieren de inhibidores de fusión del virus y la célula anteriormente descritos.^[99-101]

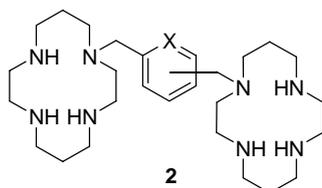
Tabla 0.6. Estructura de los biciclamos con linker alifático (1) sintetizados por Johnson Matthey. La actividad está expresada en μM como IC_{50} o concentración del 50% de inhibición de citopaticidad por VIH-1 (III_B) en células MT-4. La citotoxicidad está expresada en μM como CC_{50} o 50% de concentración citotóxica para células MT-4.^[98]



Compuesto	IC_{50} / μM	CC_{50} / μM
n=0 JM2762	18.6	349
n=1 JM2763	0.248	>622
n=4 JM2849	0.616	290

Estudios posteriores permitieron mejorar la actividad antirretroviral de esta familia de compuestos utilizando un *linker* aromático (2) (Tabla 0.7), en particular el *p*-fenilenbismetileno. Este compuesto, conocido como AMD3100 o JM3100, presenta unos valores de actividad, expresados como EC_{50} , del orden de 1 a 10 ng/mL, no muestra toxicidad hasta concentraciones de 500 $\mu\text{g/mL}$ y por tanto su índice de selectividad es mayor o igual a 100,000.^[102]

Tabla 0.7. Estructura de los biciclamos con linker aromático (2) sintetizados por Johnson Matthey. La actividad está expresada en $\mu\text{g/mL}$ como EC_{50} o concentración efectiva para inhibir el 50% de muerte inducida por VIH-1 (III_B) en células MT-4. La citotoxicidad está expresada en $\mu\text{g/mL}$ como CC_{50} o 50% de concentración citotóxica para células MT-4.^[102]



Compuesto	EC_{50} / $\mu\text{g}\cdot\text{mL}^{-1}$	CC_{50} / $\mu\text{g}\cdot\text{mL}^{-1}$
<i>meta</i> , X=CH JM2986	0.04	>500
<i>para</i> , X=CH JM3100	0.005	>500
<i>meta</i> , X=N JM3106	0.03	>500

En el mismo grupo de investigación realizaron estudios sobre el efecto del tamaño del macrociclo y sustituyentes sobre el *linker* aromático. Concluyeron que no es necesaria la presencia de dos macrociclos idénticos a ambos lados del *linker*, estos sustituyentes poliaminados deben ser estructuras cíclicas para obtener compuestos con buena actividad anti-VIH, el tamaño óptimo del macrociclo es de 14 eslabones y la sustitución sobre el *linker* aromático con grupos aceptores o dadores apenas influye sobre la actividad, excepto en el caso de sustituyentes voluminosos, como el fenilo, que causan impedimento estérico y restringen la rotación de los macrociclos provocando una disminución de la actividad.^[103] Seguidamente, también estudiaron el efecto de diferentes *linkers* heteroaromáticos sobre la actividad antirretroviral de los biciclamos. Utilizaron piridinas y pirazinas con diferentes esquemas de sustitución; de todos ellos, los sistemas con *linkers* 2,6- y 3,5-piridínicos presentaron los mejores valores de actividad anti-VIH aunque en ningún caso fueron superiores a los del biciclamo de referencia AMD3100.^[104] Posteriormente evaluaron la influencia de los átomos de nitrógeno del macrociclo poliaminado, sustituyendo alguno de estos átomos por azufre u oxígeno. Obtuvieron biciclamos con actividad reducida respecto a los compuestos análogos con átomos de nitrógeno e incluso presentaban un aumento en la citotoxicidad, por tanto se descartó esta modificación del *lead* AMD3100. Por otra parte comprobaron que una sustitución del macrociclo con anillos pirimidínicos causaba un efecto beneficioso para la actividad antiviral, como en el caso de AMD3329 (Figura 0.15),^[105] o incluso la formación de complejos metálicos del macrociclo, en particular de Zn^{2+} .^[106,107]

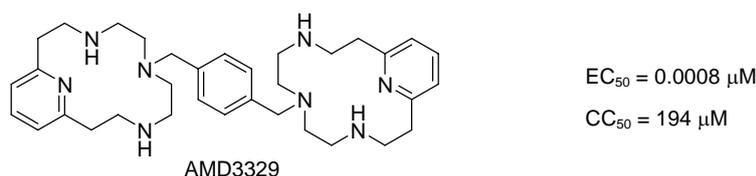


Figura 0.15. Estructura, actividad y citotoxicidad de AMD3329 evaluadas en células MT-4 infectadas por VIH-1 (III_B).^[105]

0.4.2. Desarrollo de monociclamos.

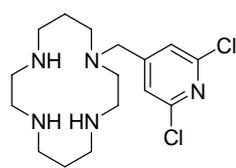
Durante el desarrollo de AMD3100 se vio que presentaba baja biodisponibilidad oral y toxicidad cardíaca relacionadas con su elevada carga positiva a pH fisiológico.^[108] La administración intravenosa es necesaria para el AMD3100, lo que impone limitaciones importantes en cuanto a su uso clínico para terapia anti-VIH a largo plazo. Sin embargo, también se descubrió que el AMD3100 era capaz de movilizar células madre de la médula ósea inhibiendo la unión entre SDF-1 α y CXCR4 (AMD3100, también llamado plerixafor y

comercializado por Genzyme bajo el nombre de Mozobil®).^[109] Para esta indicación la administración parenteral está aceptada, ya que como se ha comentado anteriormente, el AMD3100 no es oralmente biodisponible.

Para mejorar la biodisponibilidad oral de los biciclamos, reduciendo la carga positiva a pH fisiológico, se diseñaron una serie de monociclamos, que sólo contienen un anillo de ciclamo en su estructura, entre los que destacan AMD3451 y AMD3465.

El análogo *N*-piridinilmetilciclamo AMD3451 resultó ser un antagonista dual específico de CXCR4/CCR5, por tanto, inhibe la etapa de entrada del virus en la célula bloqueando ambos correceptores (Tabla 0.8).^[110]

Tabla 0.8. Estructura del monociclamo AMD3451 y la actividad frente a la replicación de diferentes cepas de VIH en diferentes líneas celulares que expresan el correceptores CXCR4 o CCR5. La actividad está expresada en μM como IC_{50} o inhibición del 50% de replicación celular.^[110]



AMD3451

Cepa de virus	Correceptor	Línea celular	IC_{50} / μM
VIH-1 ADA	CCR5	PBMC	6.0
VIH-1 III _B	CXCR4	MT-4	1.2
VIH-1 NL4.3	CXCR4	MT-4	1.8

Sin embargo, el monomacrociclo *N*-piridilmetilenciclamo AMD3465 inhibía exclusivamente CXCR4 con una actividad comparable o ligeramente mejorada respecto a AMD3100 (Figura 0.16).^[108]

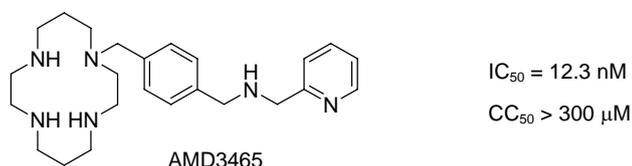


Figura 0.16. Estructura, actividad y citotoxicidad de AMD3465 evaluadas en células MT-4 infectadas por VIH-1 (III_B).^[108]

0.4.3. Modo de acción del AMD3100.

Ensayos de tiempo de adición del AMD3100 sobre células infectadas con VIH habían mostrado que este compuesto actuaba en una etapa temprana del ciclo de replicación del VIH, pero no se había conseguido identificar cuál era su diana terapéutica.^[102] El grupo de De Clercq realizó un estudio con AMD3100 y JM2763 que permitió identificar la entrada del virus en la célula como la etapa sobre la que actuaban estos biclamos y postularon que su diana terapéutica era gp120.^[111] Al cabo de un año descubrieron que AMD3100 era inactivo frente a cepas de VIH M-trópicas (que utilizan como correceptor para la entrada en la célula el receptor de quimiocinas CCR5), mientras que era activo frente a cepas T-trópicas (que utilizan el receptor de quimiocinas CXCR4 como correceptor de entrada en la célula). Este hallazgo permitió identificar la diana terapéutica del biclamo AMD3100: es un antagonista selectivo del receptor de quimiocinas CXCR4.^[112-114] Esté y colaboradores mostraron que la evolución del VIH-1 hacia cepas más patogénicas, las que utilizan CCR5 como correceptor para la entrada en la célula, se puede prevenir mediante el bloqueo del correceptor CXCR4.^[115] En base a los estudios de Dealwis *et al.*, que revelan que la unión entre SDF-1 α y CXCR4 se debe a la interacción electrostática entre la superficie positivamente cargada del ligando natural del receptor de quimiocinas y los *loops* extracelulares cargados negativamente del correceptor del VIH-1,^[116] se realizaron estudios de mutagénesis para determinar la afinidad de unión entre los mono- y biclamos y el correceptor CXCR4. El resultado mostró que tanto Asp¹⁷¹ en el dominio TM-IV y Asp²⁶² en el dominio TM-VI son esenciales para la unión.^[117,118] Posteriormente también se identificó un aminoácido adicional de CXCR4 que interaccionaba con AMD3100, Glu²⁸⁸ en el dominio TM-VII (Figura 0.17).^[119]

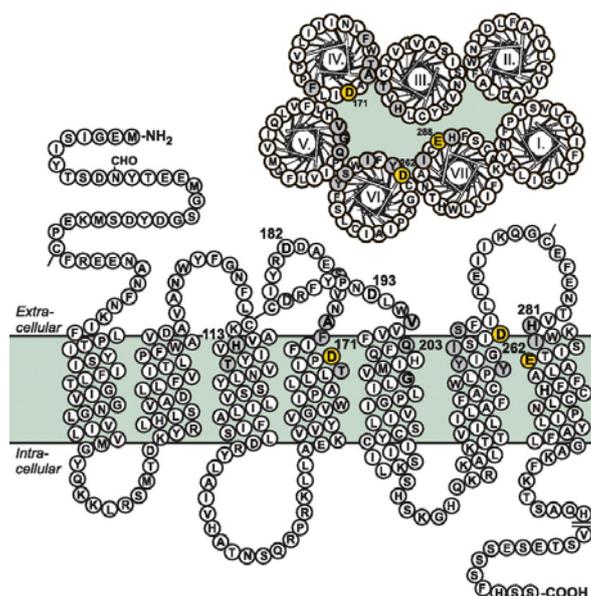


Figura 0.17. Diagrama del correceptor CXCR4 con los aminoácidos de los siete dominios transmembrana (TM) importantes para la unión de AMD3100 identificados por mutagénesis. El color gris/amarillo indica los residuos mutados. Los aminoácidos coloreados de amarillos indican los sitios de interacción con el AMD3100 identificados.^[120]

0.5. Investigación sobre el sida en el IQS.

El presente trabajo se enmarca dentro del proyecto de investigación sobre el diseño, síntesis y evaluación de la actividad antirretroviral de nuevos compuestos anti-VIH, que se lleva a cabo en los laboratorios de Síntesis y Diseño Molecular del GEM (Grup d'Enginyeria Molecular) en el IQS en colaboración con el laboratorio de Retrovirología de la Fundació IrsiCaixa en el Hospital Universitari Germans Trias i Pujol. Dicho proyecto ha sido financiado en parte por la Marató de TV3 (proyecto 020930) y en parte por el Ministerio de Educación y Ciencia (proyecto SAF-2007-6322). En el marco de este proyecto se engloba el presente trabajo, concretamente en el diseño y síntesis de nuevos análogos al AMD3100 como inhibidores de la entrada del virus por bloqueo del correceptor CXCR4.

En el laboratorio de Diseño Molecular del GEM se ha desarrollado un *software* llamado PRALINS (*Program for the Rational Analysis of Libraries in Silico*), que permite realizar selección de compuestos de una quimioteca aplicando criterios de diversidad o similitud de manera combinatoria o no combinatoria (Figura 0.18).^[121,122]

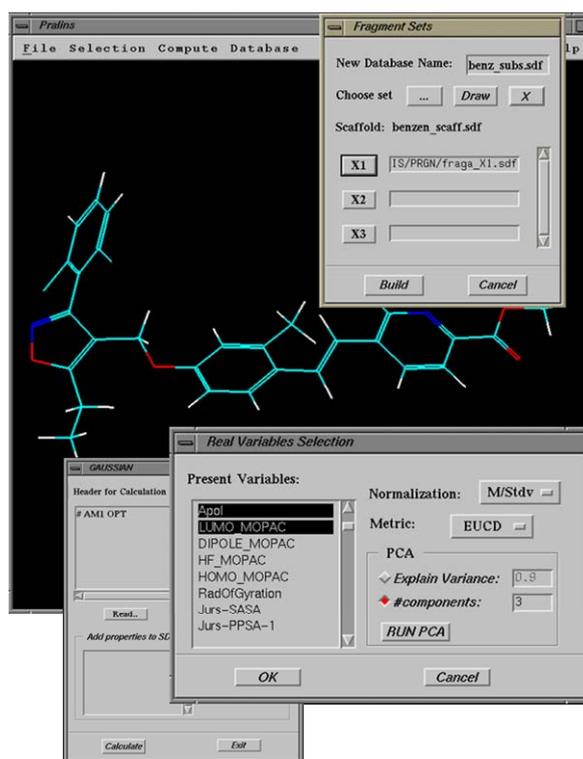


Figura 0.18. Interfaz gráfica del programa PRALINS.

También se tiene experiencia en selección de compuestos por cribado virtual de quimiotecas, de gran utilidad en la selección de compuestos de una quimioteca a partir de modelos basados en la estructura de otros compuestos activos para la misma diana (*ligand-based*) o a partir de la estructura del receptor diana de los compuestos que se quieren sintetizar (*structure-based*).^[123,124] El cribado virtual ha surgido en los últimos años como una alternativa o aproximación a los métodos de *high-throughput screening* (HTS) de quimiotecas formadas por un número elevado de moléculas. Además, se dispone de la estructura del correceptor CXCR4 (Figura 0.19) que se ha modelado por homología a la rodopsina bovina (también de la familia de las GPCRs) debido a la falta de la estructura determinada por difracción de Rayos-X del correceptor.^[125] La aplicación de estas herramientas permite realizar un diseño y selección racional de compuestos anti-VIH.

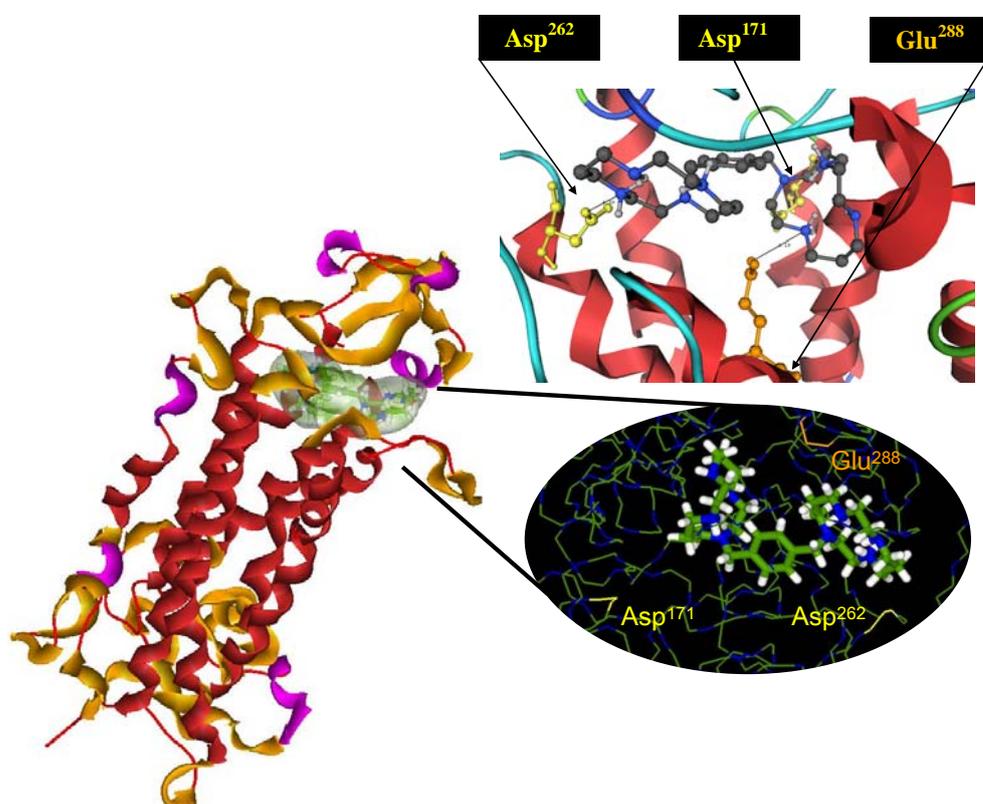
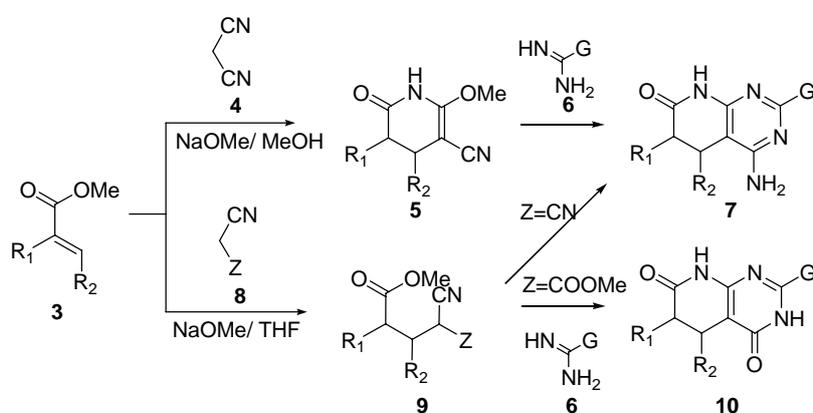


Figura 0.19. Estructura del correceptor CXCR4 y su interacción con el AMD3100 donde se han destacado los aminoácidos Asp¹⁷¹, Asp²⁶² y Glu²⁸⁸.^[125]

Otro trabajo llevado a cabo en el laboratorio de Síntesis del GEM, englobado dentro del proyecto del sida, concretamente en la síntesis de sistemas heterocíclicos inhibidores del correceptor de entrada CXCR4, se basa en la obtención de sistemas bis(pirido[2,3-d]pirimidínicos) con un espaciador *p*-fenilenbismetilénico **12**,^[126] ya que en el grupo se dispone de experiencia en la síntesis de estos sistemas pirido[2,3-d]pirimidínicos **7** y **10** (Figura 0.20). También se estudia la obtención de compuestos donde el espaciador *p*-fenilenbismetilénico está monosustituido por un sistema pirido[2,3-d]pirimidínico para mejorar la solubilidad respecto a los sistemas disustituidos y por analogía a los monociclamos anteriormente descritos.

a)



b)

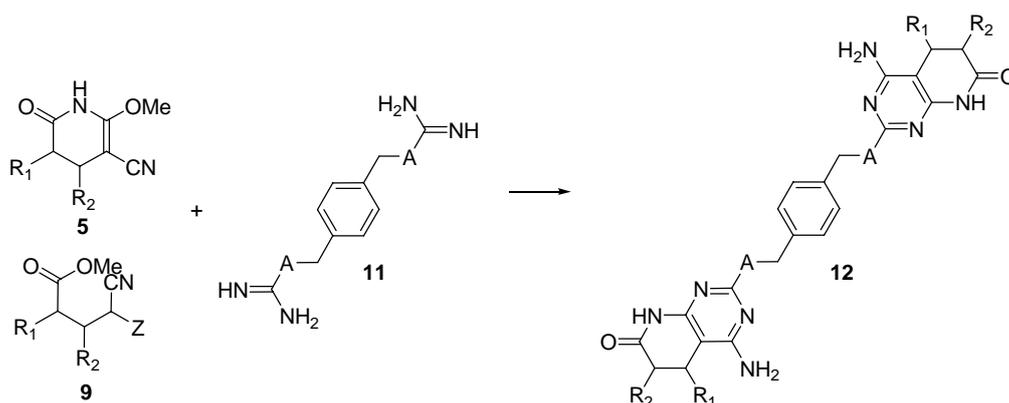


Figura 0.20. a) Ruta sintética para la obtención de sistemas pirido[2,3-d]pirimidínicos **7** y **10**. b) Síntesis de sistemas bis(pirido[2,3-d]pirimidínicos) **12**, donde A = NH, S; $R_1 = H, Me, Ph$ y $R_2 = H, Me$.

Enmarcado también dentro del proyecto de investigación sobre el diseño, síntesis y evaluación de la actividad antirretroviral de nuevos compuestos anti-VIH del laboratorio de Síntesis del GEM, con el objetivo sintetizar compuestos capaces de inhibir la etapa de transcripción inversa, se ha desarrollado una ruta sintética para obtener análogos de HEPT **23**, 1-[(2-hidroxietoxi)metil]-6-(tiofenil)timina, ver Figura 0.21.^[127,128] El HEPT se sintetizó por primera vez en 1989 por Tanaka y colaboradores^[129,130] para actuar como inhibidor competitivo del sustrato natural de la enzima RT.

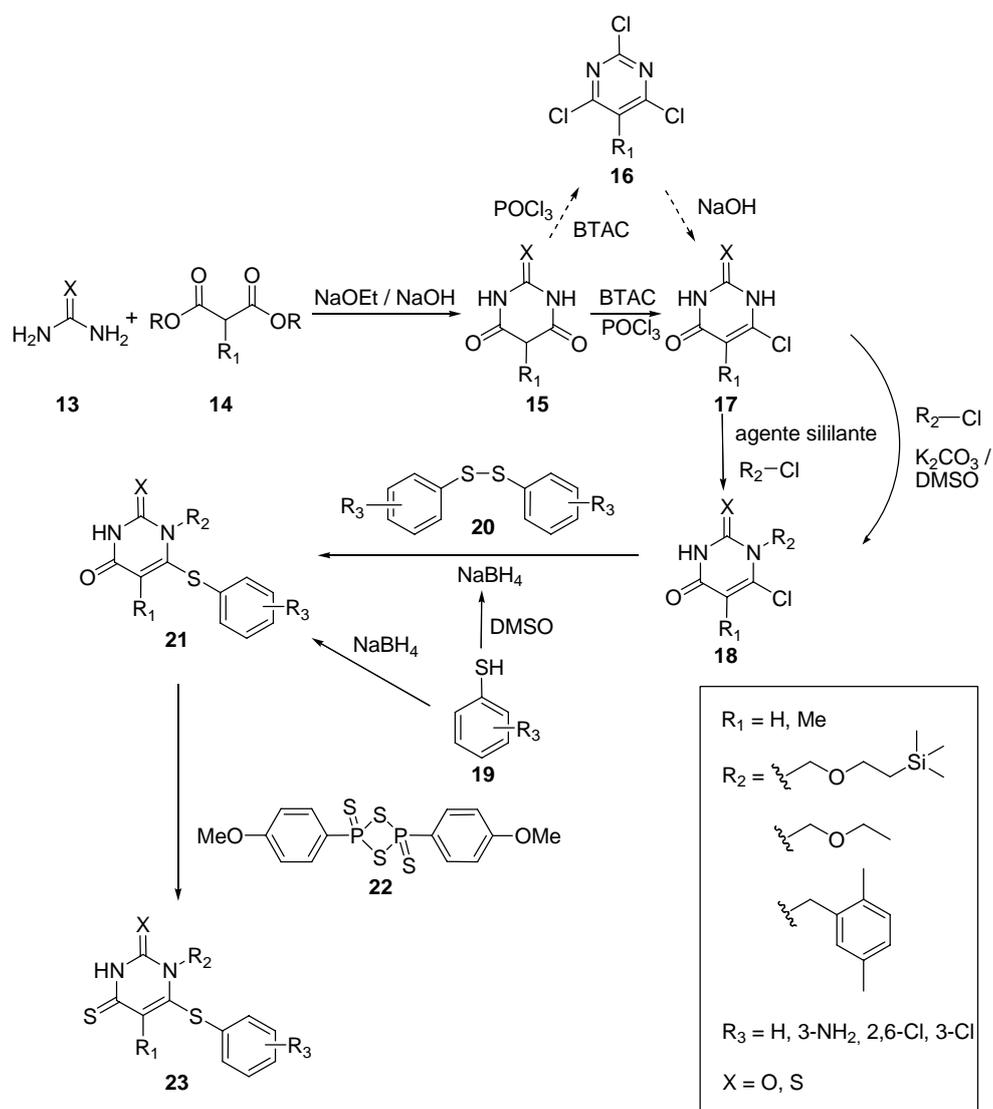


Figura 0.21. Ruta sintética desarrollada por el Sr. Puig de la Bellacasa para obtener análogos de HEPT **23**. La estructura del HEPT corresponde a la estructura general **23** donde $\text{X} = \text{O}$, $\text{R}^1 = \text{Me}$, $\text{R}^2 = \text{CH}_2\text{OCH}_2\text{CH}_2\text{OH}$ y $\text{R}^3 = \text{H}$.^[128]

El presente trabajo se centra en la aplicación de las técnicas de computacionales de cribado virtual y de selección de compuestos de quimiotecas combinatorias para la síntesis de análogos del AMD3100, concretamente en los siguientes objetivos.

0.6. Objetivos.

- Diseño, selección y síntesis de una quimioteca de nuevos compuestos anti-VIH no macrocíclicos análogos al AMD3100 que actúen previsiblemente como inhibidores del receptor de quimiocinas CXCR4. Análisis SAR de los compuestos sintetizados a partir de sus actividades anti-VIH.
- Diseño, selección y síntesis de monociclamos sustituidos como nuevos inhibidores de la unión del VIH-1 al correceptor de entrada del virus CXCR4.
- Cálculo y validación de modelos QSAR a partir de los valores de actividad anti-VIH de compuestos previamente sintetizados como inhibidores de la unión del virus a CXCR4. Uso de estos modelos para predecir la actividad de nuevos compuestos anti-VIH de la quimioteca diseñada inicialmente.
- Desarrollo y validación de modelos farmacofóricos de inhibidores de la unión del VIH-1 al receptor de quimiocinas CXCR4, y posterior predicción de actividad anti-VIH de nuevos compuestos.

Bibliografía.

- [1] M. S. Gottlieb, R. Schroff, H. M. Schanker, J. D. Weisman, P. T. Fan, R. A. Wolf, A. Saxon; Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *N. Engl. J. Med.* **1981**, *305*, 1425-1431.
- [2] K. C. Davis, C. R. Jr. Horsburgh, U. Hasiba, A. L. Schocket, C. H. Kirkpatrick; Acquired immunodeficiency syndrome in a patient with hemophilia. *Ann. Intern. Med.* **1983**, *98*, 284-286.
- [3] H. Masur, M. A. Michelis, J. B. Greene, I. Onorato, R. A. Stouwe, R. S. Holzman, G. Wormser, L. Brettman, M. Lange, H. W. Murray, S. Cunningham-Rundles; An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction. *N. Engl. J. Med.* **1981**, *305*, 1431-1438.
- [4] B. J. Poiesz, F. W. Ruscetti, A. F. Gazdar, P. A. Bunn, J. D. Minna, R. C. Gallo; Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 7415-7419.
- [5] R. C. Gallo, P. S. Sarin, E. P. Gelmann, M. Robert-Guroff, E. Richardson, V. S. Kalyanaraman, D. Mann, G. D. Sidhu, R. E. Stahl, S. Zolla-Pazner, J. Leibowitch, M. Popovic; Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science* **1983**, *220*, 865-867.
- [6] R. C. Gallo, S. Z. Salahuddin, M. Popovic, G. M. Shearer, M. Kaplan, B. F. Haynes, T. J. Palker, R. Redfield, J. Oleske, B. Safai; Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* **2008**, *224*, 500-503.
- [7] M. Popovic, M. G. Sarngadharan, E. Read, R. C. Gallo; Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science* **1984**, *224*, 497-500.
- [8] F. Barre-Sinoussi, J. C. Chermann, F. Rey, M. T. Nugeyre, S. Chamaret, J. Gruest, C. Dauguet, C. Axler-Blin, F. Vézinet-Brun, C. Rouzioux, W. Rozenbaum, L. Montagnier; Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **1983**, *220*, 868-871.
- [9] J. Coffin, A. Haase, J. A. Levy, L. Montagnier, S. Oroszlan, N. Teich, H. Temin, K. Toyoshima, H. Varmus, P. Vogt; What to call the AIDS virus? *Nature* **1986**, *321*, 10.
- [10] F. Barre-Sinoussi; HIV: A discovery opening the road to novel scientific knowledge and global health improvement (Nobel Lecture). *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 5809-5814.
- [11] L. Montagnier; 25 years after HIV discovery: prospects for cure and vaccine (Nobel Lecture). *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 5815-5826.
- [12] J. B. Alimonti, T. B. Ball, K. R. Fowke; Mechanism of CD4⁺ T lymphocyte cell death in human immunodeficiency virus infection and AIDS. *J. Gen. Virol.* **2003**, *84*, 1649-1661.
- [13] J. M. Costin; Cytopathic mechanism of HIV-1. *Virol. J.* **2007**, *4*, 100-121.

- [14] S. A. Schwartz, M. P. N. Nair; Current concepts in human immunodeficiency virus infection and AIDS. *Clin. Diagn. Lab. Immunol.* **1999**, *6*, 295-305.
- [15] E. A. Berger, P. M. Murphy, J. M. Farber; Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. *Annu. Rev. Immunol.* **1999**, *17*, 657-700.
- [16] Situación de la epidemia del sida. <http://www.unaids.org> (fecha de consulta 10-9-2008)
- [17] W. K. Wang, M. Y. Chen, C. Y. Chuang, K. T. Jeang, L. M. Huang; Molecular biology of human immunodeficiency virus type 1. *J. Microbiol. Immunol. Infect.* **2000**, *33*, 131-140.
- [18] S. M. Kingsman, A. J. Kingsman; The regulation of human immunodeficiency virus type-1 gene expression. *Eur. J. Biochem.* **1996**, *240*, 491-507.
- [19] R. A. Subbramanian, E. A. Cohen; Molecular biology of the human immunodeficiency virus accessory proteins. *J. Virol.* **1994**, *68*, 6831-6835.
- [20] W. A. Haseltine; Molecular biology of the human immunodeficiency virus type 1. *FASEB J.* **1991**, *5*, 2349-2360.
- [21] M. H. Nielsen, F. S. Pedersen, J. Kjems; Molecular strategies to inhibit HIV-1 replication. *Retrovirology* **2005**, *2*, 10.
- [22] S. Nisole, A. Saïb; Early steps of retrovirus replicative cycle. *Retrovirology* **2004**, *1*, 9.
- [23] J. Blanco, B. Bosch, M. T. Fernández-Figueras, J. Barretina, B. Clotet, J. A. Esté; High level of coreceptor-independent HIV transfer induced by contacts between primary CD4 T cells. *J. Biol. Chem.* **2004**, *279*, 51305-51314.
- [24] W. Hübner, G. P. McNerney, P. Chen, B. M. Dale, R. E. Gordon, F. Y. S. Chuang, X. D. Li, D. M. Asmuth, T. Huser, B. K. Chen; Quantitative 3D video microscopy of HIV transfer across T cell virological synapses. *Science* **2009**, *323*, 1743-1747.
- [25] C. Jolly, K. Kashefi, M. Hollinshead, Q. J. Sattentau; HIV-1 cell to cell transfer across an Env-induced, actin-dependent synapse. *J. Exp. Med.* **2004**, *199*, 283-293.
- [26] D. McDonald, L. Wu, S. Bohks, V. KewalRamani, D. Unutmaz, T. Hope; Recruitment of HIV and its receptors to dendritic cell-T cell junctions. *Science* **2003**, *300*, 1295-1297.
- [27] D. M. Phillips; The role of cell-to-cell transmission in HIV infection. *AIDS* **1994**, *8*, 719-731.
- [28] A. Dalgleish, P. Beverley, P. Clapham, D. H. Crawford, M. F. Greaves, R. A. Weiss; The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* **1984**, *312*, 763-767.
- [29] D. Klatzmann, E. Champagne, S. Chamaret, J. Gruest, D. Guetard, T. Hercend, J. C. Gluckman, L. Montagnier; T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature* **1984**, *312*, 767-768.
- [30] R. Bourgeois, J. Mercier, I. Paquette-Brooks, E. A. Cohen; Association between disruption of CD4 receptor dimerization and increased human immunodeficiency virus type I entry. *Retrovirology* **2006**, *3*, 31-38.
- [31] R. W. Doms, D. Trono; The plasma membrane as a combat zone in the HIV battlefield. *Genes Dev.* **2000**, *14*, 2677-2688.
- [32] V. Briz, E. Poveda, V. Soriano; HIV entry inhibitors: mechanisms of action and resistance pathways. *J. Antimicrob. Chemother.* **2006**, *57*, 619-627.

- [33] D. C. Chan, P. S. Kim; HIV entry and its inhibition. *Cell* **1998**, 93, 681-684.
- [34] P. Clapham, A. McNight; Cell surface receptors, virus entry and tropism of primate lentiviruses. *J. Gen. Virol.* **2002**, 83, 1809-1829.
- [35] R. W. Doms, J. P. Moore; HIV-1 membrane fusion: targets of opportunity. *J. Cell Biol.* **2000**, 151, F9-F13.
- [36] C. Gomez, T. J. Hope; The ins and outs of HIV replication. *Cell. Microbiol.* **2005**, 7, 621-626.
- [37] C. C. LaBranche, G. Galasso, J. P. Moore, D. P. Bolognesi, M. S. Hirsch, S. M. Hammer; HIV fusion and its inhibition. *Antiviral Res.* **2001**, 50, 95-115.
- [38] J. Sodroski; HIV-1 entry inhibitors in the side pocket. *Cell* **1999**, 99, 243-246.
- [39] R. Wyatt, J. Sodroski; The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. *Science* **1998**, 280, 1884-1888.
- [40] C. C. Bleul, M. Farzan, H. Choe, C. Parolin, I. Clark-Lewis, J. Sodroski, T. A. Springer; The lymphocyte chemoattractant SDF-1 is a ligand for LESTR/fusin and block HIV-1 entry. *Nature* **1996**, 382, 829-833.
- [41] Y. Feng, C. C. Broder, P. E. Kennedy, E. A. Berger; HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **1996**, 272, 872-876.
- [42] E. Oberlin, A. Amara, F. Bachelierie, C. Bessia, J. L. Virelizier, F. Arenzana-Seisdedos, O. Schwartz, J. M. Heard, I. Clark-Lewis, D. F. Legler, M. Loetscher, M. Baggiolini, B. Moser; The CXC chemokine SDF-1 is the ligand for LESTR/fusin and prevents infection by T-cell-line-adapted HIV-1. *Nature* **1996**, 382, 833-835.
- [43] G. Alkhatib, C. Combadiere, C. C. Broder, Y. Feng, P. E. Kennedy, P. M. Murphy, E. A. Berger; CC CKR5: a RANTES, MIP-1 α , MIP-1 β receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science* **1996**, 272, 1955-1958.
- [44] H. Deng, R. Liu, W. Willmeier, S. Choe, D. Unutmaz, M. Burkhart, P. Di Marzio, S. Marmon, R. E. Sutton, C. M. Hill, C. B. Davis, S. C. Peiper, T. J. Schall, D. R. Littman, N. R. Landau; Identification of a major co-receptor for primary isolates of HIV-1. *Nature* **1996**, 381, 661-666.
- [45] T. Dragic, V. Litwin, G. P. Allaway, S. R. Martin, Y. Huang, K. A. Nagashima, C. Cayanan, P. J. Maddon, R. A. Koup, J. P. Moore, W. A. Paxton; HIV-1 entry into CD4⁺ cells is mediated by the chemokine receptor CC-CKR-5. *Nature* **1996**, 381, 667-673.
- [46] A. Trkola, T. Dragic, J. Arthos, J. M. Binley, W. C. Olson, G. P. Allaway, C. Cheng-Mayer, J. Robinson, P. J. Maddon, J. P. Moore; CD4-dependent, antibody-sensitive interactions between HIV-1 and its co-receptor CCR-5. *Nature* **1996**, 384, 184-187.
- [47] J. M. Busillo, J. L. Benovic; Regulation of CXCR4 signalling. *Biochim. Biophys. Acta* **2007**, 1768, 952-963.
- [48] A. Ben-Baruch, D. F. Michiel, J. J. Oppenheim; Signals and receptors involved in recruitment of inflammatory cells. *J. Biol. Chem.* **1995**, 270, 11703-11706.
- [49] S. G. Ward, J. Westwick; Chemokines: understanding their role in T-lymphocyte biology. *Biochem. J.* **1998**, 333, 457-470.

- [50] J. C. Stephens, D. E. Reich, D. B. Goldstein, H. D. Shin, M. W. Smith, M. Carrington, C. Winkler, G. A. Huttley, R. Allikmets, L. Schriml, B. Gerrard, M. Malasky, M. D. Ramos, S. Morlot, M. Tzetis, C. Oddoux, F. S. di Giovine, G. Nasioulas, D. Chandler, M. Aseev, M. Hanson, L. Kalaydjieva, D. Glavac, P. Gasparini, E. Kanavakis, M. Claustres, M. Kambouris, H. Ostrer, G. Duff, V. Baranov, H. Sibul, A. Metspalu, D. Goldman, N. Martin, D. Duffy, J. Schmidtke, X. Estivill, S. J. O'Brien, M. Dean; Dating the origin of the CCR5-Delta32 AIDS-resistance allele by the coalescence of haplotypes. *Am. J. Hum. Genet.* **1998**, *62*, 1507-1515.
- [51] D. S. Dimitrov, X. Xiao, D. J. Chabot, C. C. Broder; HIV coreceptors. *J. Membrane Biol.* **1998**, *166*, 75-90.
- [52] M. M. Goodenow, R. G. Collman; HIV-1 coreceptor preference is distinct from target cell tropism: a dual-parameter nomenclature to define viral phenotypes. *J. Leukoc. Biol.* **2006**, *80*, 965-972.
- [53] J. P. Moore, A. Trkola, T. Dragic; Co-receptors for HIV-1 entry. *Curr. Opin. Immunol.* **1997**, *9*, 551-562.
- [54] P. Lusso; HIV and the chemokine system: 10 years later. *EMBO J.* **2006**, *25*, 447-456.
- [55] T. Dragic; An overview of the determinants of CCR5 and CXCR4 co-receptor function. *J. Gen. Virol.* **2001**, *82*, 1807-1814.
- [56] B. D. Preston, B. J. Poiesz, L. A. Loeb; Fidelity of HIV-1 reverse transcriptase. *Science* **1988**, *242*, 1168-1171.
- [57] A. S. Perelson, A. U. Neumann, M. Markowitz, J. M. Leonard, D. D. Ho; HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* **1996**, *271*, 1582-1586.
- [58] B. F. Keele, F. Van Heuverswyn, Y. Li, E. Bailes, J. Takehisa, M. L. Santiago, F. Bibollet-Ruche, Y. Chen, L. V. Wain, F. Liegeois, S. Loul, E. M. Ngole, Y. Bienvenue, E. Delaporte, J. F. Brookfield, P. M. Sharp, G. M. Shaw, M. Peeters, B. H. Hahn; Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science* **2006**, *313*, 523-526.
- [59] B. Korber, M. Muldoon, J. Theiler, F. Gao, R. Gupta, A. Lapedes, B. H. Hahn, S. Wolinsky, T. Bhattacharya; Timing the ancestor of the HIV-1 pandemic strains. *Science* **2000**, *288*, 1789-1796.
- [60] M. Worobey, M. Gemmel, D. E. Teuwen, T. Haselkorn, K. Kunstman, M. Bunce, J. J. Muyembe, J. M. Kabongo, R. M. Kalengayi, E. Van Marck, M. T. Gilbert, S. M. Wolinsky; Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature* **2008**, *455*, 661-664.
- [61] F. Van Heuverswyn, Y. Li, C. Neel, E. Bailes, B. F. Keele, W. Liu, S. Loul, C. Butel, F. Liegeois, Y. Bienvenue, E. M. Ngole, P. M. Sharp, G. M. Shaw, E. Delaporte, B. H. Hahn, M. Peeters; Human immunodeficiency viruses: SIV infection in wild gorillas. *Nature* **2006**, *444*, 164.
- [62] J. D. Reeves, R. W. Doms; Human immunodeficiency virus type 2. *J. Gen. Virol.* **2002**, *83*, 1253-1265.
- [63] H. I. Requejo; Worldwide molecular epidemiology of HIV. *Rev. Saúde Pública* **2006**, *40*, 331-345.
- [64] L. Buonaguro, M. L. Tornesello, F. M. Buonaguro; Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. *J. Virol.* **2007**, *81*, 10209-10219.

- [65] Drugs used in the treatment of HIV infection. www.fda.gov (fecha de consulta 7-9-2009)
- [66] E. de Clercq; Molecular targets for antiviral agents. *J. Pharmacol. Exp. Ther.* **2001**, *297*, 1-10.
- [67] E. de Clercq; Emerging anti-HIV drugs. *Expert Opin. Emerging Drugs* **2005**, *10*, 241-274.
- [68] L. D. Madhu Vajpayee; Antiretroviral drugs in HIV infection. *Indian J. Pharmacol.* **1999**, *31*, 96-103.
- [69] H. Mitsuya, K. J. Weinhold, P. A. Furman, M. H. St Clair, S. N. Lehrman, R. C. Gallo, D. Bolognesi, D. W. Barry, S. Broder; 3'-Azido-3'-deoxythymidine (BW A509U): an antiviral agent that inhibits the infectivity and cytopathic effect of human T-lymphotropic virus type III/lymphadenopathy-associated virus in vitro. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 7096-7100.
- [70] E. de Clercq; New approaches toward anti-HIV chemotherapy. *J. Med. Chem.* **2005**, *48*, 1297-1313.
- [71] J. S. Cervia, M. A. Smith; Enfuvirtide (T-20): a novel human immunodeficiency virus type 1 fusion inhibitor. *Clin. Infect. Dis.* **2008**, *37*, 1102-1106.
- [72] C. Derdeyn, J. M. Decker, J. N. Sfakianos, X. Wu, W. A. O'Brien, L. Ratner, J. Kappes, G. M. Shaw, E. Hunter; Sensitivity of human immunodeficiency virus type 1 to the fusion inhibitor T-20 is modulated by coreceptor specificity defined by the V3 loop of gp120. *J. Virol.* **2000**, *74*, 8358-8367.
- [73] S. Liu, H. Lu, J. Niu, Y. Xu, S. Wu, S. Jiang; Different from the HIV fusion inhibitor C34, the anti-HIV drug Fuzeon (T-20) inhibits HIV-1 entry by targeting multiple sites in gp41 and gp120. *J. Biol. Chem.* **2005**, *280*, 11259-11273.
- [74] G. Moyle; Stopping HIV fusion with enfuvirtide: the first step to extracellular HAART. *J. Antimicrob. Chemother.* **2003**, *51*, 213-217.
- [75] S. Veiga, S. Henriques, N. C. Santos, M. Castanho; Putative role of membranes in the HIV fusion inhibitor enfuvirtide mode of action at the molecular level. *Biochem. J.* **2004**, *377*, 107-110.
- [76] P. Dorr, M. Westby, S. Dobbs, P. Griffin, B. Irvine, M. Macartney, J. Mori, G. Rickett, C. Smith-Burchnell, C. Napier, R. Webster, D. Armour, D. Price, B. Stammen, A. Wood, M. Perros; Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity. *Antimicrob. Agents Chemother.* **2005**, *49*, 4721-4732.
- [77] D. K. Walker, S. Abel, P. Comby, G. J. Muirhead, A. N. R. Nedderman, D. A. Smith; Species differences in the disposition of the CCR5 antagonist, UK-427,857, a new potential treatment for HIV. *Drug. Metab. Dispos.* **2005**, *33*, 587-595.
- [78] V. Summa, A. Petrocchi, F. Bonelli, B. Crescenczi, M. Donghi, M. Ferrara, F. Fiore, C. Gardelli, O. Gonzalez Paz, D. J. Hazuda, P. Jones, O. Kinzel, R. Laufer, E. Monteagudo, E. Muraglia, E. Nizi, F. Orvieto, P. Pace, G. Pescatore, R. Scarpelli, K. Stillmock, M. V. Witmer, M. Rowley; Discovery of raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection. *J. Med. Chem.* **2008**, *51*, 5843-5855.

- [79] X. Q. Zhang, M. Sorensen, M. Fung, R. T. Schooley; Synergistic in vivo antiretroviral activity of a humanized monoclonal anti-CD4 antibody (TNX-355) and enfurvitide (T-20). *Antimicrob. Agents Chemother.* **2006**, *50*, 2231-2233.
- [80] D. R. Kuritzkes, J. Jacobson, W. G. Powderly, E. Godofsky, E. DeJesus, F. Haas, K. A. Reimann, J. L. Larson, P. O. Yarbough, V. Curt, W. R. Jr. Shanahan; Antiretroviral activity of the anti-CD4 monoclonal antibody TNX-355 in patients infected with HIV type 1. *J. Infect. Dis.* **2004**, *189*, 286-291.
- [81] A. Trkola, T. J. Ketas, K. A. Nagashima, L. Zhao, T. Cilliers, L. Morris, J. P. Moore, P. J. Maddon, W. C. Olson; Potent, broad-spectrum inhibition of human immunodeficiency virus type 1 by the CCR5 monoclonal antibody PRO 140. *J. Virol.* **2001**, *75*, 579-588.
- [82] J. M. Jacobson, M. S. Saag, M. A. Thompson, M. A. Fischl, R. Liporace, R. C. Reichman, R. R. Redfield, C. J. Fichtenbaum, B. S. Zingman, M. C. Patel, J. D. Murga, S. M. Pemrick, P. D'Ambrosio, M. Michael, H. Kroger, H. Ly, Y. Rotshteyn, R. Buice, S. A. Morris, J. J. Stavola, P. J. Maddon, A. B. Kremer, W. C. Olson; Antiviral activity of single-dose PRO 140, a CCR5 monoclonal antibody, in HIV-infected adults. *J. Infect. Dis.* **2008**, *198*, 1345-1352.
- [83] J. M. Strizki, C. Tremblay, S. Xu, L. Wojcik, N. Wagner, W. Gonsiorek, R. W. Hipkin, C. C. Chou, C. Pugliese-Sivo, Y. Xiao, J. R. Tagat, K. Cox, T. Priestley, S. Sorota, W. Huang, M. Hirsch, G. R. Reyes, B. M. Baroudy; Discovery and characterization of vicriviroc (SCH 417690), a CCR5 antagonist with potent activity against human immunodeficiency virus type 1. *Antimicrob. Agents Chemother.* **2005**, *49*, 4911-4919.
- [84] J. R. Tagat, S. W. McCombie, D. Nazareno, M. A. Labroli, Y. Xiao, R. W. Steensma, J. M. Strizki, B. M. Baroudy, K. Cox, J. Lachowich, G. Varty, R. Watkins; Piperazine-based CCR5 anagonists as HIV-1 inhibitors. IV. Discovery of 1-[(4,6-dimethyl-5-pyrimidinyl)carbonyl]-4-[4-[2-methoxy-1(R)-4-(trifluoromethyl)phenyl]ethyl]-3(S)-methyl-1-piperazinyl]-4-methylpiperidine (Sch-417690/Sch-D), a potent, highly selective, and orally bioavailable CCR5 antagonist. *J. Med. Chem.* **2004**, *47*, 2405-2408.
- [85] J. B. Crawford, G. Chen, D. Gauthier, T. Wilson, B. Carpenter, I. R. Baird, E. McEachem, A. Kaller, C. Harwig, B. Atsma, R. Skerlj, G. J. Bridger; AMD070, a CXCR4 chemokine receptor antagonist: practical large-scale laboratory synthesis. *Org. Process Res. Dev.* **2008**, *12*, 823-830.
- [86] N. D. Stone, S. B. Dunaway, C. Flexner, C. Tierney, G. B. Calandra, S. Becker, Y. J. Cao, I. P. Wiggins, J. Conley, R. T. MacFarland, J. G. Park, C. Lalama, S. Snyder, B. Kallungal, K. L. Klingman, C. W. Hendrix; Multiple-dose escalation study of the safety, pharmacokinetics, and biologic activity of oral AMD070, a selective CXCR4 receptor inhibitor, in human subjects. *Antimicrob. Agents Chemother.* **2007**, *51*, 2351-2358.
- [87] K. Shimura, E. Kodama, Y. Sakagami, Y. Matsuzaki, W. Watanabe, Y. Yamataka, Y. Ohata, S. Doi, M. Sato, M. Kano, S. Ikeda, M. Matsuoka; Broad antiretroviral activity and resistance profile of the novel human immunodeficiency virus integrase inhibitor elvitegravir (JTK-303/GS-9137). *J. Virol.* **2008**, *82*, 764-774.
- [88] P. A. Janssen, P. J. Lewi, E. Arnold, F. Daeyaert, M. de Jonge, J. Heeres, L. Koymans, M. Vinkers, J. Guillemont, E. Pasquier, M. Kukla, D. Ludovici, K. Andries, M. P. de Béthune, R. Pauwels, K. Das, A. D. Jr. Clark, Y. V. Frenker, S. H. Hughes, B. Medaer, F. De Knaep, H. Bohets, F. De Clerck, A. Lampo, P. Williams, P. Stoffels; In search of a novel anti-HIV drug: multidisciplinary coordination in the discovery of 4-[[4-[[4-[(1E)-2-cyanoethenyl]-2,6-dimethylphenyl]amino]-2-pyrimidinyl]amino]benzotrile (R278474, rilpivirine). *J. Med. Chem.* **2005**, *48*, 1901-1909.

- [89] D. Ripamonti, F. Maggiolo; Rilpivirine, a non-nucleoside reverse transcriptase inhibitor for the treatment of HIV infection. *Curr. Opin. Investig. Drugs* **2008**, *9*, 899-912.
- [90] R. C. Bethell, Y. S. Lie, N. T. Parkin; In vitro activity of SPD754, a new deoxyxytidine nucleoside reverse transcriptase inhibitor (NRTI), against 215 HIV-1 isolates resistant to other NRTIs. *Antivir. Chem. Chemother.* **2005**, *16*, 295-302.
- [91] Z. Gu, B. Allard, J. M. de Muys, J. Lippens, R. F. Rando, N. Nguyen-Ba, C. Ren, P. McKenna, D. L. Taylor, R. C. Bethell; In vitro antiretroviral activity and in vitro toxicity profile of SPD754, a new deoxyxytidine nucleoside reverse transcriptase inhibitor for treatment of human immunodeficiency virus infection. *Antimicrob. Agents Chemother.* **2006**, *50*, 625-631.
- [92] M. A. Wainberg, P. Cahn, R. C. Bethell, J. Sawyer, S. Cox; Apricitabine: a novel deoxycytidine analogue nucleoside reverse transcriptase inhibitor for the treatment of nucleoside-resistant HIV infection. *Antivir. Chem. Chemother.* **2007**, *18*, 61-70.
- [93] T. S. Lin, M. Z. Luo, M. C. Liu, Y. L. Zhu, E. Gullen, G. E. Dutschman, Y. C. Cheng; Design and synthesis of 2',3'-dideoxy-2',3'-didehydro-beta-L-cytidine (beta-L-d4C) and 2',3'-dideoxy 2',3'-didehydro-beta-L-5-fluorocytidine (beta-L-Fd4C), two exceptionally potent inhibitors of human hepatitis B virus (HBV) and potent inhibitors of human immunodeficiency virus (HIV) in vitro. *J. Med. Chem.* **1996**, *39*, 1757-1759.
- [94] J. Shi, J. J. McAtee, S. Schlueter Wirtz, P. Tharnish, A. Juodawlkis, D. C. Liotta, R. F. Schinazi; Synthesis and biological evaluation of 2',3'-didehydro-2',3'- dideoxy-5-fluorocytidine (D4FC) analogues: discovery of carbocyclic nucleoside triphosphates with potent inhibitory activity against HIV-1 reverse transcriptase. *J. Med. Chem.* **1999**, *42*, 859-867.
- [95] F. Reuss, M. Kulke, J. Braspenning, B. Heckl-Östreicher, H. G. Opitz; Fosaltudine tidoxil, a novel alovudine-derived prodrug is activated and inhibits the replication of HIV-1 in human PBMC. *Int. Conf. AIDS* **2006**, *16*, Abstract No. ThPE0025.
- [96] K. S. Harris, W. Brabant, S. Styrchak, A. Gall, R. Daifuku; KP-1212/1461, a nucleoside designed for the treatment of HIV by viral mutagenesis. *Antiviral Res.* **2005**, *67*, 1-9.
- [97] C. Herzmann, K. Arastèh, R. L. Murphy, H. Schulbin, P. Kreckel, D. Drauz, R. F. Schinazi, A. Beard, L. Cartee, M. J. Otto; Safety, pharmacokinetics, and efficacy of (+/-)-beta-2',3'-dideoxy-5-fluoro-3'-thiacytidine with efavirenz and stavudine in antiretroviral-naïve human immunodeficiency virus-infected patients. *Antimicrob. Agents Chemother.* **2005**, *49*, 2828-2833.
- [98] E. de Clercq, N. Yamamoto, R. Pauwels, M. Baba, D. Schols, H. Nakashima, J. Balzarini, Z. Debyser, B. A. Murrer, D. Schwartz, D. Thornton, G. Bridger, S. Fricker, G. Henson, M. Abrams, D. Picker; Potent and selective inhibition of human immunodeficiency virus (HIV)-1 and HIV-2 replication by a class of bicyclams interacting with a viral uncoating event. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 5286-5290.
- [99] J. Balzarini, D. Schols, J. Neyts, E. Van Damme, W. Peumans, E. de Clercq; Alpha-(1-3)- and alpha-(1-6)-D-mannose-specific plant lectins are markedly inhibitory to human immunodeficiency virus and cytomegalovirus infections in vitro. *Antimicrob. Agents Chemother.* **1991**, *35*, 410-416.
- [100] R. W. Jansen, G. Molema, R. Pauwels, D. Schols, E. de Clercq, D. K. Meijer; Potent in vitro anti-human immunodeficiency virus-1 activity of modified human serum albumins. *Mol. Pharmacol.* **1991**, *39*, 818-823.
- [101] T. Matsui, S. Kobayashi, O. Yoshida, S. Ishii, Y. Abe, N. Yamamoto; Effects of succinylated concanavalin A on infectivity and syncytial formation of human immunodeficiency virus. *Med. Microbiol. Immunol.* **1990**, *179*, 225-235.

- [102] E. de Clercq, N. Yamamoto, R. Pauwels, J. Balzarini, M. Witvrouw, K. de Vreese, Z. Debyser, B. Rosenwirth, P. Peichl, R. Datema, D. Thornton, R. Skerlj, F. Gaul, S. Padmanabhan, G. Bridger, G. Henson, M. Abrams; Highly potent and selective inhibition of human immunodeficiency virus by the bicyclam derivative JM3100. *Antimicrob. Agents Chemother.* **1994**, *38*, 668-674.
- [103] G. J. Bridger, R. T. Skerlj, D. Thornton, S. Padmanabhan, S. A. Martellucci, G. W. Henson, M. J. Abrams, N. Yamamoto, K. de Vreese, R. Pauwels, E. de Clercq; Synthesis and structure-activity relationships of phenylenebis(methylene)-linked bis-tetraazamacrocycles that inhibit HIV replication. Effects of macrocyclic ring size and substituents on the aromatic linker. *J. Med. Chem.* **1995**, *38*, 366-378.
- [104] G. J. Bridger, R. T. Skerlj, S. Padmanabhan, S. A. Martellucci, G. W. Henson, M. J. Abrams, H. C. Joao, M. Witvrouw, K. de Vreese, R. Pauwels, E. de Clercq; Synthesis and structure-activity relationships of phenylenebis(methylene)-linked bis-tetraazamacrocycles that inhibit human immunodeficiency virus replication. 2. Effect of heteroaromatic linkers on the activity of bicyclams. *J. Med. Chem.* **1996**, *39*, 109-119.
- [105] G. J. Bridger, R. T. Skerlj, S. Padmanabhan, S. A. Martellucci, G. W. Henson, S. Struyf, M. Witvrouw, D. Schols, E. de Clercq; Synthesis and structure-activity relationships of phenylenebis(methylene)-linked bis-azamacrocycles that inhibit HIV-1 and HIV-2 replication by antagonism of the chemokine receptor CXCR4. *J. Med. Chem.* **1999**, *42*, 3971-3981.
- [106] J. A. Esté, C. Cabrera, E. de Clercq, S. Struyf, J. Van Damme, G. Bridger, R. T. Skerlj, M. J. Abrams, G. Henson, A. Gutierrez, B. Clotet, D. Schols; Activity of different bicyclam derivatives against human immunodeficiency virus depends on their interaction with the CXCR4 chemokine receptor. *Mol. Pharmacol.* **1999**, *55*, 67-73.
- [107] L. O. Gerlach, J. S. Jacobsen, K. P. Jensen, M. M. Rosenkilde, R. Skerlj, U. Ryde, G. J. Bridger, T. W. Schwartz; Metal ion enhanced binding of AMD3100 to Asp(262) in the CXCR4 receptor. *Biochemistry* **2003**, *42*, 710-717.
- [108] S. Hatse, K. Princen, E. de Clercq, M. M. Rosenkilde, T. W. Schwartz, P. E. Hernandez-Abad, R. T. Skerlj, G. J. Bridger, D. Schols; AMD3465, a monomacrocyclic CXCR4 antagonist and potent HIV entry inhibitor. *Biochem. Pharmacol.* **2005**, *70*, 752-761.
- [109] G. K. Reddy, J. Crawford, V. K. Jain; The role of Plerixafor (AMD3100) in mobilizing hematopoietic progenitor cells in patients with hematologic malignancies. *Support Cancer Ther.* **2006**, *3*, 73-76.
- [110] K. Princen, S. Hatse, K. Vermeire, S. Aquaro, E. de Clercq, L. O. Gerlach, M. M. Rosenkilde, T. W. Schwartz, R. Skerlj, G. Bridger, D. Schols; Inhibition of human immunodeficiency virus replication by a dual CCR5/CXCR4 antagonist. *J. Virol.* **2004**, *78*, 12996-13006.
- [111] K. de Vreese, D. Reymen, P. Griffin, A. Steinkasserer, G. Werner, G. J. Bridger, J. Esté, W. James, G. W. Henson, J. Desmyter, J. Anné, E. de Clercq; The bicyclams, a new class of potent human immunodeficiency virus inhibitors, block viral entry after binding. *Antiviral Res.* **1996**, *29*, 209-219.
- [112] D. Schols, S. Struyf, J. Van Damme, J. A. Esté, G. Henson, E. de Clercq; Inhibition of T-tropic HIV strains by selective antagonization of the chemokine receptor CXCR4. *J. Exp. Med.* **1997**, *186*, 1383-1388.
- [113] D. Schols, J. A. Esté, G. Henson, E. de Clercq; Bicyclams, a class of potent anti-HIV agents, are targeted at the HIV coreceptor fusin/CXCR-4. *Antiviral Res.* **1997**, *35*, 147-156.

- [114] B. Labrosse, A. Brelot, N. Heveker, N. Sol, D. Schols, E. de Clercq, M. Alizon; Determinants for sensitivity of human immunodeficiency virus coreceptor CXCR4 to the bicyclam AMD3100. *J. Virol.* **1998**, *72*, 6381-6388.
- [115] J. A. Esté, C. Cabrera, J. Blanco, A. Gutierrez, G. Bridger, G. Henson, B. Clotet, D. Schols, E. de Clercq; Shift of clinical human immunodeficiency virus type 1 isolates from X4 to R5 and prevention of emergence of the syncytium-inducing phenotype by blockade of CXCR4. *J. Virol.* **1999**, *73*, 5577-5585.
- [116] C. Dealwis, E. J. Fernandez, D. A. Thompson, R. J. Simon, M. A. Siani, S. Loul; Crystal structure of chemically synthesized [N33A] stromal cell-derived factor 1alpha, a potent ligand for the HIV-1 "fusin" coreceptor. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6941-6946.
- [117] L. O. Gerlach, R. Skerlj, G. Bridger, T. W. Schwartz; Molecular interactions of cyclam and bicyclam non-peptide antagonists with the CXCR4 chemokine receptor. *J. Biol. Chem.* **2001**, *276*, 14153-14160.
- [118] S. Hatse, K. Princen, L. O. Gerlach, G. Bridger, G. Henson, E. de Clercq, T. W. Schwartz, D. Schols; Mutation of Asp(171) and Asp(262) of the chemokine receptor CXCR4 impairs its coreceptor function for human immunodeficiency virus-1 entry and abrogates the antagonistic activity of AMD3100. *Mol. Pharmacol.* **2001**, *60*, 164-173.
- [119] M. M. Rosenkilde, L. O. Gerlach, J. S. Jakobsen, R. Skerlj, G. J. Bridger, T. W. Schwartz; Molecular mechanism of AMD3100 antagonism in the CXCR4 receptor. *J. Biol. Chem.* **2004**, *279*, 3033-3041.
- [120] M. M. Rosenkilde, L. O. Gerlach, S. Hatse, R. Skerlj, D. Schols, G. J. Bridger, T. W. Schwartz; Molecular mechanism of action of monocyclam versus bicyclam non-peptide antagonists in the CXCR4 chemokine receptor. *J. Biol. Chem.* **2007**, *282*, 27354-27365.
- [121] R. Pascual; Estudios sobre la metodología computacional para el diseño de bibliotecas combinatorias: Desarrollo del programa Pralins y aplicaciones en el campo del HIV. Tesis doctoral, IQS (Universitat Ramon Llull), Barcelona, **2003**.
- [122] R. Pascual, J. I. Borrell, J. Teixido; Analysis of selection methodologies for combinatorial library design. *Mol. Divers.* **2003**, *6*, 121-133.
- [123] V. I. Pérez-Nueno; Herramientas de cribado virtual aplicadas a inhibidores de entrada del VIH. Diseño de nuevos compuestos anti-VIH. Tesis doctoral, IQS (Universitat Ramon Llull), Barcelona, **2009**.
- [124] M. O. Rabal; Herramientas de cribado virtual aplicadas a inhibidores de tirosina quinasas. Contribución al desarrollo del programa PRALINS para el diseño de quimiotecas combinatorias. Tesis doctoral, IQS (Universitat Ramon Llull), Barcelona, **2006**.
- [125] V. I. Pérez-Nueno, D. W. Ritchie, O. Rabal, R. Pascual, J. I. Borrell, J. Teixidó; Comparison of ligand-based and receptor-based virtual screening of HIV entry inhibitors for the CXCR4 and CCR5 receptors using 3D ligand shape matching and ligand-receptor docking. *J. Chem. Inf. Model.* **2008**, *48*, 509-533.
- [126] L. Ros; Quimiotecas de biimidazoles i sistemes polinitrogenats com a inhibidors d'entrada del VIH. Tesis doctoral en curso, IQS (Universitat Ramon Llull), Barcelona, **2009**.
- [127] L. Giménez; Disseny d'una ruta sintètica per obtenir anàlegs HEPT (Part 1). Trabajo final de carrera, IQS (Universitat Ramon Llull), Barcelona, **2005**.

- [128] R. Puig de la Bellacasa; Disseny de quimiotèques d'inhibidors de la transcriptasa inversa anàlegs d'AZT i HEPT. Tesis doctoral en curs, IQS (Universitat Ramon Llull), Barcelona, **2009**.
- [129] M. Baba, H. Tanaka, E. de Clercq, R. Pauwels, J. Balzarini, D. Schols, H. Nakashima, C. F. Perno, R. T. Walker, T. Miyasaka; Highly specific inhibition of human immunodeficiency virus type 1 by a novel 6-substituted acyclouridine derivative. *Biochem. Biophys. Res. Commun.* **1989**, 165, 1375-1381.
- [130] T. Miyasaka, H. Tanaka, M. Baba, H. Hayakawa, R. T. Walker, J. Balzarini, E. de Clercq; A novel lead for specific anti-HIV-1 agents: 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)-thymine. *J. Med. Chem.* **1989**, 32, 2507-2509.

CAPÍTULO 1. ANÁLOGOS POLINITROGENADOS NO CICLÁMICOS DEL AMD3100

1.1. Diseño de una quimioteca de análogos polinitrogenados del AMD3100.

La entrada y la fusión del VIH a la célula se han validado como dianas para la intervención terapéutica contra la infección.^[1,2] Los actuales *leads* con mayor actividad en este ámbito, como el AMD3100 (**24**) (antagonista de CXCR4 de la familia de los biclamos), TAK-779 o SCH-D (antagonistas de CCR5), presentan sistemas heterocíclicos nitrogenados separados por un espaciador aromático o alifático.

Entre todos los compuestos en estudio, los biclamos con el AMD3100 como *lead* representan la familia que muestra una mayor actividad anti-VIH. Sin embargo, este compuesto ha presentado falta de biodisponibilidad oral y cardiotoxicidad relacionados con su elevada carga positiva a pH fisiológico.^[3] Por tanto, es necesario sintetizar una familia de compuestos análogos al AMD3100 que solvente estas desventajas. En el presente trabajo se ha diseñado una biblioteca de poliaminas **25*** que presentan las principales características del AMD3100 (**24**) (Figura 1.1):

- un nitrógeno bencílico a cada lado del *linker* 1,4-fenilenbismetilénico,
- una distancia similar entre los átomos de nitrógeno del anillo de ciclamo del AMD3100 y los átomos de nitrógeno presentes a cada lado del *linker* 1,4-fenilenbismetilénico y los del sistema heterocíclico **26**.

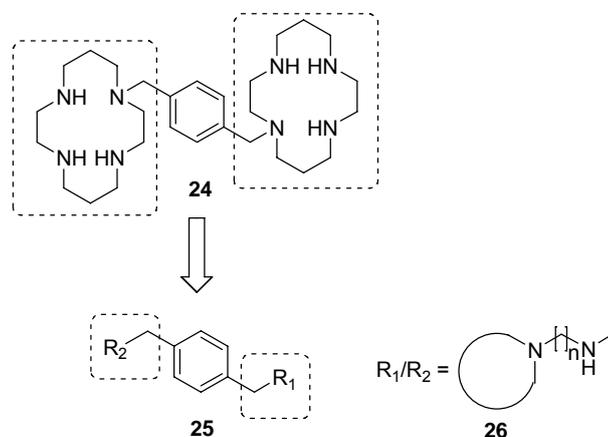


Figura 1.1. Estructura del AMD3100 (**24**) y estructura general de las poliaminas **25** de la quimioteca diseñada.

* Durante el desarrollo del presente trabajo, el grupo de investigación de D.C. Liotta publica un trabajo^[4] que utiliza la misma estrategia de diseño de análogos del AMD3100 de estructura general **25**, con una aproximación más general ya que R_1 y R_2 no se restringen a los sistemas polinitrogenados **26**. El compuesto más activo de su quimioteca ($R_1 = R_2 = 2$ -aminopiridina) presenta mejor inhibición de la unión de la quimiocina SDF-1 al receptor CXCR4 que el AMD3100, sin embargo en ensayos realizados en cultivos celulares para determinar la propagación del VIH resultó ser débilmente activo.

El análisis retrosintético de los compuestos objetivo de estructura general **27** lleva como primer paso a una adición de grupo funcional para obtener las iminas (si $n \neq 0$) o hidrazonas (si $n = 0$) **28** correspondientes; estas hidrazonas también se han incluido en la quimioteca. A continuación se procede a una desconexión del enlace imínico, obteniéndose así el precursor tereftalaldehído (**29**) y dos unidades de sistemas polinitrogenados **26** (iguales si los sistemas están simétricamente sustituidos y diferentes si los sistemas no están simétricamente sustituidos) como se muestra en la Figura 1.2.

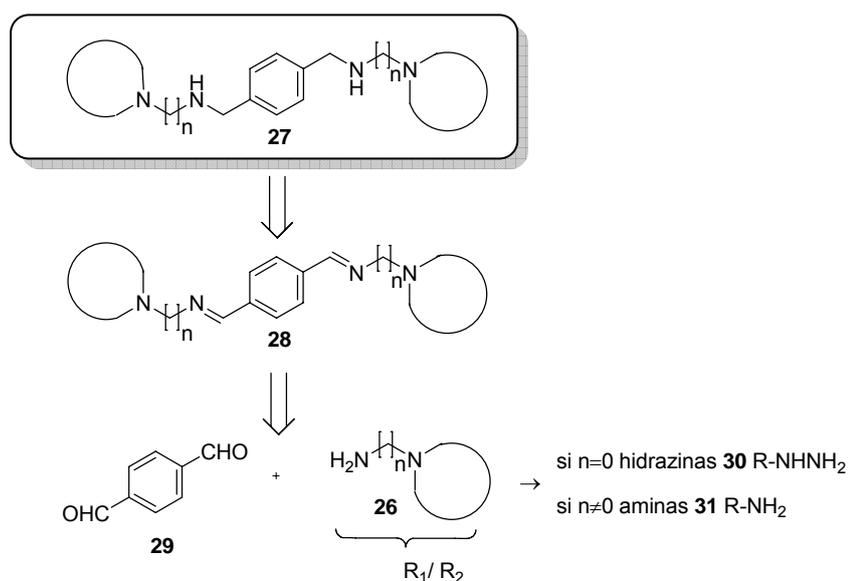


Figura 1.2. Análisis retrosintético de los compuestos de la quimioteca de poliaminas **27** diseñada.

La vía sintética que se utiliza para la obtención de los compuestos de sustitución simétrica **32** y **33** de la quimioteca (cuando $R_1 = R_2$) es la aminación reductora del tereftalaldehído (**29**) con las diferentes aminas **31** o acoplamiento del tereftalaldehído (**29**) con las hidrazinas **30** (Figura 1.3).

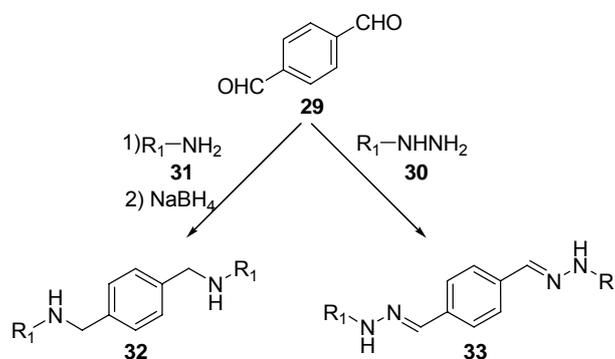


Figura 1.3. Esquema sintético para la obtención de los compuestos simétricos **32** y **33** de la quimioteca.

En el caso de los compuestos de la quimioteca de sustitución no simétrica **32**, **37** y **33** (cuando $R_1 \neq R_2$) se modifica ligeramente la estrategia sintética (Figura 1.4). Las diaminas asimétricas **32** se sintetizan por aminación reductora del 4-(dietoximetil)benzaldehído (**34**) con la amina correspondiente, desprotección del acetal para rendir el intermedio aminobenzaldehído **35** y posterior aminación reductora con otra amina. Las hidrazonas asimétricas **33** se obtienen por condensación del tereftalaldehído (**29**) con la hidrazina correspondiente, separación cromatográfica del producto monosustituido (hidrazonobenzaldehído **36**) y nuevo acoplamiento con otra hidrazina. Los productos del tipo aminohidrazona **37** se pueden obtener a partir del intermedio aminobenzaldehído **35** por acoplamiento con una hidrazina, o a partir del intermedio hidrazonobenzaldehído **36** por reacción de aminación reductora con una amina.

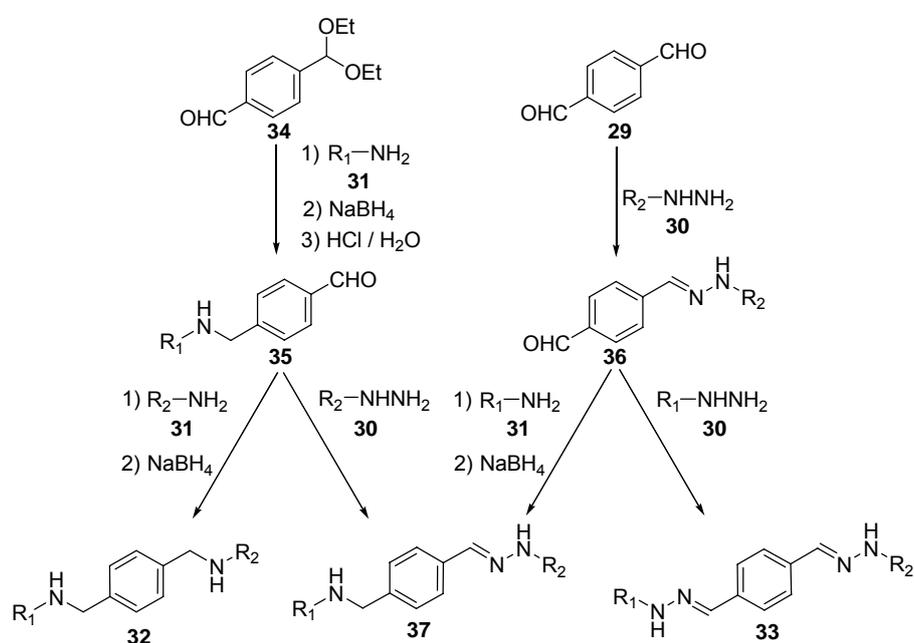


Figura 1.4. Esquema sintético para la obtención de los compuestos no simétricos **32**, **37** y **33** de la quimioteca.

Inicialmente, la selección de *building blocks* polinitrogenados hidrazinas **30** y aminas **31** se basa en la disponibilidad comercial de los compuestos que cumplen las siguientes condiciones:

- presencia de un heterociclo nitrogenado (piperidina, piperazina, morfolina, pirrolidina, imidazol y triazol),
- unión del heterociclo nitrogenado a una cadena alquílica (o sin cadena alquílica en el caso de hidrazinas),
- grupo amino terminal unido a la cadena alquílica (o al heterociclo en el caso de hidrazinas).

El resultado de la búsqueda de compuestos nitrogenados comerciales **30** y **31** que cumplen las condiciones anteriormente detalladas resulta en 18 compuestos formados por los heterociclos piperidina, piperazina, morfolina, pirrolidina, imidazol y triazol, separados por un espaciador $-(CH_2)_n-$ (donde $n = 0, 2, 3$) de un grupo amino terminal, sus estructuras se muestran en la Figura 1.5. No obstante, se prevén potenciales problemas de solubilidad en los disolventes utilizados para realizar los ensayos de actividad anti-VIH (DMSO, etanol y agua) para los *building blocks* **30**{12,17-18} **31**{13-16}. Para comprobarlo, se sintetizaron las correspondientes poliaminas simétricas **32** y **33** y resultaron insolubles, por lo que se excluyeron del diseño. Por otra parte, los *building blocks* utilizados para generar el *core* de las moléculas objetivo son tereftalaldehído (**29**) y 4-(dietoximetil)benzalaldehído (**34**).

Una vez seleccionados los *building blocks* polinitrogenados, 3 hidrazinas **30**{1-3} y 8 aminas **31**{4-11} para la sustitución en las posiciones R_1 y R_2 de la moléculas objetivo **32**, **33** y **37**, se construye y enumera la quimioteca virtual mediante el módulo Combichem del programa Cerius2.^[5]

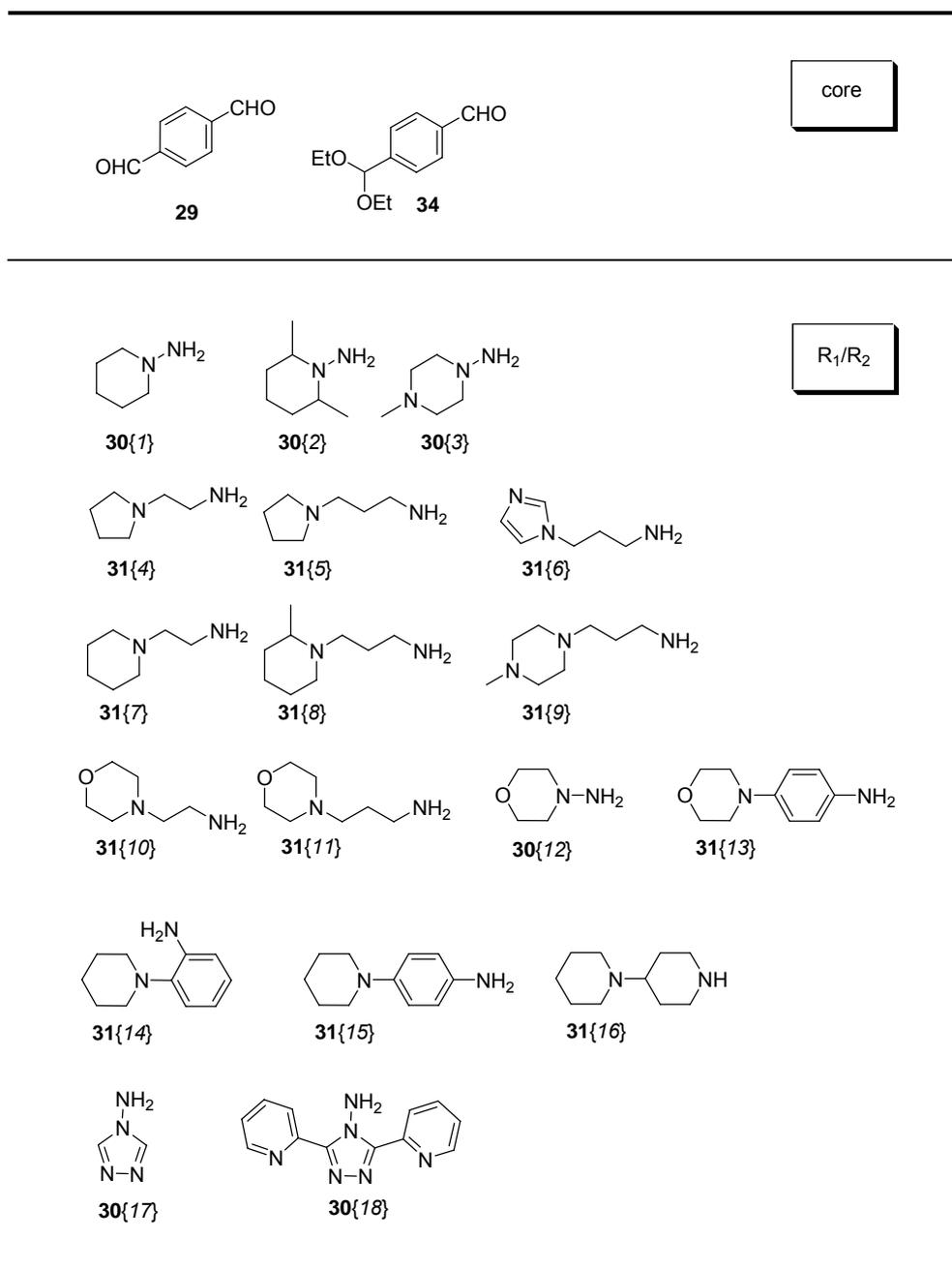
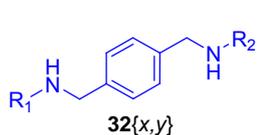


Figura 1.5. Estructura de los building blocks utilizados para generar el core: tereftalaldehído (**29**) y 4-(dietoximetil)benzaldehído (**34**). Estructura de los building blocks utilizados para los grupos poliamínicos formados por aminas **31** e hidrazinas **30**.

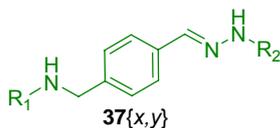
Se obtiene una biblioteca de 11×11 compuestos de estructura general **32**, **33** y **37** que se reduce a 66 moléculas debido a la simetría presente en su sustitución, es decir, un compuesto donde la sustitución es $R_1=30\{1\}$ y $R_2=31\{5\}$ es el mismo que $R_1=31\{5\}$ y $R_2=30\{1\}$. De manera que si se representa la quimioteca en una tabla donde las columnas corresponden a R_1 y las filas a R_2 se pueden suprimir los compuestos por encima o por debajo de la diagonal porque son equivalentes, ver Tabla 1.1.

Tabla 1.1. Representación de los compuestos de la quimioteca de estructura general **32**, **33** y **37**, donde R_1 y R_2 corresponden a los building blocks poliaminados **30** y **31** utilizados. Se indican en rojo las dihidrazonas de estructura general **33**, en verde las aminohidrazonas de estructura general **37** y en azul las poliaminas de estructura general **32**.

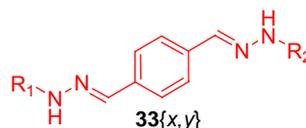
R_1 R_2	1	2	3	4	5	6	7	8	9	10	11
1	1,1										
2	1-2	2-2									
3	1-3	2-3	3-3								
4	1-4	2-4	3-4	4-4							
5	1-5	2-5	3-5	4-5	5-5						
6	1-6	2-6	3-6	4-6	5-6	6-6					
7	1-7	2-7	3-7	4-7	5-7	6-7	7-7				
8	1-8	2-8	3-8	4-8	5-8	6-8	7-8	8-8			
9	1-9	2-9	3-9	4-9	5-9	6-9	7-9	8-9	9-9		
10	1-10	2-10	3-10	4-10	5-10	6-10	7-10	8-10	9-10	10-10	
11	1-11	2-11	3-11	4-11	5-11	6-11	7-11	8-11	9-11	10-11	11-11



$x=4-11, y=4-11$



$x=4-11, y=1-3$



$x=1-3, y=1-3$

1.2. Selección de compuestos de la quimioteca virtual con PRALINS.

La selección de compuestos se ha realizado mediante el uso del programa PRALINS^[6-8] (*Program for Rational Analysis of Libraries in Silico*) desarrollado en el Laboratorio de Diseño Molecular del GEM en el IQS.

Se han seguido las siguientes etapas:

- Optimización de la estructura 3D de los compuestos de la quimioteca mediante el *force field* MMFF94 hasta un gradiente de 0.001 kcal/mol utilizando el programa MOE.^[9]
- Cálculo de descriptores moleculares 2D y 3D a partir de la estructura optimizada (Anexo 1).
- Reducción de la dimensionalidad mediante un análisis de componentes principales (PCA) utilizando el programa MOE.
- Selección de los compuestos diversos utilizando PRALINS a partir de los PCA calculados.

1.2.1. Descriptores moleculares.

Los descriptores moleculares son el resultado de un proceso lógico y matemático que transforma la información química codificada en una representación simbólica de una molécula en un número o en el resultado de un experimento estandarizado.^[10]

Se caracterizan los compuestos de la quimioteca utilizando diferentes descriptores: descriptores 2D (físicoquímicos, topológicos, topológicos basados en teoría de la información) y descriptores 3D (de energía potencial, de superficie, forma y volumen), ver Anexo 1. Los descriptores 2D utilizan solamente los átomos y la información de la conexión de la molécula para el cálculo, por tanto, no dependen de la conformación de la molécula, mientras que los descriptores 3D utilizan información sobre las coordenadas internas de la molécula sin tener en cuenta la rotación y traslación de la conformación (para los descriptores 3D internos), e información sobre las coordenadas absolutas de una molécula (para los descriptores 3D externos).

1.2.2. Análisis de componentes principales (PCA).

El objetivo del análisis de componentes principales es reducir la dimensionalidad de un conjunto de descriptores moleculares por transformación lineal de los datos. Frecuentemente, los descriptores moleculares están correlacionados entre ellos y difieren en su escala relativa. Esto puede suponer un problema al realizar un análisis de diversidad basado en la distancia porque no se puede usar una métrica euclídea directamente y al realizar regresiones ya que se pueden dar efectos de correlación. El análisis por PCA calcula un nuevo conjunto de descriptores no correlacionados y normalizados (media 0 y varianza 1).

El proceso que se ha seguido para realizar esta selección es en primer lugar caracterizar los compuestos de la quimioteca utilizando diferentes descriptores: descriptores 2D (físicoquímicos, topológicos, topológicos basados en teoría de la información) y descriptores 3D (de energía potencial, de superficie, forma y volumen) y a continuación reducir la dimensionalidad mediante un análisis de componentes principales (PCA). Se seleccionan las cinco primeras componentes principales (que explican un 90% de la varianza) como parámetros de entrada para la agrupación en *clusters* de los compuestos de la quimioteca con PRALINS.

1.2.3. Selección de compuestos con PRALINS.

Una vez caracterizada la biblioteca mediante las componentes principales, se procede a la selección de 20 compuestos diversos utilizando el programa PRALINS, mediante la agrupación de los compuestos de la quimioteca en *clusters* jerárquicos (HRC). Se utilizan diferentes métodos de *clustering* implementados en el *software*: HRC average (o método centroide), HRC single (o método del vecino más cercano), HRC complete (o método del vecino más lejano), HRC ward (o método de Ward) y HRC groupaverage (o método de la media de grupo).

Los pasos que sigue el programa PRALINS para el *clustering* son los siguientes:

1. Calcular la matriz triangular de distancias euclídeas entre todos los pares de compuestos; inicialmente cada molécula se considera un *cluster* aislado o *singleton*.
2. Buscar el par de *clusters* con menor distancia y unirlos en un *cluster* común, reemplazando el original.
3. Actualizar la matriz de distancias según el método deseado (fórmula de Lance-Williams, ver Ecuación 1.1 y Tabla 1.2),^[11]

$$d_{k,(i,j)} = \alpha d_{k,i} + \beta d_{k,j} + \delta d_{i,j} + \gamma |d_{k,i} - d_{k,j}| \quad [1.1]$$

donde *i* y *j* son los *clusters* que se juntan y se actualiza la distancia a otro *cluster* *k*.

Tabla 1.2. Constantes de la ecuación de Lance-Williams para los distintos tipos de clustering jerárquico. Las variables i y j representan los clusters que se juntan en un nuevo cluster k y n_i , n_j , n_k representan el número de compuestos en los clusters i , j y k respectivamente.

	α	β	δ	γ
HRC complete	0.5	0.5	0	0.5
HRC groupaverage	0.5	0.5	-0.25	0
HRC single	0.5	0.5	0	-0.5
HRC average	$\frac{n_i}{n_i + n_j}$	$\frac{n_j}{n_i + n_j}$	$\frac{-n_i n_j}{(n_i + n_j)^2}$	0
HRC ward	$\frac{n_k + n_i}{n_k + n_i + n_j}$	$\frac{n_k + n_j}{n_k + n_i + n_j}$	$\frac{-n_k}{n_k + n_i + n_j}$	0

4. Repetir los pasos 2 y 3 hasta obtener el número de *clusters* deseados.

Se representa la agrupación de los compuestos de la quimioteca en *clusters*, en función de la distancia euclídea entre cada uno de los compuestos o grupo de compuestos formados en la iteración anterior, mediante dendogramas (Figura 1.6).

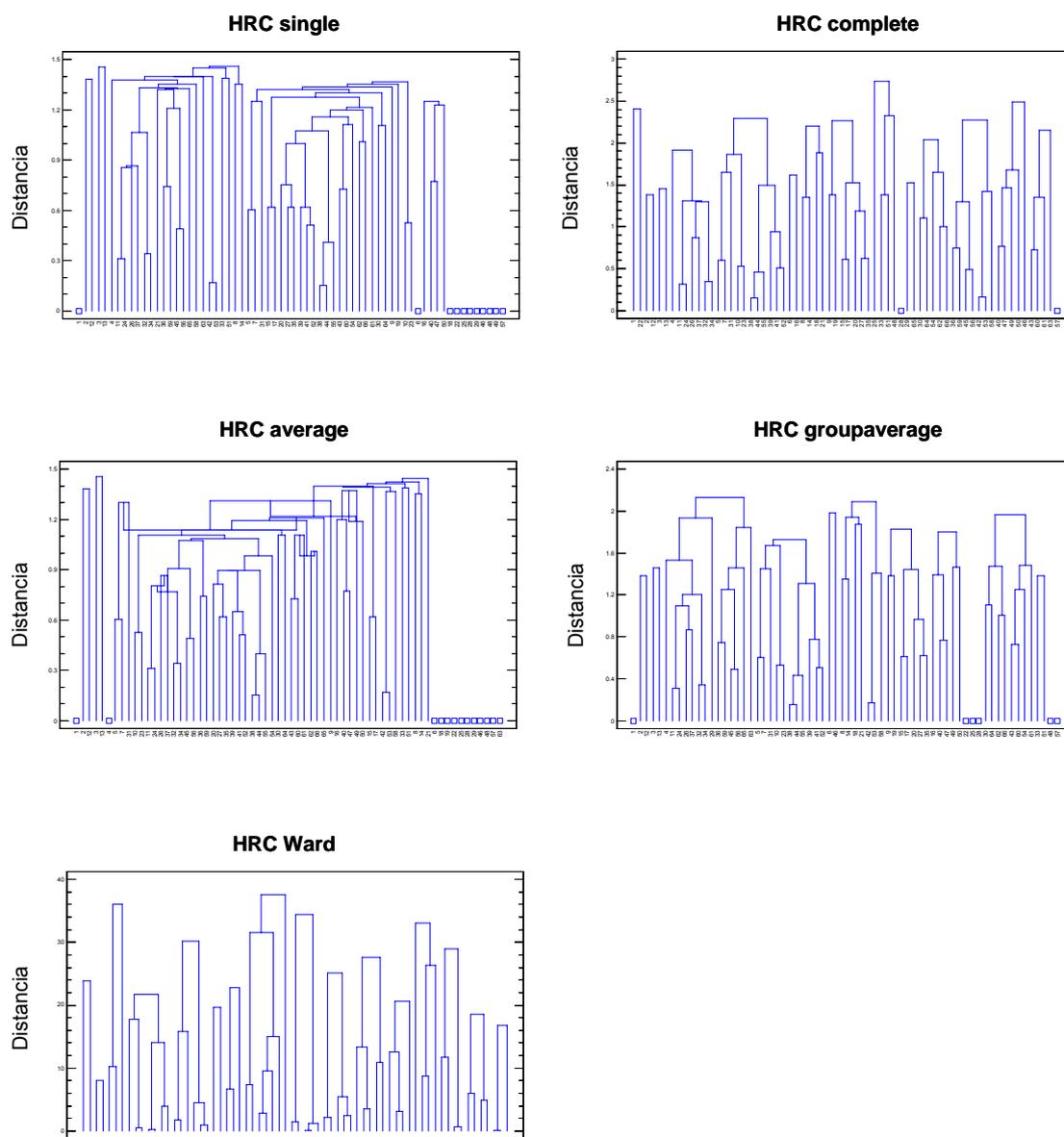


Figura 1.6. Dendogramas de la agrupación en clusters con PRALINS utilizando los diferentes métodos de clustering jerárquico: *single*, *complete*, *average*, *groupaverage* y *Ward* sobre la quimioteca diseñada.

Se compara el número de compuestos obtenidos en cada *cluster* según el método utilizado y se escoge el que realiza una agrupación más equilibrada, es decir, que distribuye los compuestos de la quimioteca diseñada de una manera más equitativa dentro de los diferentes grupos. El resultado obtenido de la distribución de los compuestos de la quimioteca aplicando el método HRC average no es equilibrado, como se puede observar en la Figura 1.7, la mayor parte de los compuestos están agrupados en un único *cluster* mientras que el resto de *clusters* están formados por un solo compuesto. Lo mismo ocurre con el método HRC single, aunque en vez de agrupar la mayor parte de los compuestos en un mismo *cluster*, los distribuye en dos *clusters* principales y el resto están prácticamente formados por un solo compuesto. Los tres métodos restantes, HRC complete, Ward y groupaverage realizan una distribución equilibrada de los compuestos de la quimioteca. Se selecciona el método de Ward por referencias a otros ensayos realizados en nuestro grupo por la Dra. Rabal sobre otro tipo de quimiotecas,^[8] ya que en general ha permitido obtener unos resultados más equilibrados.

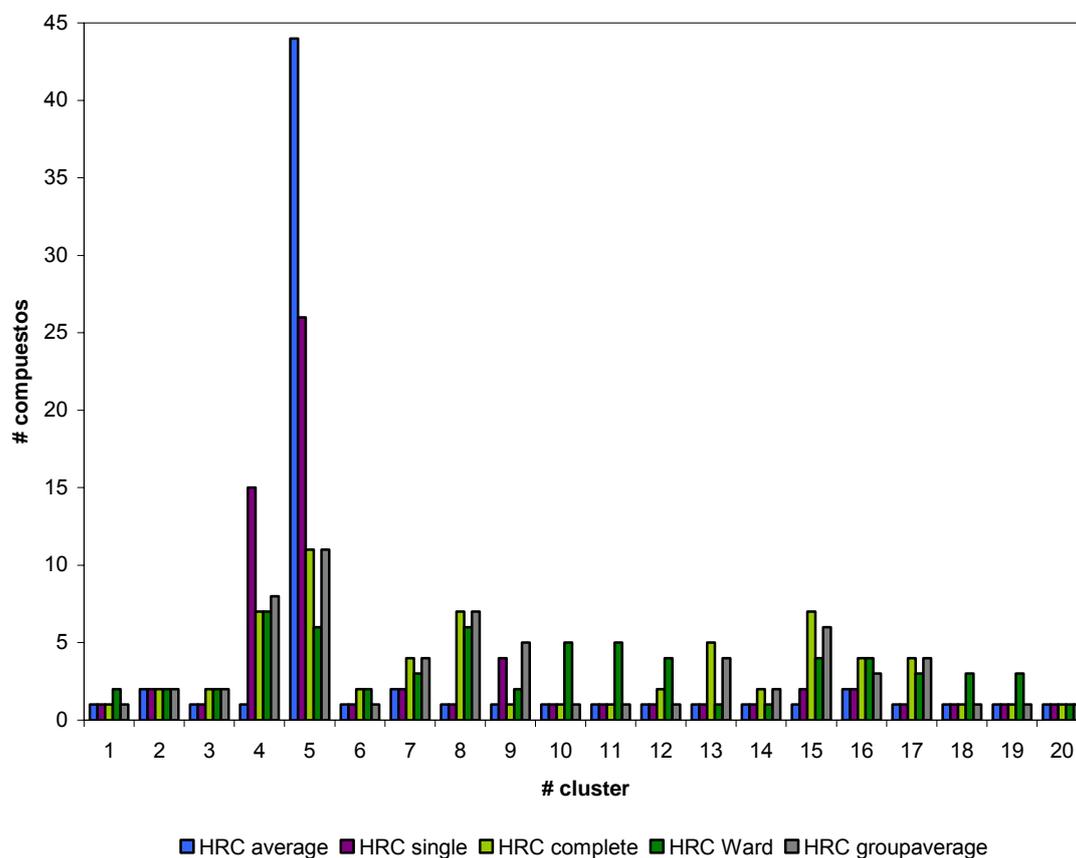


Figura 1.7. Gráfico del número de compuestos por cada cluster según el método de clustering jerárquico utilizado.

Una vez escogido el método de Ward, se selecciona un compuesto de cada *cluster* generado con este método. En la Figura 1.8 se muestran los veinte compuestos diversos seleccionados por PRALINS para explorar el espacio químico representado por la quimioteca diseñada. De esta manera se ha reducido el número de compuestos a sintetizar sin reducir presumiblemente la probabilidad de encontrar *hits* y/o *leads*.

R ₁ R ₂	1	2	3	4	5	6	7	8	9	10	11
1	1-1										
2		2-2									
3	1-3										
4		2-4	3-4								
5			3-5		5-5						
6			3-6			6-6					
7						6-7					
8			3-8		5-8			8-8			
9			3-9						9-9		
10					5-10			8-10			
11					5-11	6-11					11-11

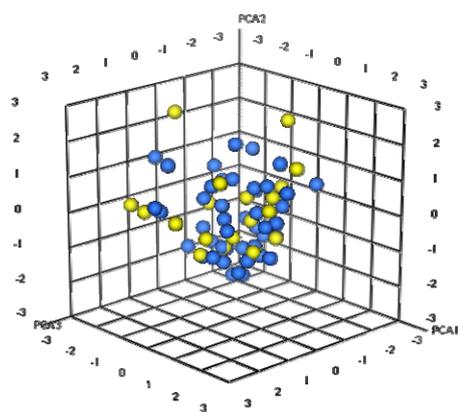


Figura 1.8. Izquierda: compuestos seleccionados de la quimioteca mediante PRALINS aplicando un criterio de diversidad y utilizando el método de Ward. Derecha: representación en función de las tres primeras componentes principales los compuestos seleccionados (amarillo) y del resto de compuestos de la quimioteca (azul).

1.3. Predicción de propiedades ADME-Tox.

1.3.1. Introducción.

Las interacciones de un fármaco con un sistema biológico se pueden agrupar de la siguiente manera.^[12]

- Propiedades farmacocinéticas o efecto del sistema biológico sobre el fármaco:
 - Absorción
 - Distribución
 - Metabolismo
 - Excreción
- Propiedades farmacodinámicas o efecto del fármaco sobre el sistema biológico:
 - Actividad
 - Toxicidad

Las propiedades moleculares de absorción, distribución, metabolismo, excreción (ADME) y toxicidad (Tox) son cruciales en el diseño de fármacos. Muchos candidatos a fármacos son descartados debido a propiedades ADME-Tox insatisfactorias. Se han desarrollado métodos de modelado *in silico* de las propiedades ADME-Tox que permiten seleccionar aquellos compuestos que presentan unas propiedades farmacocinéticas y farmacodinámicas adecuadas y así reducir los costes y el tiempo invertidos en las fases posteriores de desarrollo de un fármaco.

Un parámetro importante en el diseño de nuevos fármacos es la biodisponibilidad, que se define como la cantidad de fármaco disponible en los tejidos para interactuar con su receptor. Está relacionado con la solubilidad del fármaco, su unión a las proteínas plasmáticas, su metabolismo hepático y la eliminación renal.^[13]

Después de la administración un fármaco debe llegar a su diana terapéutica. La forma más habitual de administración es por vía oral, y por lo tanto, el fármaco debe cruzar la membrana intestinal para llegar al torrente sanguíneo. La absorción intestinal ocurre después de que el fármaco se haya disuelto. El siguiente paso es la distribución de un fármaco hasta su diana terapéutica, además de los diferentes tejidos y órganos del organismo, y depende de su unión a las proteínas plasmáticas, en general a las albúminas. Esta unión es un proceso dinámico y generalmente reversible donde la disociación del complejo proteína-fármaco ocurre rápidamente de manera que no es un factor limitante en la absorción del fármaco en su órgano diana.^[14] El metabolismo se define como las transformaciones que sufre un fármaco para facilitar su eliminación del organismo. La mayoría de estas biotransformaciones ocurren en el hígado por acción de unas enzimas de la familia del citocromo P450, cuya inhibición causa la

acumulación de fármaco en el organismo y por lo tanto toxicidad.^[15] Finalmente, los metabolitos generados o el fármaco sin transformar se eliminan del organismo por excreción generalmente a través de la orina o de heces (Figura 1.9).

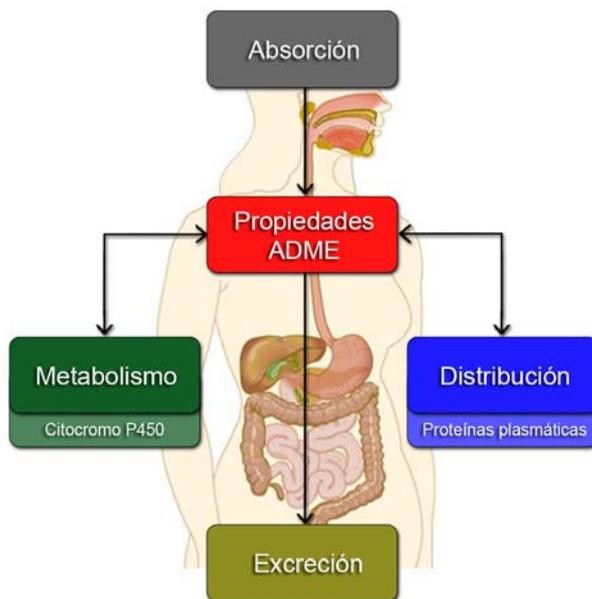


Figura 1.9. Esquema ADME de un fármaco.

1.3.2. Métodos utilizados para la predicción de propiedades ADME-Tox.

La predicción de las propiedades ADME-Tox en el presente trabajo se ha realizado con el programa Accord para Excel.^[16] La absorción intestinal humana después de la administración oral del fármaco se define como el porcentaje de fármaco que pasa al torrente sanguíneo. El modelo que se utiliza para esta predicción se desarrolló a partir de 182 compuestos y utiliza los descriptores *AlogP98* (coeficiente de partición octanol/agua) y *PSA* (área polar de la superficie 2D). El resultado obtenido es una clasificación de los compuestos en cuatro niveles, desde una buena absorción (nivel 0) a una absorción muy baja (nivel 3). Se define que un compuesto con buena absorción debe presentar al menos un 90% de paso al torrente sanguíneo.^[17]

La solubilidad acuosa de un fármaco se puede relacionar directamente con su biodisponibilidad oral, de manera que cuanto mayor es su solubilidad, mayor es su biodisponibilidad oral. El modelo utilizado en la predicción de la solubilidad acuosa (basado en descriptores 2D) fue desarrollado con un *training set* de 775 compuestos cuyo peso molecular está comprendido entre 50 y 800 Da, e incluye gran variedad de grupos funcionales.^[18] Este

modelo permite clasificar los compuestos en cinco niveles según su solubilidad, desde extremadamente poco solubles (nivel 0) a muy solubles (nivel 5).

Otro parámetro en estudio es la penetración de la barrera hematoencefálica. En el caso de fármacos cuya diana terapéutica no es el sistema nervioso central, es importante que no atraviesen la barrera hematoencefálica ya que pueden ocasionar graves efectos secundarios. Esta propiedad se estima mediante un modelo desarrollado a partir de 102 compuestos y permite clasificar las moléculas en cuatro niveles, desde una penetración baja (nivel 3) a una muy alta (nivel 0).

También se estudia el efecto de los compuestos como inhibidores de la enzima CYP2D6 del citocromo P450. El modelo de clasificación binaria utilizado en la predicción fue desarrollado a partir de 100 compuestos diversos y expresa el resultado como inhibidor (clase 1) o no inhibidor (clase 0).^[19,20] El resultado deseado para los compuestos estudiados es que no presenten inhibición de esta enzima.

El modelo de predicción de hepatotoxicidad se desarrolló a partir de 382 compuestos con toxicidad hepática conocida y permite clasificar nuevos compuestos en tóxicos (clase 1) y no tóxicos (clase 0).^[19]

Finalmente, la unión de moléculas a las proteínas plasmáticas se predice según el valor del descriptor *AlogP98*, de manera que clasifica los compuestos en tres niveles: nivel 0 corresponde a una unión inferior al 90%, nivel 1 corresponde a una unión entre el 90 y 95%, y el nivel 2 corresponde a una unión superior o igual al 95%.^[21] Este parámetro influye en la distribución del fármaco en el cuerpo, si el fármaco no se une a las proteínas plasmáticas no llega a su diana terapéutica, interesa tener un valor elevado de unión a proteínas plasmáticas pero no una unión irreversible, siempre debe haber un equilibrio para que el fármaco se pueda liberar y actuar sobre su diana.

1.3.3. Resultados de la predicción de propiedades ADME-Tox.

Se predicen las propiedades ADME de los compuestos de la quimioteca diseñada con potencial actividad anti-VIH. El resultado obtenido determina que ninguno de estos compuestos es inhibidor de la enzima CYP2D6. Sólo uno de ellos, **32**{6,6}, y el compuesto de referencia AMD3100 resultan hepatotóxicos. En general todos presentan buena absorción intestinal y una unión a las proteínas plasmáticas superior al 90% (Tabla 1.3). La predicción determina que solamente **33**{1,2}, **33**{2,2} y **37**{2,8} presentan una penetración alta de la barrera hematoencefálica, propiedad poco deseable en un fármaco cuya diana terapéutica no sea el sistema nervioso central por los efectos secundarios que puede ocasionar sobre el mismo. La mayoría de compuestos de la quimioteca presentan buena solubilidad acuosa, excepto **33**{2,2} que resulta poco soluble en la predicción.

Tabla 1.3. Predicción de la solubilidad y la unión a proteínas plasmáticas de los compuestos con actividad antiviral de la quimioteca diseñada, se incluye la actividad antiviral expresada como EC_{50} en $\mu\text{g/mL}$. Los niveles de penetración de la barrera hematoencefálica oscilan entre 0 (muy alta) y 3 (baja). Los niveles de solubilidad oscilan entre 0 (muy poco soluble) y 5 (muy soluble). Los niveles de unión a proteínas plasmáticas son de 0 (si es < 90%), 1 (si es \geq 90%) o 2 (si es \geq 95%).

Compuesto	Penetración hematoencefálica	Solubilidad acuosa	Unión a proteínas plasmáticas
33{1,1}	1	2	0
33{1,2}	0	2	1
33{1,3}	1	3	0
37{1,4}	1	3	1
37{1,5}	1	3	1
37{1,6}	2	3	2
37{1,7}	1	2	2
37{1,8}	1	2	1
37{1,9}	1	3	1
37{1,10}	2	3	1
37{1,11}	2	3	1
33{2,2}	0	1	2
33{2,3}	1	2	0
37{2,4}	1	2	1
37{2,5}	1	2	1
37{2,6}	1	3	0
37{2,7}	1	2	1
37{2,8}	0	2	1
37{2,9}	1	3	1
37{2,10}	1	3	1
37{2,11}	1	3	1
33{3,3}	2	3	0
37{3,4}	2	3	1
37{3,5}	2	3	1
37{3,6}	2	3	2
37{3,7}	1	3	2
37{3,8}	1	3	1
37{3,9}	2	3	1
37{3,10}	2	3	1
37{3,11}	2	3	1
32{4,4}	2	3	2
32{4,5}	1	3	2
32{4,6}	2	4	2

Compuesto	Penetración hematoencefálica	Solubilidad acuosa	Unión a proteínas plasmáticas
32{4,7}	2	4	2
32{4,8}	2	4	2
32{4,9}	2	3	1
32{4,10}	2	4	2
32{4,11}	2	4	2
32{5,5}	1	3	1
32{5,6}	2	4	2
32{5,7}	1	3	2
32{5,8}	1	3	1
32{5,9}	2	3	1
32{5,10}	2	3	1
32{5,11}	2	4	1
32{6,6}	3	4	2
32{6,7}	2	3	2
32{6,8}	2	3	2
32{6,9}	3	4	2
32{6,10}	3	4	2
32{6,11}	3	4	2
32{7,7}	1	3	2
32{7,8}	1	3	2
32{7,9}	2	3	2
32{7,10}	2	3	3
32{7,11}	2	3	2
32{8,8}	1	3	1
32{8,9}	1	3	1
32{8,10}	2	3	2
32{8,11}	2	3	1
32{9,9}	2	4	1
32{9,10}	3	4	1
32{9,11}	3	4	1
32{10,10}	3	4	1
32{10,11}	3	4	1
32{11,11}	3	4	1
AMD3100	3	3	0

Se compara el resultado de las predicciones de los compuestos de la quimioteca diseñada con las del compuesto de referencia AMD3100. Este compuesto no inhibe la enzima CYP2D6, aunque sí resulta hepatotóxico. Además presenta una unión a las proteínas plasmáticas inferior al 90%, una solubilidad acuosa media y una baja penetración de la barrera hematoencefálica. Se observa que en general, los compuestos de la quimioteca diseñada presentan una mayor unión a las proteínas plasmáticas, de manera que pueden distribuirse de manera más eficaz por el organismo para llegar a su posible diana terapéutica. La solubilidad de los compuestos de la quimioteca, en general es media-alta, esto resulta un factor positivo en cuanto a la biodisponibilidad oral de estos compuestos aunque la predicción de la solubilidad del AMD3100 también es media y se ha demostrado que este compuesto presenta una baja biodisponibilidad oral.^[3] Por último, la predicción de la penetración de la barrera hematoencefálica de los compuestos de la quimioteca diseñada oscila entre baja y muy alta. *A priori* sólo se eliminaría **33{2,2}** de la quimioteca debido a su baja solubilidad, pero como forma parte del conjunto de 20 compuestos seleccionados con PRALINS se decide sintetizarlo y se comprobar su solubilidad experimentalmente. Por lo tanto se consideran todos los compuestos de la quimioteca diseñada como inhibidores potenciales de la unión del VIH al correceptor CXCR4.

1.4. Metodología sintética.

1.4.1. Aminación reductora.

La aminación reductora es la reacción de formación de una amina a partir de un compuesto carbonílico, del tipo aldehído o cetona, y una amina. En la literatura se han descrito reacciones de aminación reductora utilizando amoníaco, aminas primarias y aminas secundarias.

En el caso de llevar a cabo la reacción de aminación reductora con una amina primaria (Figura 1.10), la primera etapa es la formación de un intermedio carbinolamina a partir del compuesto carbonílico y la amina primaria, que seguidamente deshidrata para formar la imina correspondiente (etapa lenta). La segunda etapa (la más rápida) es la reducción de la imina formada en la etapa anterior para rendir el producto final.^[22]

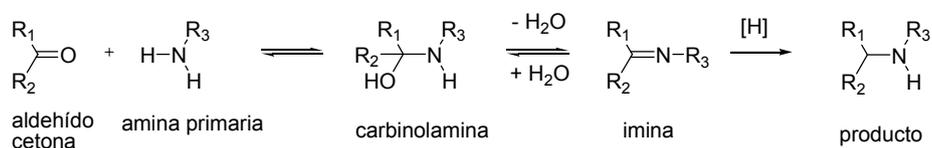


Figura 1.10. Mecanismo de aminación reductora entre un aldehído o cetona y una amina primaria.

En el caso de aminas secundarias hay dos mecanismos posibles en función de la presencia o ausencia de átomos de hidrógeno en posición α del carbonilo (Figura 1.11). Si hay un átomo de hidrógeno en dicha posición, se forma una carbinolamina intermedia a partir de la amina secundaria y el compuesto carbonílico, que seguidamente deshidrata dando lugar a una enamina intermedia. Esta enamina intermedia se reduce para obtener la amina deseada. Mientras que si no hay un átomo de hidrógeno en posición α de carbonilo, no se puede formar la enamina intermedia, en este caso el mecanismo procede vía hidrogenólisis de la carbinolamina formada a partir del compuesto carbonílico y la amina secundaria.^[22]

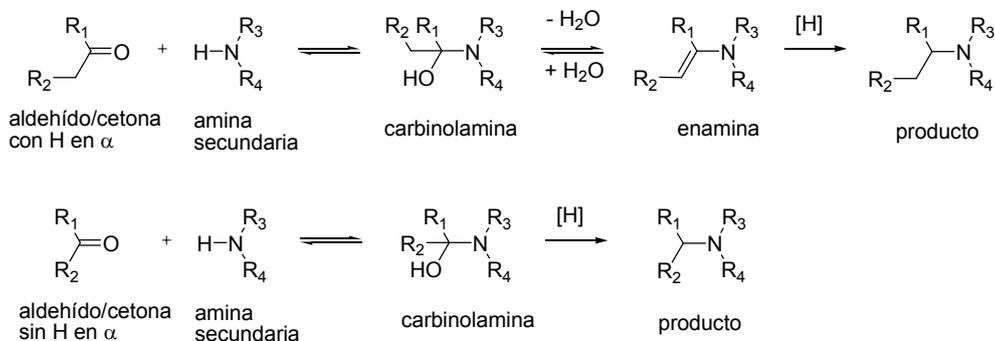


Figura 1.11. Mecanismo de reacción de aminación reductora. Arriba: de un aldehído o cetona con H en α y una amina secundaria. Abajo: de un aldehído o cetona sin H en α y una amina secundaria.

La etapa de formación de la imina a partir de una amina primaria y un aldehído o cetona, o de formación de la enamina a partir de un aldehído o cetona con hidrógeno en α y una amina secundaria, se puede acelerar mediante la eliminación de agua del medio de reacción ya sea con destilación azeotrópica, tamiz molecular, MgSO_4 o Na_2SO_4 anhidros.

Las iminas con un grupo arilo se conocen con el nombre de bases de Schiff, son compuestos estables e incluso se pueden aislar, en comparación con las iminas sin sustituyente arilo, que descomponen o polimerizan fácilmente.

La aminación reductora de aldehídos y cetonas se ha llevado a cabo con una extensa selección de reactivos. La selección de un reductor apropiado es crucial si se quiere llevar a cabo una aminación reductora directa, ya que debe reducir selectivamente a las iminas formadas frente a los aldehídos o cetonas utilizados como reactivo de partida para obtener el producto deseado. En cambio, si se lleva a cabo una aminación reductora indirecta o por etapas, es decir, en primer lugar se forma la imina a partir del aldehído o cetona y la amina correspondientes y a continuación se reduce dicha imina, la selección del agente reductor no resulta tan crítico.

En este ámbito, los métodos más comunes son la hidrogenación catalítica,^[23,24] Zn/AcOH ,^[25] $\text{Et}_3\text{SiH-TFA}$,^[26] Bu_3SnH ^[27] y varios agentes reductores de la familia de los borohidruros, como NaBH_4 ,^[28-34] NaBH_3CN ,^[35,36] $\text{NaBH}(\text{OAc})_3$,^[28,37-39] resina de intercambio de borohidruro^[40] y borano-piridina.^[41-43] En algunos casos se emplea *in situ* un ácido de Lewis como ZnCl_2 , TiCl_4 y $\text{Ti}(\text{O}i\text{Pr})_4$ para promover la condensación entre la amina con el compuesto carbonílico.^[44]

En la aminación reductora con hidruros, la hidrogenación con catálisis de platino, paladio o níquel es un método económico y efectivo, pero su principal inconveniente es que no es selectivo y presenta problemas de reacciones secundarias en presencia de otros grupos funcionales susceptibles de reducirse. Por lo tanto, se prefieren agentes reductores más selectivos.

El cianoborohidruro sódico es un reductor estable en soluciones ácidas (aproximadamente hasta pH 3), es soluble en disolventes hidroxílicos, como MeOH , y presenta selectividad en función del pH. A pH entre 3 y 4 reduce aldehídos y cetonas de manera efectiva, pero dicha reacción se enlentece al aumentar el pH. A pH entre 6 y 8 las iminas se protonan preferentemente a los aldehídos/cetonas y se reducen más rápido que estos últimos. Por tanto, este reductor permite llevar a cabo una reducción selectiva de iminas a pH básico.^[35,45] Los principales inconvenientes en el uso de cianoborohidruro sódico como agente reductor es el gran exceso de amina de partida que se debe utilizar en la reacción, contaminación del producto deseado con subproductos, su elevada toxicidad y la toxicidad de los subproductos generados en el *work up* de la reacción como HCN y NaCN . Dados estos inconvenientes se prefiere el uso de otro agente reductor.

Otro reductor selectivo de iminas es el triacetoxiborohidruro sódico, que resuelve los problemas que ocasiona el reductor anterior e incluso se puede generar *in situ* a partir de NaBH_4 y el correspondiente ácido carboxílico.^[38,46] Sin embargo, en la reacción de aminación reductora de aldehídos con aminas primarias es muy frecuente la dialquilación como reacción secundaria. Además, actúa como reductor selectivo de iminas frente a aldehídos cuando la reacción se realiza en disolventes polares apróticos, pero no en metanol, donde la reducción del carbonilo es una reacción en competencia con la reducción de la imina formada, o en agua, que reacciona con el triacetoxiborohidruro sódico. Dada la dificultad de separación del producto de dialquilación del de monoalquilación y la limitación en la elección del disolvente, se descarta el uso de este agente reductor.^[37]

Por tanto, la alternativa preferida es la aminación reductora indirecta, es decir, en primer lugar se lleva a cabo una primera etapa de condensación entre el aldehído y la amina para generar la imina correspondiente, seguida de una segunda etapa de reducción de la imina formada con NaBH_4 para obtener la amina objetivo. Abdel-Magid y colaboradores estudiaron la velocidad de formación de iminas a partir de aldehídos y aminas primarias en diferentes disolventes, THF, DCE y MeOH. Concluyeron que las reacciones realizadas en medio metanólico eran más rápidas y se obtenían conversiones prácticamente cuantitativas en comparación con los otros disolventes ensayados.^[28]

En conclusión, se realizarán aminaciones reductoras indirectas de aldehídos con aminas primarias y secundarias. Constan de una primera etapa de formación de las iminas donde se utilizará MeOH como disolvente, y tamiz molecular 4 Å o Na_2SO_4 como agente deshidratante, y una segunda etapa de reducción de las iminas formadas utilizando NaBH_4 como agente reductor.

1.4.2. Condensación de aldehídos o cetonas con hidrazinas.

La formación de hidrazonas ocurre por condensación entre un compuesto carbonílico del tipo aldehído o cetona y una hidrazina (Figura 1.12). La reacción transcurre de manera análoga a la condensación entre la amina y el compuesto carbonílico de la aminación reductora. Sin embargo, dado que las hidrazonas son más estables que las iminas por contribución del nitrógeno adicional en la deslocalización electrónica (Figura 1.13),^[47,48] éstas se pueden aislar y guardar sin problemas, siendo innecesaria la inmediatez de la etapa posterior de reducción.

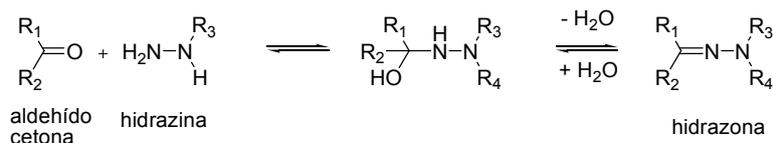


Figura 1.12. Mecanismo de formación de hidrazonas a partir de un aldehído o cetona y una hidrazina.

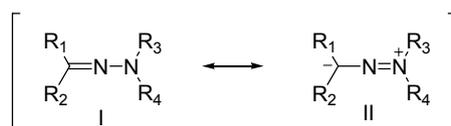


Figura 1.13. Deslocalización electrónica de las hidrazonas.

1.4.3. Estrategia de obtención de los compuestos de la quimioteca diseñada.

1.4.3.1. Estrategia de obtención de diaminas simétricas **32**{x,x}.

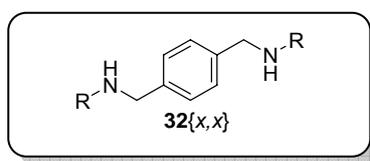


Figura 1.14. Estructura general de las diaminas simétricas **32**{x,x}.

La obtención de las diaminas simétricas **32**{x,x} se lleva a cabo por reacción de aminación reductora por etapas de 1 mol de tereftalaldehído (**29**) y 2 moles de amina **31**{4-11} en MeOH anhidro y en presencia de tamiz molecular como agente deshidratante. En la primera etapa se forma la diimina intermedia **38**, y a continuación se lleva a cabo la segunda etapa de

reducción de dicha diimina **38** con 2 moles de NaBH_4 para rendir la diamina objetivo **32** $\{x,x\}$ (Figura 1.15).

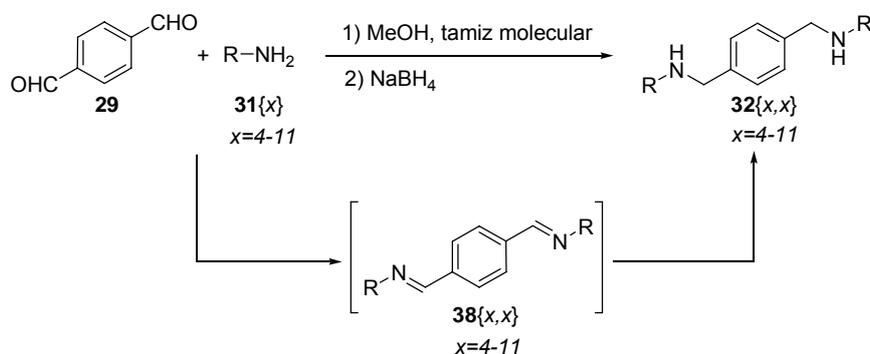


Figura 1.15. Obtención de diaminas simétricas **32** $\{x,x\}$ por reacción de aminación reductora por etapas de tereftaldehído (**29**) y la amina correspondiente **31** $\{x\}$.

La imina intermedia **38** se puede aislar dado que la conjugación del enlace $\text{C}=\text{N}$ con el anillo aromático estabiliza este tipo de compuestos. Sin embargo, puesto que el objetivo es la diamina **32** $\{x,x\}$ y no el intermedio diimínico **38**, se reduce directamente este intermedio.

El seguimiento de la reacción no se puede realizar por CCF porque las iminas **38** formadas se hidrolizan sobre la fase estacionaria (sílica o alúmina) y como productos de hidrólisis se obtienen los propios reactivos, es decir, tereftaldehído (**29**) y amina **31**. Es necesario realizar el seguimiento de la reacción mediante ^1H -RMN donde se puede identificar la formación de la imina **38** por la aparición de un singlete correspondiente al protón del grupo $-\text{CH}=\text{N}-$ hacia 8 ppm y la desaparición del singlete del protón de los grupos $-\text{CHO}$ hacia 10 ppm.

1.4.3.2. Estrategia de obtención de dihidrazonas simétricas **33** $\{x,x\}$.

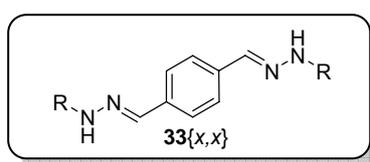


Figura 1.16. Estructura general de las dihidrazonas simétricas **33** $\{x,x\}$.

La obtención de las hidrazonas simétricas **33**{x,x} se lleva a cabo por condensación de 1 mol de tereftalaldehído (**29**) con 2 moles de hidrazina **30**{1-3} en MeOH anhidro y en presencia de tamiz molecular (Figura 1.17).

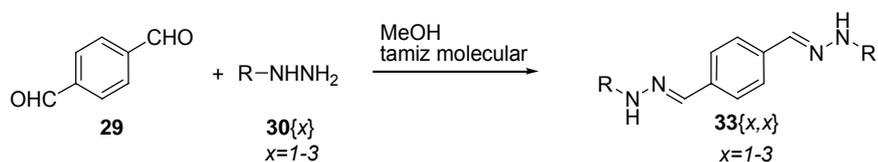


Figura 1.17. Obtención de hidrazonas simétricas **33**{x,x} por condensación de tereftalaldehído (**29**) y la hidrazina **30**{x} correspondiente.

El seguimiento de la reacción por $^1\text{H-RMN}$ permite identificar la formación de la hidrazona **33**{x,x} por la aparición de la señal correspondiente al protón del $-\text{CH}=\text{N}-$ como un singlete hacia 8 ppm y la desaparición del singlete del protón de los grupos $-\text{CHO}$ hacia 10 ppm.

1.4.3.3. Estrategia de obtención de diaminas no simétricas **32**{x,y}.

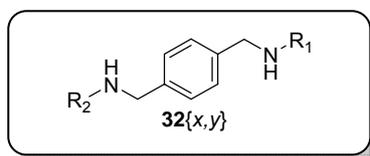


Figura 1.18. Estructura general de las diaminas no simétricas **32**{x,y}.

La síntesis de compuestos de sustitución no simétrica se puede abordar mediante dos estrategias diferentes de monofuncionalización:

- condiciones de reacción que favorezcan la monosustitución
- uso de un grupo protector

La obtención de la monoimina **39** sería posible utilizando unas condiciones de reacción que favorezcan la monosustitución frente a la disustitución (en que se obtendría la diimina simétrica **38**), como pueden ser el uso de un exceso de tereftalaldehído (**29**) respecto a la amina **31**, realizar la reacción en alta dilución, realizar una adición lenta de la amina **31** correspondiente, etc. El inconveniente de esta vía es que, aún favoreciendo la monosustitución, se obtendrán también el producto disustituido **38** y quedará un exceso de tereftalaldehído (**29**).

La manera de aislar el producto monosustituido **39** sería por cromatografía de columna, pero no se puede realizar con las iminas porque se hidrolizan sobre la fase estacionaria (gel de sílice). Otra opción sería reducir la imina **39** formada para obtener la amina **40** correspondiente, pero en este caso también se reduciría el grupo aldehído a alcohol y no estaría presente el grupo aldehído necesario para realizar la segunda aminación reductora y poder obtener la diamina de sustitución no simétrica **32** $\{x,y\}$ (Figura 1.19).

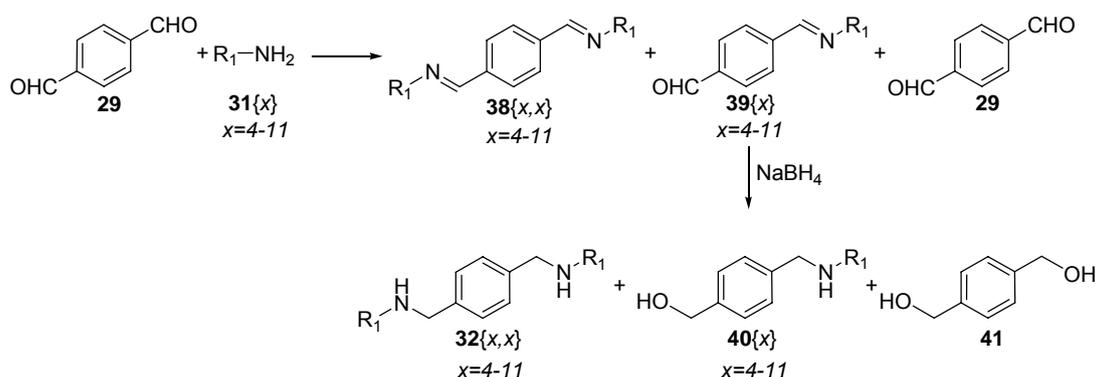


Figura 1.19. Reacción de obtención de la monoimina **39** a partir de tereftalaldehído (**29**) y la amina **31** correspondiente, se obtendrían también la diimina simétrica **38** y quedaría exceso de tereftalaldehído (**29**). Una etapa posterior de reducción con NaBH₄ reduciría las iminas a grupos amino pero también los grupos aldehído a alcoholes.

El uso de un grupo protector (GP) sobre uno de los grupos aldehído del tereftalaldehído permitiría realizar la aminación reductora sobre el aldehído libre **42** con una primera amina **31** y, tras una etapa de desprotección que permitiese recuperar el otro grupo aldehído **44**, realizar una segunda aminación reductora con otra amina **31** diferente de la primera, rindiendo así el compuesto objetivo del tipo diamina no simétrica **32** $\{x,y\}$ (Figura 1.20).

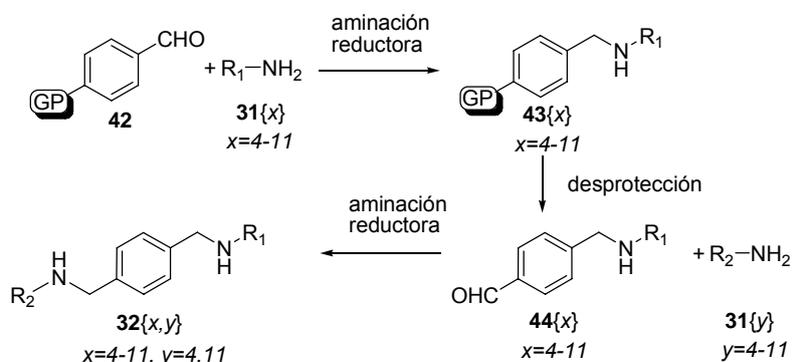


Figura 1.20. Estrategia de síntesis de compuestos del tipo diaminas no simétricas **32** $\{x,y\}$ con uso de grupo protector (GP).

Esta segunda estrategia posibilita la obtención de las diaminas **32**{x,y} objetivo mediante una síntesis de tres pasos.

Una vez seleccionada esta segunda estrategia sintética, se debe escoger un grupo protector de aldehídos que resista las condiciones de la reacción de aminación reductora y el aislamiento: disolución metanólica básica, disolución acuosa básica y el reductor NaBH₄. Los acetales cumplen todos estos requisitos, y se desprotegen en condiciones de reacción suaves que no interaccionan con la amina formada y con rendimientos elevados. Generalmente los acetales se desprotegen mediante hidrólisis ácida utilizando disoluciones acuosas de ácidos próticos,^[49-51] concretamente se utilizará HCl acuoso diluido.^[52] Se realiza una búsqueda de acetales en bases de datos de reactivos comerciales y se selecciona el 4-(dietoximetil)benzaldehído (**34**) como reactivo para generar el core.

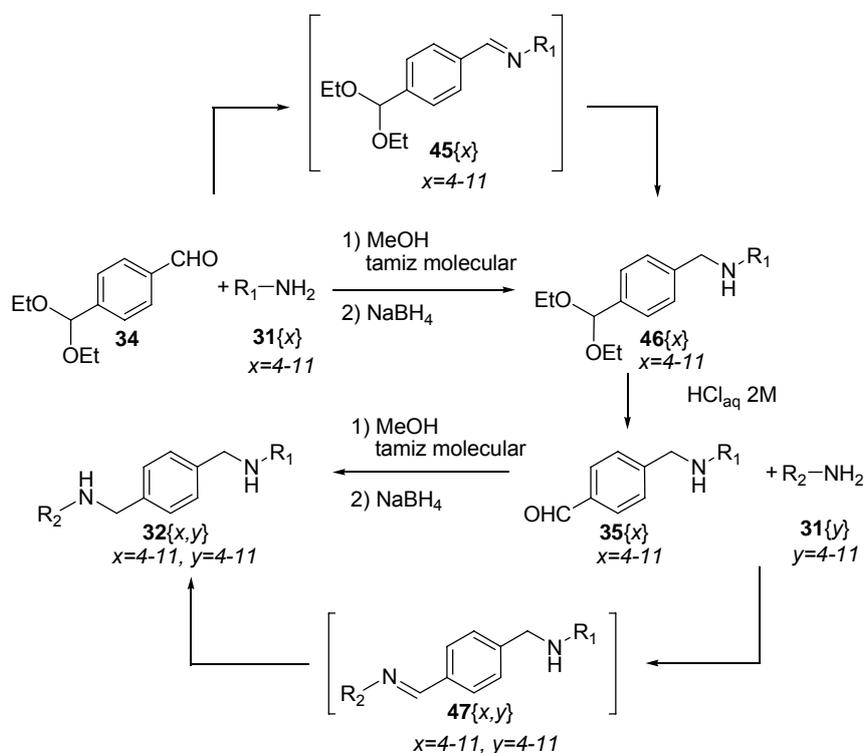


Figura 1.21. Vía sintética para la obtención de las diaminas no simétricas **32**{x,y}, siendo $x \neq y$.

Experimentalmente se realiza la aminación reductora a partir de cantidades equimolares de 4-(dietoximetil)benzaldehído (**34**) y de una primera amina **31**{x}, seguida de la reducción de la imina **45** formada con una cantidad equimolar de NaBH₄ para obtener un intermedio aminoacetálico **46** que se desprotege en HCl_{aq} rindiendo un compuesto del tipo aminoaldehído **35**. Sobre dicho compuesto se lleva a cabo una segunda aminación reductora con una cantidad equimolar de otra amina **31**{y} diferente de la primera, de manera que se obtienen las diaminas no simétricas **32**{x,y} (Figura 1.21).

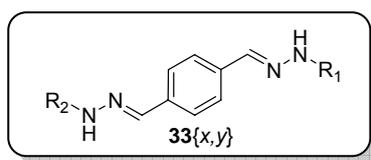
1.4.3.4. Estrategia de obtención de dihidrazonas no simétricas **33**{x,y}.

Figura 1.22. Estructura general de las dihidrazonas no simétricas **33**{x,y}.

En la síntesis de dihidrazonas no simétricas **33**{x,y} no se puede utilizar la estrategia anterior, en que se protege uno de los grupos aldehído con un acetal dietílico, porque en las condiciones de reacción de desprotección de dicho acetal, medio acuoso ácido, también se produce la reacción de hidrólisis de la hidrazona al aldehído y la hidrazina correspondientes. Las hidrazonas son más estables que las iminas y no se hidrolizan sobre sílica o alúmina, y por tanto se pueden purificar por cromatografía de columna. Por tanto se puede utilizar la primera estrategia sintética planteada para las diaminas no simétricas **32**{x,y}, es decir, unas condiciones de reacción que favorezcan la formación de la monohidrazona **36** frente a la dihidrazona **33**{x,x}. Concretamente, se trabaja en presencia de un exceso dos veces superior de tereftalaldehído (**29**) frente a hidrazina **30** y se realiza una adición lenta de dicha hidrazina **30** sobre una disolución en MeOH del aldehído **29**. Una vez acabada la reacción se obtiene el producto monosustituído deseado (hidrazonoaldehído **36**), exceso de tereftalaldehído (**29**) y una pequeña cantidad de producto de disustitución **33**{x,x} cuya formación no se ha podido evitar. Dicha mezcla se puede separar por cromatografía de columna (gel de sílice) y se aísla el producto de monosustitución **36**. A continuación, se realiza una condensación con cantidades equimolares del intermedio hidrazonoaldehído **36** obtenido con una hidrazina **30** diferente de la primera para obtener la dihidrazona no simétrica **33**{x,y} (Figura 1.23).

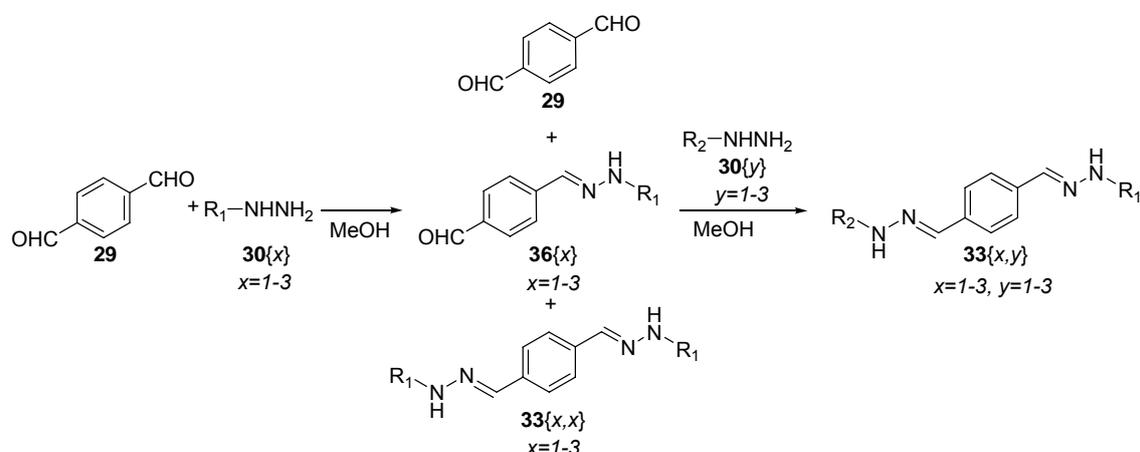


Figura 1.23. Vía sintética utilizada en la obtención de dihidrazonas no simétricas **33**{x,y}.

1.4.3.5. Estrategia de obtención de aminohidrazonas $37\{x,y\}$.

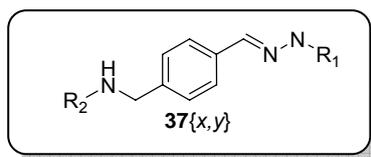


Figura 1.24. Estructura general de las aminohidrazonas no simétricas $37\{x,y\}$.

Las aminohidrazonas $37\{x,y\}$ se pueden obtener a partir de los intermedios aminoaldehído **35** o hidrazonoaldehído **36** sintetizados para la obtención de diaminas no simétricas **32** $\{x,y\}$ (Figura 1.18) y dihidrazonas no simétricas **33** $\{x,y\}$ (Figura 1.22) respectivamente. Si se utiliza el aminoaldehído **35** como reactivo, se lleva a cabo una reacción de condensación con una hidrazina **30** $\{1-3\}$ (en cantidades equimolares) para rendir la aminohidrazona **37** $\{x,y\}$ (Figura 1.25). En cambio, si se utiliza el hidrazonoaldehído **36** se realiza una reacción de aminación reductora de dicho compuesto con la amina **31** $\{4-11\}$ correspondiente (en cantidades equimolares) y posterior reducción de la imina **48** formada con NaBH_4 para rendir la aminohidrazona **37** $\{x,y\}$ (Figura 1.25).

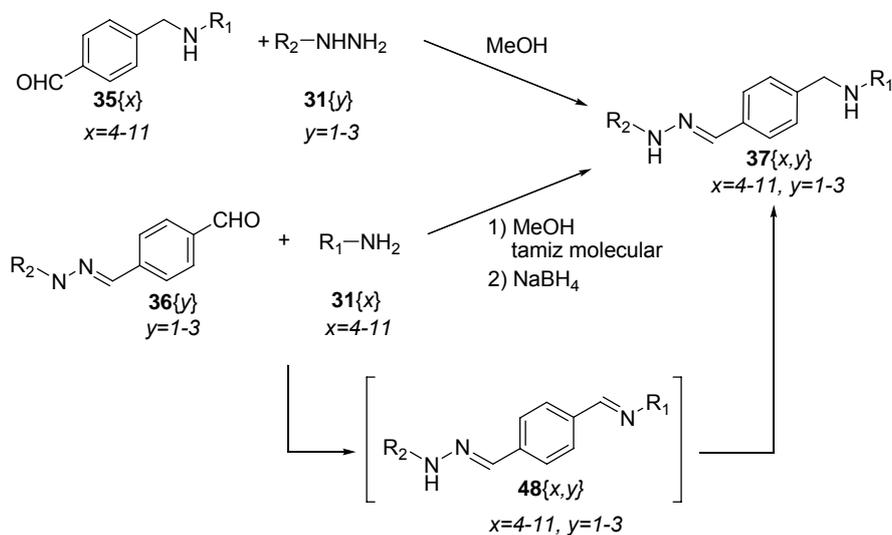


Figura 1.25. Vía sintética utilizada en la obtención de aminohidrazonas $37\{x,y\}$.

1.5. Evaluación de la actividad anti-VIH y del modo de acción de un fármaco.

La evaluación de la actividad anti-VIH y del modo de acción de los compuestos sintetizados se realiza en el grupo del Dr. Esté en el Laboratorio de Retrovirología de la Fundación IrsiCaixa del Hospital Universitari Germans Trias i Pujol.

La determinación de la actividad anti-VIH se realiza mediante un ensayo de colorimetría que utiliza el MTT, bromuro de 3-(4,5-dimetiltiazol-2-il)-2,5-difeniltetrazolo, de color amarillo a sal de formazán, de color liloso. Las células vivas y metabólicamente activas son capaces de reducir el MTT a formazán. Se utilizan células MT-4 (línea celular de células T humanas aisladas de un paciente con leucemia de células T) y se infectan con la cepa salvaje de VIH-1_{NL4-3}, que presenta tropismo X4.

Se determina la actividad anti-VIH-1 expresada como EC₅₀ (concentración efectiva 50 o concentración necesaria para inhibir el 50% de muerte celular inducida por el VIH-1) y la toxicidad celular expresada como CC₅₀ (concentración citotóxica 50 o concentración necesaria para inducir el 50% de muerte celular inducida por la toxicidad del compuesto) en base a la viabilidad de las células MT-4 infectadas o no infectadas con el virus respectivamente, a diferentes concentraciones de los compuestos a analizar disueltos en DMSO. Las células MT-4 se dejan proliferar durante 5 días y se cuantifica el número de células viables mediante el método colorimétrico MTT.^[53,54]

Como compuestos de referencia se utiliza el biciclamo AMD3100 que es un inhibidor del correceptor CXCR4 y cuyo valor de EC₅₀ es 0.001 µg/mL, y el polisacárido sulfato de dextrano (DS) que es un inhibidor de la unión del virus a la célula por interacción con la proteína viral gp120 y cuyo valor de EC₅₀ es 0.011 µg/mL.

El modo de acción de un compuesto se determina mediante ensayos de tiempo de adición del compuesto y por inhibición de la unión de anticuerpos monoclonales específicos de un correceptor en presencia del compuesto.

El ensayo del tiempo de adición consiste en sincronizar una infección de células y posteriormente se adicionan los compuestos a ensayar a diferentes tiempos de postinfección. El objetivo del ensayo es determinar en qué momento deja de inhibir la infección cada compuesto y se evalúa mediante la cuantificación de la producción de antígeno viral CAp24 (proteína p24 del core de la cápside viral) en cada punto a 30 horas postinfección (que es el tiempo en que se ha dado un ciclo completo de replicación). En la Figura 1.26 se observa el perfil de las curvas de un inhibidor de fusión del virus y la célula (C34), un inhibidor de la unión de la proteína viral gp120 al receptor celular CD4 (DS), un inhibidor de la unión al correceptor CXCR4 (AMD3100) y un inhibidor de la RT (AZT).

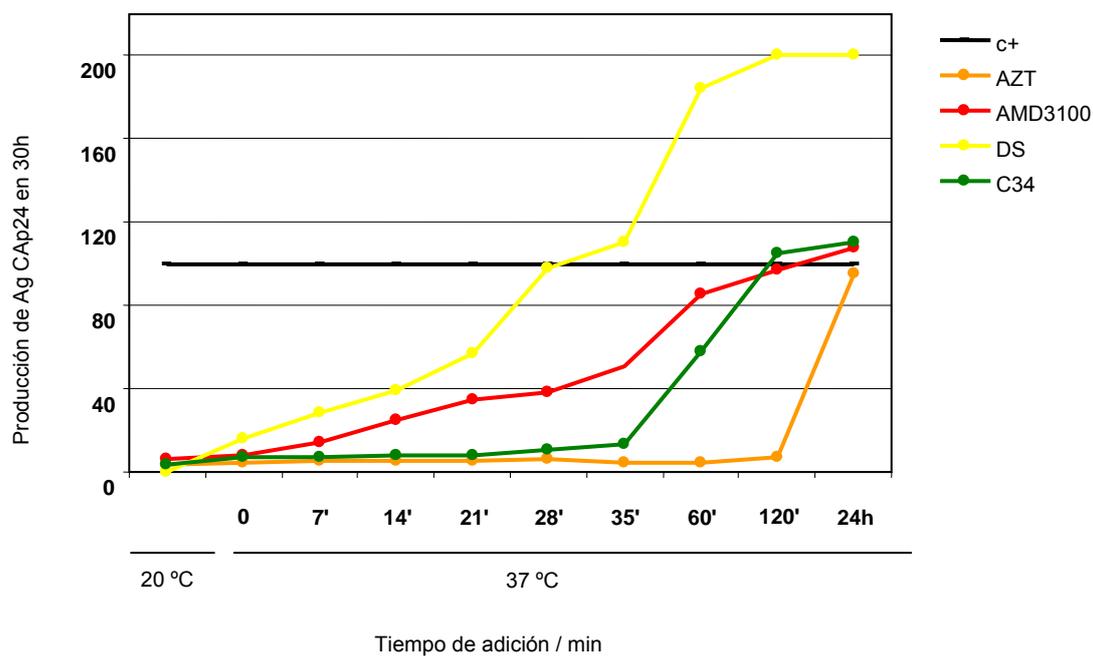


Figura 1.26. Gráfico del tiempo de adición de compuestos con mecanismo de acción conocido. Se cuantifica mediante la producción del antígeno viral CAp24 a las 30 horas de postinfección de las células MT-4 con la cepa salvaje VIH-1_{NL4-3}.

Otro ensayo para determinar el mecanismo de acción de los compuestos sintetizados es mediante un ensayo de marcaje de la unión de anticuerpos monoclonales específicos de tres receptores implicados en la infección por VIH-1 (CD4, CXCR4 y CCR5) y un receptor control (CD45) en células CEM-CCR5 (línea celular T CD4⁺ humana que expresa en su superficie los receptores CD4, CXCR4, CCR5 y CD45) y la inhibición de esta unión por parte de los compuestos ensayados.

1.6. Resultados y discusión.

1.6.1. Síntesis y actividad anti-VIH.

Inicialmente se seleccionan 20 compuestos diversos de la quimioteca diseñada (ver Figura 1.8) con PRALINS para explorar el espacio químico representado por la quimioteca sin tener que sintetizar todas las moléculas que la componen. De los 20 compuestos seleccionados se sintetizan inicialmente 8 de ellos: **33**{1,1}, **33**{2,2}, **37**{3,4}, **32**{5,5}, **32**{6,6}, **32**{6,11}, **32**{9,9} y **32**{11,11}. Una vez sintetizados se evalúan sus actividades anti-VIH. El resultado muestra que tres de ellos presentan elevada actividad antiviral, **32**{5,5}, **32**{6,11} y **32**{9,9}, presentan valores de EC₅₀ entre 0.8 y 1.9 µg/mL (Figura 1.27). Ya que el compuesto de referencia AMD3100 presenta un valor de 0.001 µg/mL, se intenta optimizar el resultado obtenido mediante la síntesis y el análisis de otros compuestos de la quimioteca diseñada.

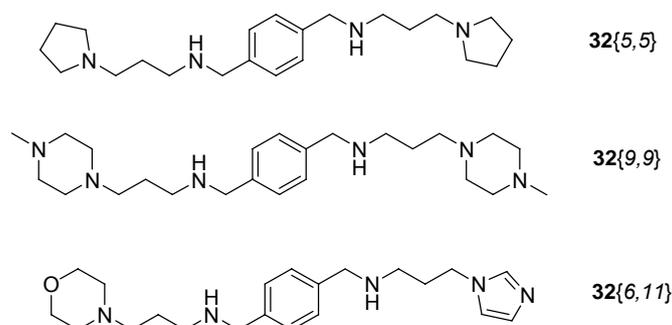


Figura 1.27. Estructura de los compuestos **32**{5,5}, **32**{6,11} y **32**{9,9}.

En la siguiente etapa se consideran las 12 moléculas restantes que forman parte del grupo inicialmente seleccionado con PRALINS y aquellas que forman parte de las familias de los tres compuestos activos identificados en la primera etapa, es decir aquellos 26 compuestos de la quimioteca que contienen los *building blocks* **31**{5}, **31**{6}, **31**{9} y **31**{11} y que no se habían incluido en la primera selección de diversidad.

De estas 38 moléculas, en primer lugar se sintetiza un subconjunto de 17 compuestos: **33**{1,3}, **37**{1,5}, **37**{2,5}, **37**{3,5}, **37**{3,6}, **37**{3,8}, **37**{3,9}, **32**{4,5}, **32**{5,6}, **32**{5,7}, **32**{5,8}, **32**{5,9}, **32**{5,10}, **32**{5,11}, **32**{6,7}, **32**{9,10} y **32**{9,11}. Doce compuestos de este subconjunto presentaron valores de EC₅₀ entre 0.2 y 2.7 µg/mL, incluso dos de ellos destacan por su potente actividad anti-VIH, **32**{5,6} y **32**{5,8}, ambos con un valor de EC₅₀ de 0.2 µg/mL.

A continuación se sintetiza un segundo subconjunto de 28 moléculas que incluye el los 3 compuestos restantes de la selección inicial de 20 compuestos diversos, **37**{2,4}, **32**{8,8} y **32**{8,10}, los 18 compuestos no sintetizados de las familias de los tres compuestos activos

identificados en la primera etapa, **37{1,6}**, **37{1,9}**, **37{1,11}**, **37{2,6}**, **37{2,9}**, **37{2,11}**, **37{3,11}**, **32{4,6}**, **32{4,9}**, **32{4,11}**, **32{6,8}**, **32{6,9}**, **32{6,10}**, **32{7,9}**, **32{7,11}**, **32{8,9}**, **32{8,11}** y **32{10,11}**, los 4 compuestos simétricamente sustituidos no sintetizados previamente, **33{3,3}**, **32{4,4}**, **32{7,7}** y **32{10,10}**, y 3 compuestos seleccionados mediante métodos computacionales de *docking* sobre el correceptor CXCR4, modelos QSAR y farmacofóricos calculados a partir de los activos previamente identificados (Capítulos 3 y 4), **37{1,8}**, **37{2,8}** y **32{7,8}**.^[55] Tres compuestos de este grupo presentan valores de EC₅₀ de 0.03 µg/mL, **32{6,8}** y **32{8,9}** y de 0.008 µg/mL, **32{8,8}** (Figura 1.28), este último es del mismo orden que el compuesto de referencia AMD3100 (0.001 µg/mL) y no muestran citotoxicidad a las concentraciones ensayadas de máximo 25 µg/mL.

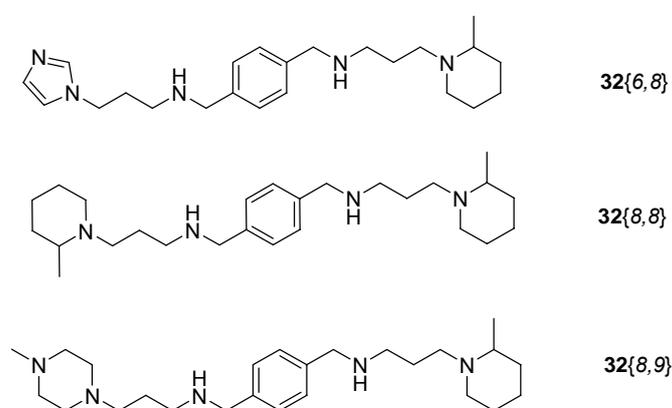


Figura 1.28. Estructura de los compuestos **32{6,8}**, **32{8,8}** y **32{8,9}**.

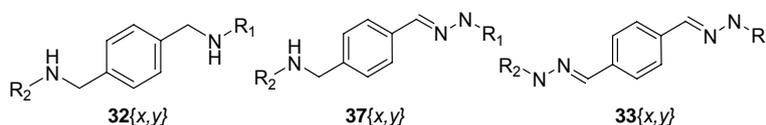
El resultado de los ensayos de actividad anti-VIH y de citotoxicidad de los 53 compuestos sintetizados **32**, **37** y **33** se muestra en la Tabla 1.4.

Tabla 1.4. Valores de actividad anti-VIH (EC₅₀) y de citotoxicidad (CC₅₀) expresados en µg/mL de los compuestos sintetizados y de los compuestos de referencia AMD3100 (inhibidor de la unión del virus a CXCR4) y DS (inhibidor de la unión del virus a la célula). Se resaltan en negrita los tres compuestos más activos **32{6,8}**, **32{8,8}** y **32{8,9}**.

Compuesto	R ₁ NH ₂	R ₂ NH ₂	EC ₅₀ / µg·mL ⁻¹	CC ₅₀ / µg·mL ⁻¹
33{1,1}	30{1}	30{1}	>125	>125
33{1,3}	30{1}	30{3}	>25	>25
37{1,5}	30{1}	31{5}	2.7	10.1
37{1,6}	30{1}	31{6}	>25	>25
37{1,8}	30{1}	31{8}	>4.1	4.1

Compuesto	R ₁ NH ₂	R ₂ NH ₂	EC ₅₀ / µg·mL ⁻¹	CC ₅₀ / µg·mL ⁻¹
37{1,9}	30{1}	31{9}	>9.8	9.8
37{1,11}	30{1}	31{11}	10.6	>25
33{2,2}	30{2}	30{2}	>125	>125
37{2,4}	30{2}	31{4}	14.7	>25
37{2,5}	30{2}	31{5}	2.0	9.8
37{2,6}	30{2}	31{6}	>25	>25
37{2,8}	30{2}	31{8}	0.6	14.6
375{2,9}	30{2}	31{9}	3.8	19.2
37{2,11}	30{2}	31{11}	15.7	>25
33{3,3}	30{3}	30{3}	>125	>125
37{3,4}	30{3}	31{4}	>84.9	84.9
37{3,5}	30{3}	31{5}	1.8	14.3
37{3,6}	30{3}	31{6}	11.2	>25
37{3,8}	30{3}	31{8}	1.4	10.3
37{3,9}	30{3}	31{9}	11.7	>25
37{3,11}	30{3}	31{11}	8.1	>25
32{4,4}	31{4}	31{4}	10.2	>25
32{4,5}	31{4}	31{5}	1.7	>25
32{4,6}	31{4}	31{6}	4.8	>25
32{4,9}	31{4}	31{9}	8.2	>25
32{4,11}	31{4}	31{11}	>25	>25
32{5,5}	31{5}	31{5}	0.9	32.4
32{5,6}	31{5}	31{6}	0.2	>25
32{5,7}	31{5}	31{7}	1.7	>25
32{5,8}	31{5}	31{8}	0.2	>25
32{5,9}	31{5}	31{9}	0.5	>25
32{5,10}	31{5}	31{10}	2.4	>25
32{5,11}	31{5}	31{11}	1.6	>25
32{6,6}	31{6}	31{6}	>59.5	59.5
32{6,7}	31{6}	31{7}	2.0	>25
32{6,8}	31{6}	31{8}	0.03	>25
32{6,9}	31{6}	31{9}	>25	>25
32{6,10}	31{6}	31{10}	>25	>25
32{6,11}	31{6}	31{11}	18.4	>125
32{7,7}	31{7}	31{7}	>11.7	11.7
32{7,8}	31{7}	31{8}	0.5	>25
32{7,9}	31{7}	31{9}	2.5	>25
32{7,11}	31{7}	31{11}	2.7	>25

Compuesto	R ₁ NH ₂	R ₂ NH ₂	EC ₅₀ / μg·mL ⁻¹	CC ₅₀ / μg·mL ⁻¹
32{8,8}	31{8}	31{8}	0.008	>25
32{8,9}	31{8}	31{9}	0.03	>25
32{8,10}	31{8}	31{10}	0.4	>25
32{8,11}	31{8}	31{11}	0.5	>25
32{9,9}	31{9}	31{9}	9.5	>125
32{9,10}	31{9}	31{10}	>25	>25
32{9,11}	31{9}	31{11}	9.1	>25
32{10,10}	31{10}	31{10}	>85.7	85.7
32{10,11}	31{10}	31{11}	>25	>25
32{11,11}	31{11}	31{11}	>125	>125
AMD3100	-	-	0.001	>5
DS	-	-	0.011	>125



Se analiza el resultado de los valores de citotoxicidad obtenidos para los compuestos de la quimioteca y se observa que en general los compuestos no presentan toxicidad a la máxima concentración ensayada.

Un análisis del resultado de los valores de actividad obtenidos para los compuestos de la quimioteca permite descartar las dihidrazonas **33** como compuestos anti-VIH, ya que ninguna de las sintetizadas ha mostrado actividad, sus valores de EC₅₀ son mayores que 25 μg/mL en todos los casos, aunque tampoco muestran toxicidad a esta concentración ni a concentraciones inferiores. Por lo tanto, la baja solubilidad acuosa *in vivo* predicha en las propiedades ADME para **33{2,2}** no resulta relevante. Sin embargo, cuando se sustituye un grupo hidrazono por un grupo amino (estructuras del tipo aminohidrazona **37**) se obtiene actividad anti-VIH. Dos de los tres intermedios hidrazonoaldehído **36** presentan actividades del orden de 10 μg/mL.

Al examinar los valores de EC₅₀ de los compuestos simétricos, que están representados en la diagonal de la Tabla 1.5, se observa que los *building blocks* más activos son las aminas **31{4}**, **31{5}**, **31{8}** y **31{9}**, destaca especialmente la amina **31{8}** ya que el compuesto **32{8,8}** es el más activo de la quimioteca, con un EC₅₀ en la escala de los ng/mL. Además, en el grupo de los compuestos de sustitución no simétrica, aquellos que incorporan la amina **31{8}** en su estructura han resultado más activos, entre 0.008 y 1.4 μg/mL. Incluso el intermedio aminoaldehído **35{8}** presenta un valor de EC₅₀ de 1.8 μg/mL. La excepción es **37{1,8}** cuyo EC₅₀ es superior a 4.1 μg/mL, valor que corresponde a su citotoxicidad expresada

como CC_{50} . Los compuestos de la familia de la amina **31{5}** también han mostrado unos buenos valores de actividad, entre 2.7 y 0.2 $\mu\text{g/mL}$

Se observa que las estructuras de los tres compuestos más activos, **32{6,8}**, **32{8,8}** y **32{8,9}** (Figura 1.29), tienen en común la presencia de un *linker* trimetilénico entre el heterociclo y el átomo de nitrógeno en posición bencílica. Por lo tanto se comparan los resultados de actividad obtenidos en los compuestos con *linker* trimetilénico y etilénico, es decir los constituidos por las aminas **31{4}** frente a **31{5}** y **31{10}** frente a **31{11}** (Tabla 1.6). Se observa que en los compuestos con *linker* trimetilénico en el sistema polinitrogenado presentan en todos los casos mayor actividad anti-VIH que los compuestos con *linker* etilénico.

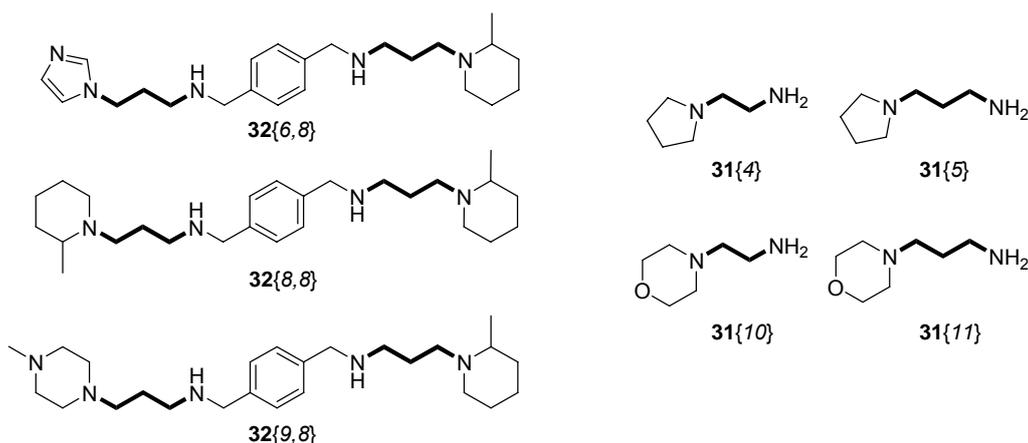


Figura 1.29. Estructura de los compuestos más activos de la quimioteca y las aminas con *linker* etilénico **31{4}** y **31{10}**, y con *linker* trimetilénico **31{5}** y **31{11}**.

La comparación de los valores de actividad antiviral obtenidos para los compuestos cuyos sistemas polinitrogenados presentan un *linker* trimetilénico y un heterociclo pirrolidínico **31{5}** o imidazólico **31{6}** (Figura 1.30) permite concluir que en general la pirrolidina rinde compuestos más activos que el imidazol, con la excepción del compuesto **32{6,8}**, una de las tres moléculas más activas (Tabla 1.6).

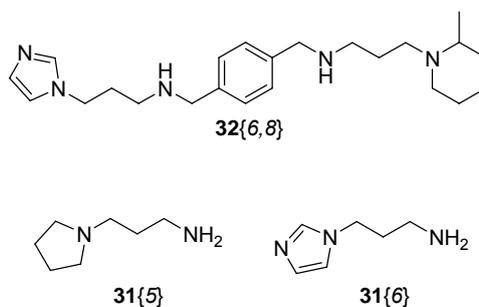


Figura 1.30. Estructura del compuesto **32{6,8}** y las aminas con *linker* trimetilénico y heterociclo de pirrolidina **31{5}** e imidazol **31{6}**.

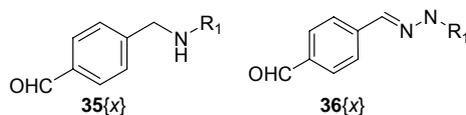
Tabla 1.5. Representación de la actividad anti-VIH, EC_{50} en $\mu\text{g/mL}$ de los compuestos en la quimioteca, donde R_1 y R_2 se refieren a los building blocks poliaminados **30** y **31** y las celdas se colorean según la actividad antiviral, azul oscuro para los más activos y azul claro para los menos activos.

R_1 R_2	1	2	3	4	5	6	7	8	9	10	11
1	>25										
2		>25									
3	>25		>25								
4		14.7	>25	10.2							
5	2.7	2.0	1.8	1.7	0.9						
6	>25	>25	11.2	4.8	0.2	>25					
7					1.7	2.0	>11.7				
8	>4.1	0.60	1.4		0.17	0.03	0.44	0.008			
9	>9.8	3.8	11.7	8.2	0.5	>25	2.5	0.03	9.5		
10					2.4	>25		0.4	>25	>25	
11	10.6	15.7	8.1	>25	1.6	18.4	2.7	0.5	9.1	>25	>25

Además se determina la actividad anti-VIH y la toxicidad de los 11 intermedios aminoaldehído **35** e hidrazonoaldehído **36**, el resultado se muestra en la Tabla 1.6. Se observa que el *building blocks* más activo **31**{8} rinde el intermedio **35** más activo, **35**{8}. Dos compuestos del tipo aminoaldehído **35**, **35**{5} y **35**{9}, presentan actividad antiviral baja y el resto de compuestos **35** no son activos a concentraciones inferiores a 25 $\mu\text{g/mL}$. De los tres hidrazonoaldehídos **36** evaluados, dos de ellos presentan actividad antiviral baja, **36**{1} y **36**{2}, y el tercero **36**{3} resulta inactivo a concentraciones inferiores a 25 $\mu\text{g/mL}$.

Tabla 1.6. Valores de actividad anti-VIH (EC_{50}) y de citotoxicidad (CC_{50}) expresados en $\mu\text{g/mL}$ de los compuestos sintetizados y de los compuestos de referencia AMD3100 (inhibidor de la unión del virus a CXCR4) y DS (inhibidor de la unión del virus a la célula).

Compuesto	$R_1\text{NH}_2$	$EC_{50} / \mu\text{g}\cdot\text{mL}^{-1}$	$CC_{50} / \mu\text{g}\cdot\text{mL}^{-1}$
36{1}	30{1}	12.6	>25
36{2}	30{2}	9.0	>25
36{3}	30{3}	>52.6	52.6
35{4}	31{4}	>25	>25
35{5}	31{5}	11.3	>25
35{6}	31{6}	>25	>25
35{7}	31{7}	>25	>25
35{8}	31{8}	1.8	12.1
35{9}	31{9}	16.1	>25
35{10}	31{10}	>25	>25
35{11}	31{11}	>25	>25
AMD3100	-	0.001	>5
DS	-	0.011	>125



1.6.2. Ampliación de la quimioteca inicial a partir del compuesto más activo 32{8,8}.

La estructura que ha presentado un mejor valor de actividad antiviral está formada por dos unidades de pipercolina, donde no se especifica la estereoisomería del carbono sobre el cual está dicho metilo y no se dispone de un referente sin metilo que permita comprobar la influencia de dicho grupo sobre la interacción con el correceptor CXCR4. Se sintetiza también un compuesto análogo 32{19,19} donde se sustituyen los heterociclos pipercolínicos por piperidínicos (Figura 1.31), sin embargo su EC_{50} es de $0.14 \mu\text{g/mL}$, por lo tanto la influencia del metilo presente en la pipercolina 31{8} incrementa considerablemente la actividad antiviral.

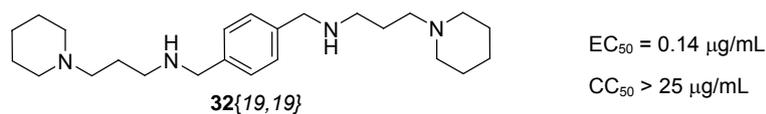


Figura 1.31. Actividad antiviral (EC₅₀) y citotoxicidad (CC₅₀) del compuesto **32{19,19}**.

Por otra parte, si se compara este compuesto **32{19,19}** (EC₅₀ = 0.14 μg/mL), cuyo heterociclo es una piperidina, con **32{5,5}** (EC₅₀ = 0.9 μg/mL), cuyo heterociclo es una pirrolidina, se observa que el primero presenta una actividad antiviral prácticamente un orden de magnitud superior. Estos resultados, junto con el obtenido para **32{6,6}** (EC₅₀ > 25 μg/mL), indican que los heterociclos con mayor flexibilidad conformacional presentan una mayor actividad antiviral.

Se ha estudiado también la influencia de la posición de los sustituyentes metilénicos del anillo aromático del *linker* en la actividad antiviral. Como referencia se ha utilizado la estructura del compuesto más activo **32{8,8}** y se ha sustituido el *linker* *p*-fenilénbismetilénico por *m*- y *o*-fenilénbismetilénico para obtener las correspondientes aminas **50{8,8}** y **52{8,8}** (Figura 1.32) mediante reacciones de aminación reductora del isoftalaldehído (**49**) y el *o*-ftalaldehído (**51**) respectivamente. En la síntesis de **50{8,8}** no se obtiene la molécula objetivo **52{8,8}** debido a la posible ciclación intramolecular producida en el término de monosustitución que rendiría la ftalimidina **53**, ya que en el espectro de ¹H-RMN no se observa la señal correspondiente a los protones de los aldehídos y en el espectro de IR se observa la banda correspondiente a la tensión C=O de una lactama de 5 eslabones a 1693 cm⁻¹.^[56]

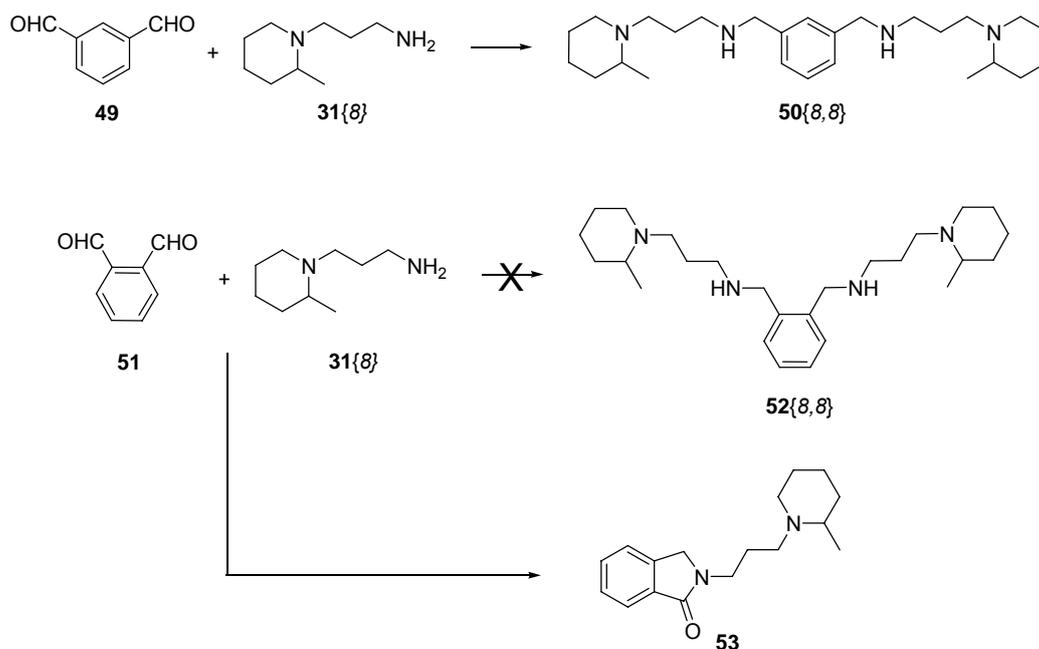


Figura 1.32. Obtención de los compuestos **50{8,8}**, **52{8,8}**, y **53**.

Se hallan descritas en la literatura reacciones de formación de ftalimidinas a partir de *o*-ftalaldehído (**51**) y aminas primarias.^[57,58] Se describen dos tipos de mecanismos: un mecanismo clásico^[59] que pasa por un intermedio tipo diol, y un mecanismo basado en un reordenamiento sigmatrópico [1,5]-H a partir de un intermedio monoimínico (Figura 1.33).^[60]

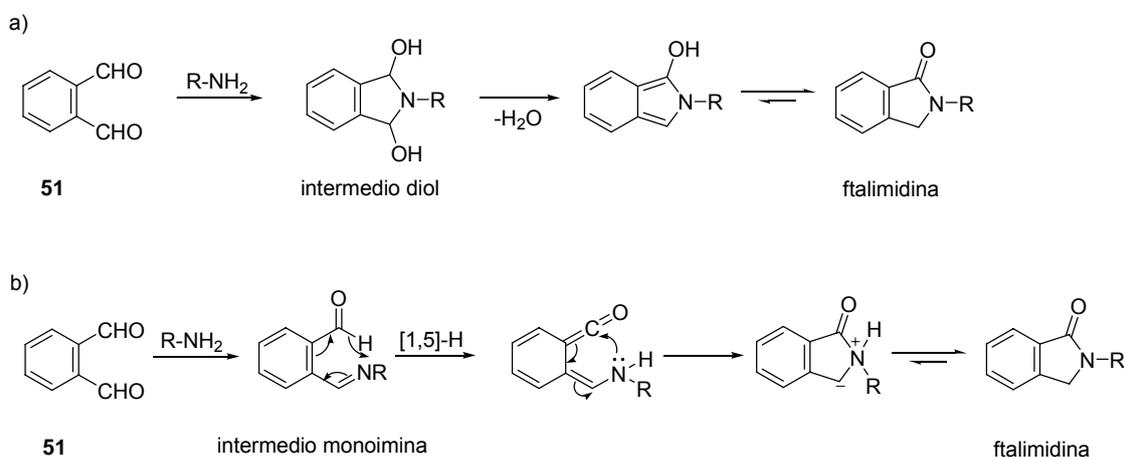


Figura 1.33. Obtención de ftalimidinas a partir de *o*-ftalaldehído (**51**) y aminas primarias. a) Mecanismo clásico a través del intermedio de tipo diol. b) Mecanismo de reordenamiento sigmatrópico [1,5]-H a partir del intermedio del tipo monoimina.

Por lo tanto se compara la actividad antiviral de **32**{8,8} y **50**{8,8}. Se observa que el término de sustitución en *meta* rinde un valor de EC_{50} de 0.07 $\mu\text{g/mL}$ (Figura 1.34), menor que el término de sustitución en *para*, de esta manera se comprueba que el *linker* *p*-fenilenbismetilénico permite una mejor interacción con el correceptor CXCR4, de manera análoga a lo que ocurre en el caso de los biciclamos.

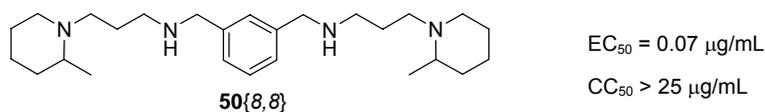


Figura 1.34. Actividad antiviral (EC_{50}) y citotoxicidad (CC_{50}) del compuesto **50**{8,8}.

Por último, en colaboración con la Sra. Corredor del Laboratorio de Síntesis del IQS,^[61] se sintetizan una serie de compuestos análogos a los de la quimioteca diseñada, donde se utiliza un *linker* con un grupo metilo sobre los carbonos en posición bencílica. La obtención de los compuestos **56** se realiza a partir de 1,4-diacetilbenceno (**54**) y la correspondiente amina **31**. Sin embargo, la presencia del metilo aumenta el impedimento estérico respecto al análogo

tereftalaldehído y además estabiliza la δ^+ del átomo de carbono del carbonilo cetónico haciéndolo menos reactivo que su análogo tereftalaldehído, de manera que la reacción de aminación reductora no presenta conversión completa. Es necesario el uso de un ácido de Lewis, el isopropóxido de titanio (IV), como catalizador para favorecer la formación del intermedio imínico **55** (Figura 1.35).^[29,31,44,62,63]

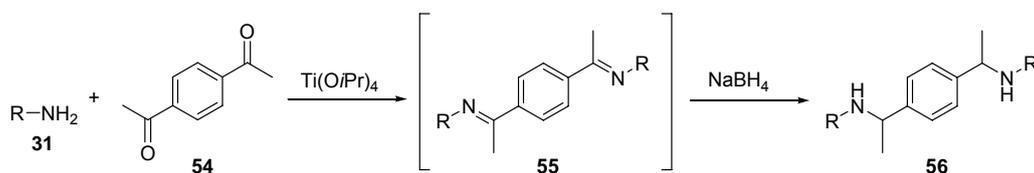


Figura 1.35. Reacción de aminación reductora de la dicetona **54** con las aminas **31** en presencia de isopropóxido de titanio (IV) para rendir las poliaminas **56** tras la etapa de reducción de las iminas **55**.

La influencia de este grupo sobre la actividad anti-VIH es diferente en cada caso estudiado (Tabla 1.7). En el caso del compuesto **56**{19,19}, no presenta ningún efecto, ya que en ambos casos, **32**{19,19} y **56**{19,19}, el valor del EC₅₀ es de 0.1 $\mu\text{g/mL}$, mientras que en **56**{8,8} el metilo hace que disminuya la actividad antiviral en un orden de magnitud, pasa de 0.008 a 0.08 $\mu\text{g/mL}$. Por último, en **56**{11,11} los metilos en posición bencílica permiten pasar de un compuesto inactivo a uno ligeramente activo con un EC₅₀ de 7.7 $\mu\text{g/mL}$.

Tabla 1.7. Valores de actividad anti-VIH (EC₅₀) y de citotoxicidad (CC₅₀) expresados en $\mu\text{g/mL}$ de los compuestos sintetizados con el linker con metilo y sin metilo en la posición bencílica.

Estructura	R	EC ₅₀ / $\mu\text{g}\cdot\text{mL}^{-1}$	CC ₅₀ / $\mu\text{g}\cdot\text{mL}^{-1}$
	H	0.14	>25
	Me	0.13	>25
	H	0.008	>25
	Me	0.08	>25
	H	>25	>25
	Me	7.7	>25

1.6.3. Modo de acción.

Para confirmar la hipótesis inicial en que se diseña una quimioteca de inhibidores de la unión del VIH-1 al correceptor de entrada CXCR4, se realiza un estudio del modo de acción con los compuestos más activos de la quimioteca sintetizada. Dicho estudio se realiza mediante dos ensayos: tiempo de adición de los compuestos e inhibición de la unión de anticuerpos específicos de los receptores CD4, CXCR4 y CCR5, ambos estudios los han realizado en el grupo del Dr. Esté del Laboratorio de Retrovirología de la Fundació irsiCaixa.^[53]

El resultado del ensayo del modo de acción se realiza con los compuestos **32{6,8}**, **32{8,8}**, **32{8,9}** y **32{8,10}** (Figura 1.36). Las curvas obtenidas para cada uno de los compuestos se comparan con: un inhibidor de la fusión del virus y la célula (C34), un inhibidor de la unión de la proteína viral gp120 al receptor celular CD4 (DS), un inhibidor de la unión al correceptor CXCR4 (AMD3100) y un inhibidor de la RT (AZT). Se observa que en los compuestos más activos de la quimioteca presentan un perfil semejante al del AMD3100, actúan después de la unión de gp120 al receptor celular CD4 (inhibida por DS) y antes de la fusión de las membranas viral y celular (inhibida por C34). Los compuestos ensayados empiezan a perder su capacidad inhibitoria a partir del minuto 14 postinfección. Se puede afirmar que actúan en el mismo tiempo del ciclo viral en que actúa AMD3100, que inhibe la unión del virus al correceptor celular CXCR4.

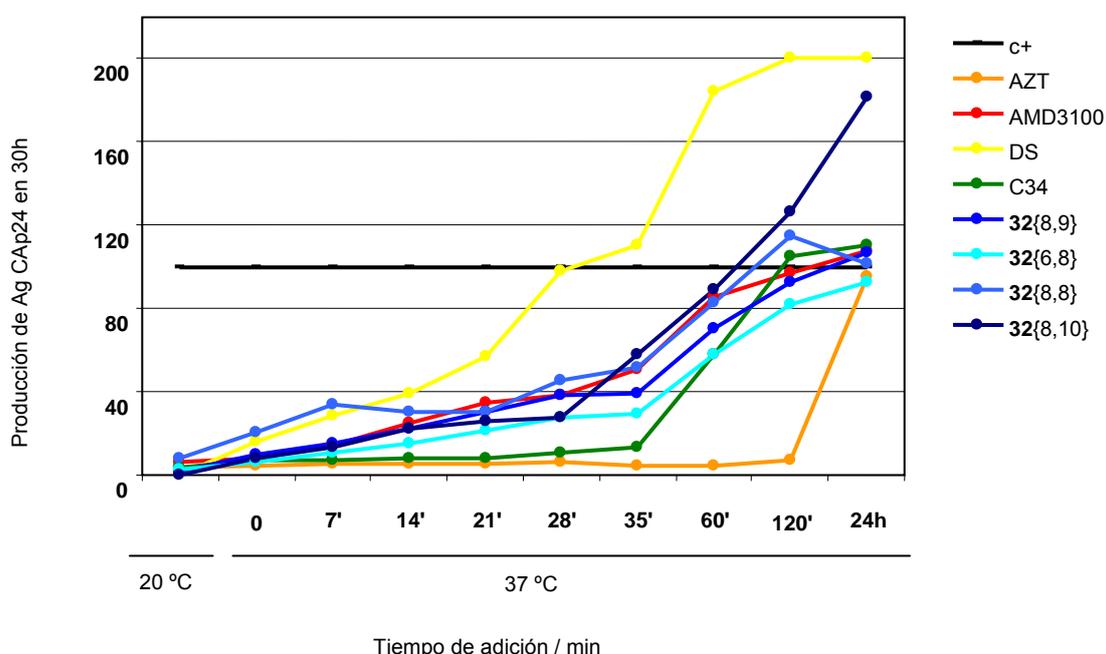


Figura 1.36. Gráfico del tiempo de adición de los compuestos más activos de la quimioteca sintetizada y comparación con fármacos de mecanismo de acción conocido. Se cuantifica mediante la producción del antígeno viral CAp24 a las 30 horas de postinfección de las células MT-4 con la cepa salvaje VIH-1_{NL4-3}.

Se realiza un estudio de inhibición de la unión de anticuerpos específicos de los receptores CD4, CXCR4 y CCR5 con los compuestos **32{6,8}**, **32{8,8}**, **32{8,9}** y **32{8,10}** y uno de actividad más reducida como control **32{7,9}**. Se evalúa el porcentaje del marcaje del anticuerpo monoclonal anti-CXCR4 humano 12G5 en células que se han incubado con cuatro concentraciones diferentes de cada uno de los compuestos ensayados (25, 5, 1 y 0.2 $\mu\text{g/mL}$), así como un control negativo (c-, nivel basal de autofluorescencia de las células sin marcar con ningún anticuerpo), un control positivo (c+, porcentaje de marcaje de cada receptor en las células CEM-CCR5 marcadas con los anticuerpos a analizar en ausencia de los compuestos) y AMD3100 en una concentración de 0.5 $\mu\text{g/mL}$ (antagonista de CXCR4). Se puede observar en la Figura 1.37 que los compuestos inhiben el marcaje del anticuerpo específico de CXCR4 de manera dosis-dependiente y muestran un porcentaje de inhibición relacionado con su capacidad de inhibición en el ensayo de actividad antiviral en células MT-4, es decir, el compuesto **32{7,9}** que presenta el menor EC_{50} de este grupo de cinco compuestos resulta ser el menos activo en la inhibición del marcaje de CXCR4, mientras que los tres compuestos más activos **32{6,8}**, **32{8,8}** y **32{8,9}** muestran una inhibición superior o igual al 50% a concentraciones de 1, 5 y 25 $\mu\text{g/mL}$.

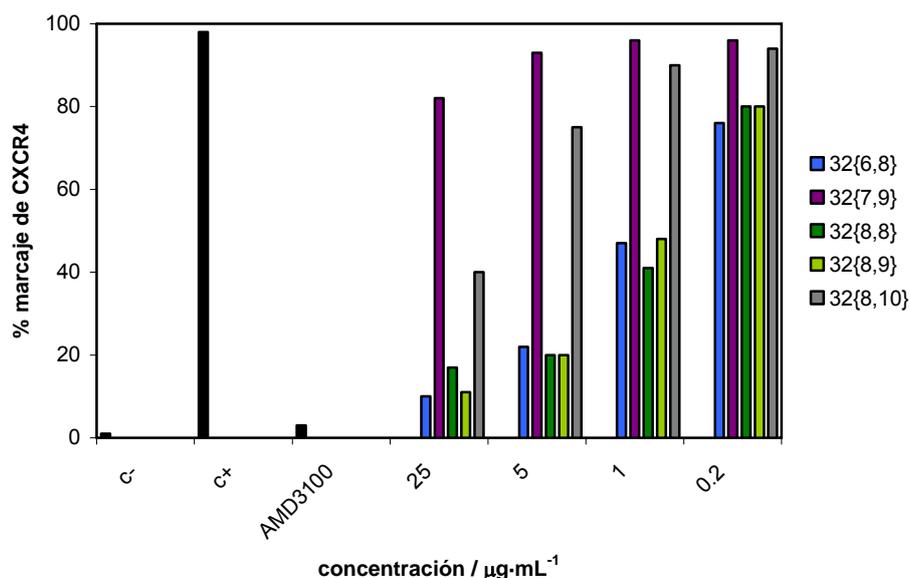


Figura 1.37. Gráfico del porcentaje de unión de anticuerpo monoclonal anti-CXCR4 en células CEM-CCR5 en presencia de los compuestos **32{6,8}**, **32{7,9}**, **32{8,8}**, **32{8,9}** y **32{8,10}** en concentraciones de 25, 5, 1 y 0.2 $\mu\text{g/mL}$.

Además, ninguno de los compuestos ensayados muestra inhibición del marcaje del resto de receptores (CD4, CCR5 y CD45), por lo tanto, se confirma la inhibición específica del correceptor CXCR4 (Figura 1.38).

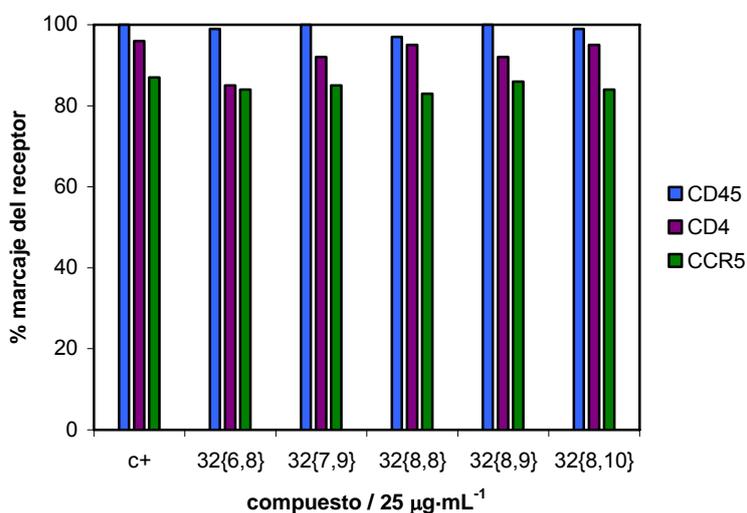


Figura 1.38. Gráfico del porcentaje de unión de los anticuerpos monoclonales anti-CD45 (azul), anti-CD4 (lila) y anti-CCR5 (verde) en células CEM-CCR5 en presencia de cinco de los compuestos de la quimioterapia a la máxima concentración ensayada de 25 µg/mL.

En conclusión, se han conseguido inhibidores del correceptor CXCR4 a partir de *building blocks* comerciales con actividad parecida a la del biciclamo AMD3100 pero solamente con 4 nitrógenos en vez de los 8 presentes en el AMD3100. Por lo tanto se estima una reducción de la toxicidad cardíaca y la falta de biodisponibilidad oral relacionados con el AMD3100 y su elevada carga positiva a pH fisiológico.

Bibliografía.

- [1] C. C. Bleul, M. Farzan, H. Choe, C. Parolin, I. Clark-Lewis, J. Sodroski, T. A. Springer; The lymphocyte chemoattractant SDF-1 is a ligand for LESTR/fusin and block HIV-1 entry. *Nature* **1996**, *382*, 829-833.
- [2] H. Deng, R. Liu, W. Willmeier, S. Choe, D. Unutmaz, M. Burkhart, P. Di Marzio, S. Marmon, R. E. Sutton, C. M. Hill, C. B. Davis, S. C. Peiper, T. J. Schall, D. R. Littman, N. R. Landau; Identification of a major co-receptor for primary isolates of HIV-1. *Nature* **1996**, *381*, 661-666.
- [3] S. Hatse, K. Princen, E. de Clercq, M. M. Rosenkilde, T. W. Schwartz, P. E. Hernandez-Abad, R. T. Skerlj, G. J. Bridger, D. Schols; AMD3465, a monomacrocyclic CXCR4 antagonist and potent HIV entry inhibitor. *Biochem. Pharmacol.* **2005**, *70*, 752-761.
- [4] W. Zhan, Z. Liang, A. Zhu, S. Kurtkaya, H. Shim, J. P. Snyder, D. C. Liotta; Discovery of small molecule CXCR4 antagonists. *J. Med. Chem.* **2007**, *50*, 5655-5664.
- [5] *Cerius2* 6.6; Accelrys Inc.: San Diego, CA, **2001**.
- [6] R. Pascual, J. I. Borrell, J. Teixido; Analysis of selection methodologies for combinatorial library design. *Mol. Divers.* **2003**, *6*, 121-133.
- [7] R. Pascual; Estudios sobre la metodología computacional para el diseño de bibliotecas combinatorias: Desarrollo del programa Pralins y aplicaciones en el campo del HIV. Tesis doctoral, IQS (Universitat Ramon Llull), Barcelona, **2003**.
- [8] M. O. Rabal; Herramientas de cribado virtual aplicadas a inhibidores de tirosina quinasas. Contribución al desarrollo del programa PRALINS para el diseño de quimiotecas combinatorias. Tesis doctoral, IQS (Universitat Ramon Llull), Barcelona, **2006**.
- [9] *MOE (Molecular Operating Environment)* 2004 Release; Chemical Computing Group, Inc.: Montreal, Canada, **2004**.
- [10] R. Todeschini, and V. Consonni; Handbook of molecular descriptors. Eds. R. Mannhold, H. Kubinyi, H. Timmerman, Wiley-VCH, Weinheim. **2000**.
- [11] G. N. Lance, W. T. Williams; A general theory of classificatory sorting strategies. *Computer J.* **1967**, *9*, 373-380.
- [12] S. Ekins, J. Mestres, B. Testa; *In silico* pharmacology for drug discovery: methods for virtual ligand screening and profiling. *Br. J. Pharmacol.* **2007**, *152*, 9-20.
- [13] T. J. Pallasch; Principles of pharmacotherapy: II. Pharmacokinetics. *Anesth. Prog.* **1988**, *35*, 133-146.
- [14] W. E. Lindup, M. C. L'E Orme; Plasma protein binding of drugs. *Br. Med. J.* **1981**, *282*, 212-214.
- [15] T. Lynch, A. Price; The effect of cytochrome P450 metabolism on drug response, interaction, and adverse effects. *Am. Fam. Physician* **2007**, *76*, 391-396.
- [16] *Accord for Excel* Version 6.2; Accelrys Software Inc.: San Diego, CA, **2005**.

- [17] W. J. Egan, K. M. Merz, J. J. Baldwin; Prediction of drug absorption using multivariate statistics. *J. Med. Chem.* **2000**, *43*, 3867-3877.
- [18] A. Cheng, K. M. Merz; Prediction of aqueous solubility of a diverse set of compounds using quantitative structure-property relationships. *J. Med. Chem.* **2003**, *46*, 3572-3580.
- [19] S. L. Dixon, H. O. Villar; Investigation of classification methods for the prediction of activity in diverse chemical libraries. *J. Comput. Aided Mol. Des.* **1999**, *13*, 533-545.
- [20] R. G. Susnow, S. L. Dixon; Use of robust classification techniques for the prediction of human cytochrome P450 2D6 inhibition. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1308-1315.
- [21] S. L. Dixon, K. M. Merz; One-dimensional molecular representations and similarity calculations: methodology and validation. *J. Med. Chem.* **2001**, *44*, 3795-3809.
- [22] S. Gomez, J. A. Peters, T. Maschmeyer; The reductive amination of aldehydes and ketones and the hydrogenation of nitriles: mechanistic aspects and selectivity control. *Adv. Synth. Catal.* **2002**, *344*, 1037-1057.
- [23] W. S. Emerson, C. A. Uraneck; Secondary and tertiary amines from nitro compounds. *J. Am. Chem. Soc.* **1941**, *63*, 749-751.
- [24] H. E. Johnson, D. G. Crosby; N-Alkylation of amides. A novel procedure. *J. Org. Chem.* **1962**, *27*, 2205-2205.
- [25] I. V. Micovic, M. D. Ivanovic, D. M. Piatak, V. D. Bojic; A simple method for preparation of secondary aromatic amines. *Synthesis* **1991**, 1043-1045.
- [26] B.-C. Chen, J. E. Sundeen, P. Guo, M. S. Bednarz, R. Zhao; Novel triethylsilane mediated reductive N-alkylation of amines: improved synthesis of 1-(4-imidazolyl)methyl-4-sulfonylbenzodiazepines, new farnesyltransferase inhibitors. *Tetrahedron Lett.* **2001**, *42*, 1245-1246.
- [27] T. Suwa, E. Sugiyama, I. Shibata, A. Baba; Reductive amination promoted by tributyltin hydride. *Synlett.* **2000**, 556-558.
- [28] A. F. Abdel-Magid, K. G. Carson, B. D. Harris, C. A. Maryanoff, R. D. Shah; Reductive amination of aldehydes and ketones with sodium triacetoxyborohydride. Studies on direct and indirect reductive amination procedures. *J. Org. Chem.* **1996**, *61*, 3849-3862.
- [29] S. Bhattacharyya; Reductive alkylation of dimethylamine using titanium(IV) isopropoxide and sodium borohydride: an efficient, safe, and convenient method for the synthesis of N,N-dimethylated tertiary amines. *J. Org. Chem.* **1995**, *60*, 4928-4929.
- [30] B. T. Cho, S. K. Kang; Direct and indirect reductive amination of aldehydes and ketones with solid acid-activated sodium borohydride under solvent-free conditions. *Tetrahedron* **2005**, *61*, 5725-5734.
- [31] H. J. Kumpaty, J. S. Williamson, S. Bhattacharyya; Synthesis of N-methyl secondary amines. *Synth. Commun.* **2003**, *33*, 1411-1416.
- [32] K. A. Schellenberg; The synthesis of secondary and tertiary amines by borohydride reduction. *J. Org. Chem.* **1963**, *28*, 3259-3261.
- [33] F. S. Varma, R. Dahiya; Sodium borohydride on wet clay: solvent-free reductive amination of carbonyl compounds using microwaves. *Tetrahedron* **1998**, *54*, 6293-6298.

- [34] G. N. Walker, M. A. Moore, B. N. Weaver; Application of sodium borohydride reduction to synthesis of substituted aminopiperidines, aminopiperazines, aminopyridines, and hydrazines. *J. Org. Chem.* **1961**, *26*, 2740-2747.
- [35] R. F. Borch, M. D. Bernstein, H. D. Durst; The cyanohydridoborate anion as a selective reducing agent. *J. Am. Chem. Soc.* **1971**, *93*, 2897-2904.
- [36] R. O. Hutchins, M. Markowitz; Tetraalkylammonium trihydridocyanoborates. Versatile, selective reagents for reductive aminations in nonpolar media. *J. Org. Chem.* **1981**, *46*, 3571-3574.
- [37] A. F. Abdel-Magid, S. J. Mehrman; A review on the use of sodium triacetoxyborohydride in the reductive amination of ketones and aldehydes. *Org. Process Res. Dev.* **2006**, *10*, 971-1031.
- [38] G. W. Gribble; The synthetic versatility of acyloxyborohydrides. *Org. Process Res. Dev.* **2006**, *10*, 1062-1075.
- [39] C. D. Gutierrez, V. Bavetsias, E. McDonald; $\text{TiCl}(\text{iOPr})_3$ and $\text{NaBH}(\text{OAc})_3$: an efficient reagent combination for the reductive amination of aldehydes by electron-deficient amines. *Tetrahedron Lett.* **2007**, *46*, 3595-3597.
- [40] N. M. Yoon, E. G. Kim, H. S. Son, J. Choi; Borohydride exchange resin, a new reducing agent for reductive amination. *Synth. Commun.* **1993**, *23*, 1595-1599.
- [41] M. D. Bomann, I. C. Guch, M. DiMare; A mild, pyridine-borane-based reductive amination protocol. *J. Org. Chem.* **1995**, *60*, 5995-5996.
- [42] A. E. Moormann; Reductive amination of piperidines with aldehydes using borane-pyridine. *Synth. Commun.* **1993**, *23*, 789-795.
- [43] A. Pelter, R. M. Rosser, M. Stuart; Reductive aminations of ketones and aldehydes using borane-pyridine. *J. Chem. Soc. Perkin. Trans. I* **1984**, 717-720.
- [44] S. Bhattacharyya; Titanium(IV) isopropoxide and sodium borohydride: a reagent of choice for reductive amination. *Tetrahedron Lett.* **1994**, *35*, 2401-2404.
- [45] R. F. Borch, H. D. Durst; Lithium cyanohydridoborate, a versatile new reagent. *J. Am. Chem. Soc.* **1969**, *91*, 3996-3997.
- [46] G. W. Gribble; Sodium borohydride in carboxylic acid media: a phenomenal reduction system. *Chem. Soc. Rev.* **1998**, *27*, 395-404.
- [47] J. Kalia, R. T. Raines; Hydrolytic stability of hydrazones and oximes. *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 7523-7526.
- [48] K. B. Wiberg, R. Glaser; Resonance interactions in acyclic systems. 4. Stereochemistry, energetics, and electron distributions in 3-centre four- π -electron systems $\text{A}=\text{B}-\text{C}$. *J. Am. Chem. Soc.* **1992**, *114*, 841-850.
- [49] J. L. Jensen, R. Siegel; Effect of conjugation on the rates of the acid-catalyzed hydrolyses of acetals. *J. Org. Chem.* **1988**, *53*, 6105-6106.
- [50] C. Ribes, E. Falomir, J. Murga; Selective cleavage of acetals with ZnBr_2 in dichloromethane. *Tetrahedron* **2006**, *62*, 1239-1244.
- [51] C. Wiles, P. Watts, S. J. Haswell; Acid-catalysed synthesis and deprotection of dimethyl acetals in a miniaturised electroosmotic flow reactor. *Tetrahedron* **2005**, *61*, 5209-5217.

- [52] C. Yue, I. Gauthier, J. Royer, H. P. Husson; Concise and stereoselective syntheses of the eight natural ant defense alkaloids (+)-tetraoponerine-1 to (+)-tetraoponerine-8 according to the CN(*R,S*) strategy. *J. Org. Chem.* **1996**, *61*, 4949-4954.
- [53] I. Clotet-Codina; Determinació de l'activitat antiviral de nous compostos inhibidors del VIH-1. Tesis doctoral, Laboratori de Retrovirologia. Fundació IrsiCaixa. Hospital Universitari Germans Trias i Pujol, Barcelona, **2007**.
- [54] G. Moncunill, M. Armand-Ugón, I. Clotet, E. Pauls, E. Ballana; Anti-HIV activity and resistance profile of the CXCR4 antagonist POL3026. *Mol. Pharmacol.* **2008**, *73*, 1264-1273.
- [55] V. I. Pérez-Nueno, S. Pettersson, D. W. Ritchie, J. I. Borrell, J. Teixidó; Discovery of novel HIV entry inhibitors for the CXCR4 receptor by prospective virtual screening. *J. Chem. Inf. Model.* **2009**, *49*, 810-823.
- [56] E. Pretsch, P. Bühlmann, C. Affolter, A. Herrera, R. Martínez; Determinación estructural de compuestos orgánicos. Springer-Verlag Ibérica, Barcelona, **2001**.
- [57] I. Takahashi, K. Nishiuchi, R. Miyamoto, M. Hatanaka, H. Uchida, K. Isa, A. Sakushima, S. Hosoi; Reaction systems peripheral to the 1:2 Mannich condensation reaction between *o*-phthalaldehyde and primary amine. *Lett. Org. Chem.* **2005**, *2*, 40-43.
- [58] J. Wan, B. Wu, Y. Pan; Novel one-step synthesis of 2-carbonyl/thiocarbonyl isoindolinones and mechanistic disclosure on the rearrangement reaction of *o*-phthalaldehyde with amide/thioamide analogs. *Tetrahedron* **2007**, *63*, 9338-9344.
- [59] T. DoMinh, A. L. Johnson, J. E. Jones, P. P. Senise; Reactions of phthalaldehyde with ammonia and amines. *J. Org. Chem.* **1977**, *42*, 4217-4221.
- [60] M. Alajarín, P. Sánchez-Andrada, C. López-Leonardo, A. Álvarez; On the mechanism of phthalimidine formation via *o*-phthalaldehyde monoimines. New [1,5]-*H* sigmatropic rearrangements in molecules with the 5-aza-2,4-pentadienal skeleton. *J. Org. Chem.* **2005**, *70*, 7617-7623.
- [61] M. Corredor; Obtenció i optimització de caps de sèrie com inhibidors d'entrada del VIH. Trabajo Final de Carrera, IQS, Barcelona, **2009**.
- [62] R. J. Mattson, K. M. Pham, D. J. Leuck, K. A. Cowen; An improved method for reductive alkylation of amines using titanium(IV) isopropoxide and sodium cyanoborohydride. *J. Org. Chem.* **1990**, *55*, 2552-2554.
- [63] K. A. Neidigh, M. A. Avery, J. S. Williamson, S. Bhattacharyya; Facile preparation of *N*-methyl secondary amines by titanium(IV) isopropoxide-mediated reductive amination of carbonyl compounds. *J. Chem. Soc. Perkin. Trans. I* **1998**, 2527-2531.

CAPÍTULO 2. MONOCICLAMOS SUSTITUIDOS

2.1. Introducción a los monociclamos sustituidos.

El biciclamo AMD3100 (Figura 2.1) es el prototipo para los antagonistas no peptídicos del correceptor CXCR4.^[1] Su estructura está formada por dos unidades de 1,4,8,11-tetraazaciclotetradecano (ciclamo) conectadas mediante un *linker* *p*-fenilenbismetilénico. Sin embargo, el AMD3100 no presenta biodisponibilidad oral debido a su elevada carga positiva a pH fisiológico, de +2 por cada anillo de ciclamo. Por lo tanto, un tratamiento anti-VIH a largo plazo supone un gran inconveniente para el paciente en cuanto a la administración del fármaco.

En la literatura se encuentran descritos algunos ejemplos de monociclamos sustituidos, entre los que destacan AMD3451 y AMD3465, cuya estructura se muestra en la Figura 2.1.

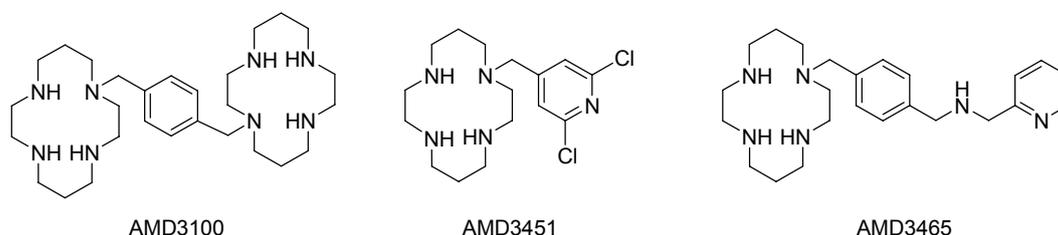


Figura 2.1. Estructura del biciclamo AMD3100 y de los monociclamos sustituidos AMD3451 y AMD3465.

El monociclamo AMD3451, desarrollado por AnorMED, es un antagonista dual de CCR5/CXCR4, es decir, presenta interacción con dichos correceptores,^[2] mientras que el AMD3465 inhibe selectivamente la unión del virus al correceptor CXCR4. Además, Hatse y colaboradores demuestran que no es esencial la presencia de dos anillos ciclámicos para que un compuesto presente actividad antiviral comparando la actividad antiviral de AMD3465 y AMD3100 (en células MT-4 infectadas con las cepas del VIH-1 III_B y NL4.3; IC₅₀ de 12.3 y 6.1 nM para el AMD3465, IC₅₀ de 12.4 y 7.4 nM para AMD3100). En este estudio, en el realizado por Rosenkilde *et al.* y en el estudio realizado en el Capítulo 1 del presente trabajo, se concluye que los ocho grupos amino no son necesarios para la actividad anti-VIH.^[3,4]

Además, en una patente publicada en 1999 (también de AnorMED) se describe la síntesis y evaluación antiviral de estos dos monociclamos junto con otros compuestos monociclámicos sustituidos por poliaminas **64**.^[5] La vía sintética descrita por AnorMED para la obtención de estos productos **64** se basa en cuatro etapas (Figura 2.2). La primera es la protección de tres de los cuatro grupos amino del ciclamo (**57**) con clorofosfonato de dietilo (**58**),^[6] con un rendimiento del 35%. Aunque también sugieren la protección de tres de los grupos amino del ciclamo (**57**) con tosilato o mesilato. A continuación realizan la alquilación del grupo amino no protegido del ciclamo **59** mediante el ataque nucleófilo sobre

α,α' -dibromo-*p*-xileno (**60**), con un rendimiento del 59%. La tercera etapa es análoga a la anterior pero en este caso el grupo amino nucleófilo es la poliamina **62** que formará el fragmento no ciclámico del compuesto objetivo **64** y el derivado halogenado es el bromoderivado ciclámico **61** que han obtenido en la etapa anterior. Y por último, desprotegen los tres nitrógenos del ciclamo **63** por tratamiento con ácido bromhídrico acuoso y ácido acético.

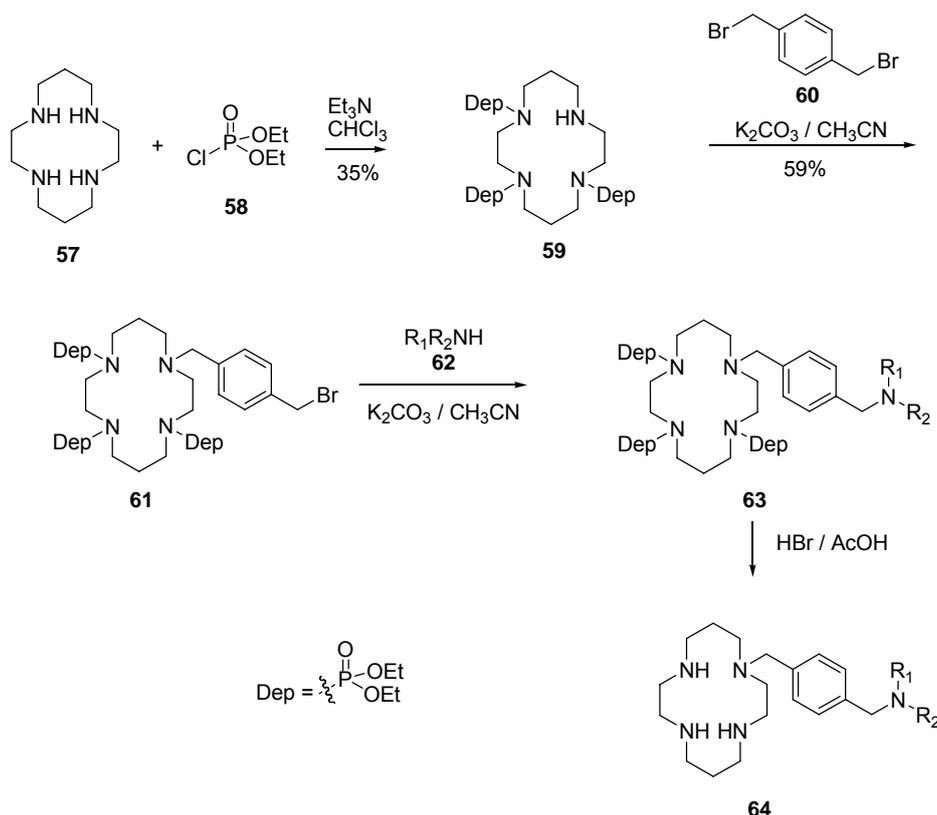
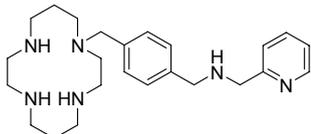
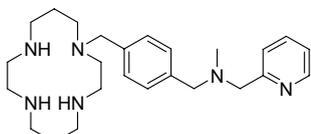
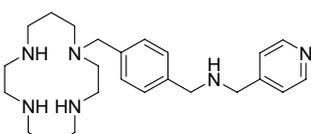
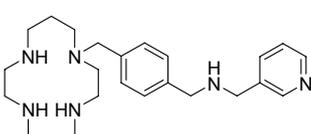
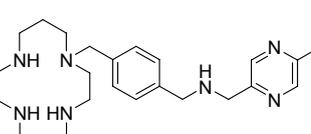
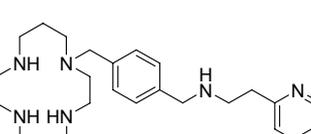


Figura 2.2. Esquema sintético utilizado por AnorMED para la obtención de compuestos monociclámicos sustituidos por poliaminas **64**. En primer lugar se protegen tres de los cuatro grupos amino del ciclamo (**57**) con clorofosfonato de dietilo (**58**). A continuación se produce el ataque nucleófilo del nitrógeno del grupo amino libre de **59** sobre α,α' -dibromo-*p*-xileno (**60**) seguido de otro ataque nucleófilo de poliamina R_1R_2NH (**62**) sobre el bromoderivado ciclámico **61**. Por último la desprotección de los grupos amino de **63** con HBr rinde los compuestos objetivo monociclámicos sustituidos por poliaminas **64**.

El resultado de los ensayos de la actividad anti-VIH de estos compuestos **64** sintetizados por AnorMED, expresada como EC_{50} en $\mu\text{g/mL}$, utilizando el método MTT en células MT-4 infectadas por VIH-1_{IIIB}, así como la citotoxicidad, expresada como CC_{50} en $\mu\text{g/mL}$, se muestra en la Tabla 2.1.

Tabla 2.1. Estructura de los monociclamos sustituidos **64** sintetizados en AnorMED, actividad anti-VIH expresada como EC_{50} y determinada sobre células MT-4 infectadas por VIH-1_{III}B y en presencia del monociclamo, y citotoxicidad expresada como CC_{50} y determinada sobre células MT-4 sin presencia de virus y en presencia del monociclamo.

Nombre	Estructura	EC_{50} / $\mu\text{g}\cdot\text{mL}^{-1}$	CC_{50} / $\mu\text{g}\cdot\text{mL}^{-1}$
AMD3645		0.008	>250
ADM3538		0.1	209
AMD3500		0.6	>250
AMD3499		1.8	>250
AMD3498		0.2	>250
AMD3497		1.8	>250

Nombre	Estructura	EC ₅₀ / μg·mL ⁻¹	CC ₅₀ / μg·mL ⁻¹
AMD3516		0.7	158
AMD3530		0.5	175
AMD3517		0.8	153
AMD3544		0.7	222
AMD3543		0.2	239
AMD3529		0.4	130

Otra estrategia, análoga a la anterior, es la utilizada por GlaxoSmithKline, que se muestra en la Figura 2.3.^[7,8] En este caso protegen tres de los grupos amino del ciclam (57) con *tert*-butilcarbamato (Boc) con un rendimiento del 19%.^[9] Después llevan a cabo la alquilación del nitrógeno libre de 66 con α,α' -dibromo-*p*-xileno (60) con un rendimiento del 90%. A continuación desplazan el bromo en posición bencílica del compuesto intermedio 67 con el nucleófilo apropiado, R₁R₂NH (62), y describen un rendimiento del 86% en el caso de utilizar

1-acetilpiperazina. Finalmente, realizan la desprotección de los grupos amino de **68** en medio ácido para rendir los compuestos objetivo **64**, en el caso de la 1-acetilpiperazina el rendimiento de esta etapa es del 34%.

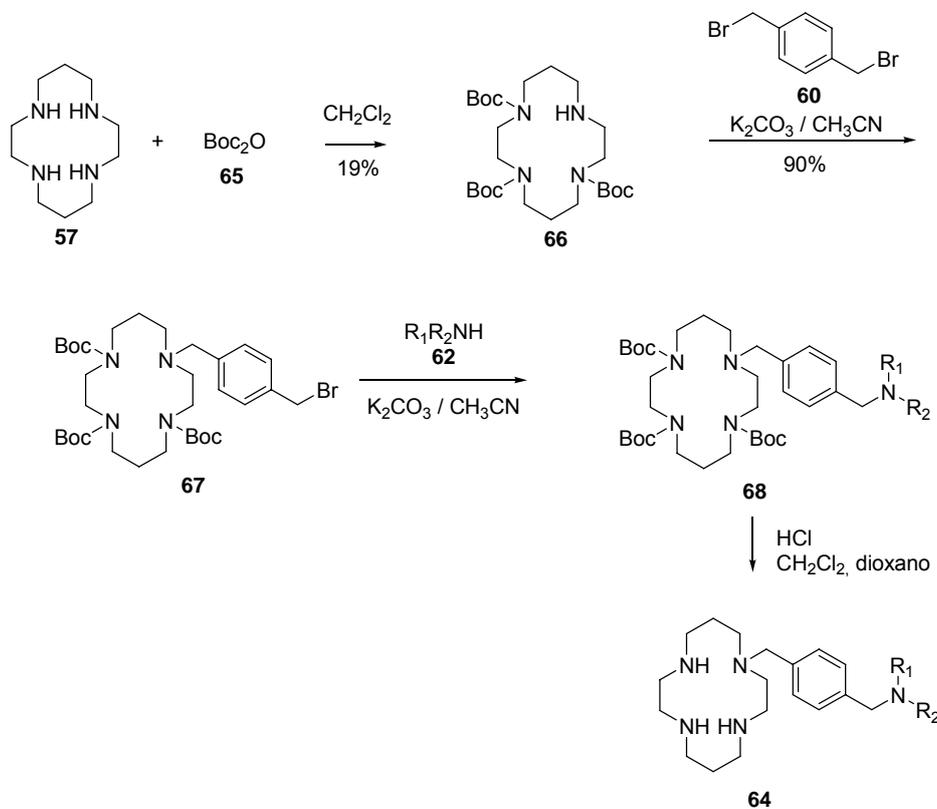


Figura 2.3. Esquema sintético utilizado por GlaxoSmithKline para la obtención de compuestos monociclámicos sustituidos por poliaminas **64**. En primer lugar se protegen tres de los cuatro grupos amino del ciclamo (**57**) con tert-butylcarbarnato. A continuación se produce el ataque nucleófilo del nitrógeno del grupo amino libre de **66** sobre α, α' -dibromo-*p*-xileno (**60**) seguido de otro ataque nucleófilo de poliamina R_1R_2NH (**62**) sobre el bromoderivado ciclámico **67**. Por último la desprotección de los grupos amino de **68** con HCl rinde los compuestos objetivo monociclámicos sustituidos por poliaminas **64**.

En el Laboratorio de Síntesis del GEM, el Sr. Palomo ha realizado la síntesis del biciclamo AMD3100 utilizando una estrategia análoga a la anterior a partir del ciclamo (**57**) y derivado bromado **60** (Figura 2.4). La primera etapa es la protección de tres de los grupo amino del ciclamo con Boc (33% de rendimiento), seguida de una segunda etapa de alquilación del ciclamo con α, α' -dibromo-*p*-xileno (**60**) (90% de rendimiento) y finalmente una tercera etapa de desprotección de los grupos Boc de **69** (90% de rendimiento).^[10] De esta manera se obtiene el biciclamo AMD3100, compuesto de referencia en los ensayos de actividad anti-VIH, con un rendimiento global del 27%.

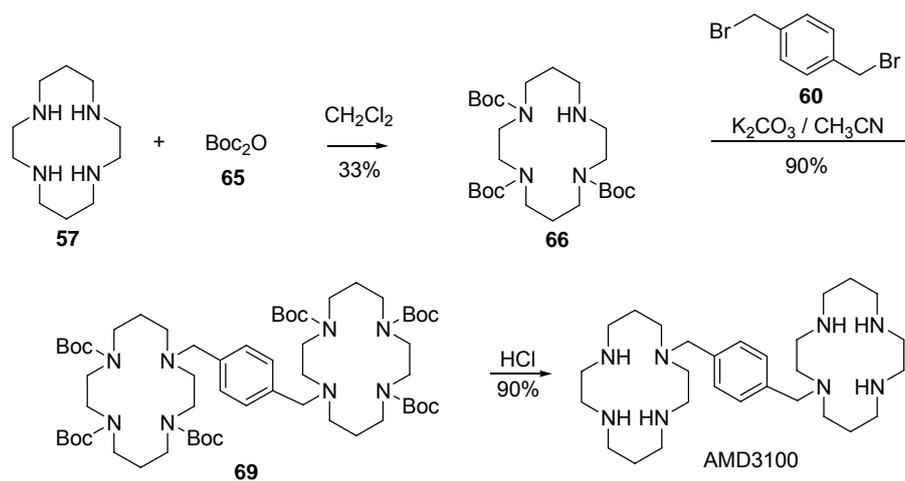


Figura 2.4. Esquema sintético utilizado por el Sr. Palomo del Laboratorio de Síntesis del IQS para la obtención del biciclamo AMD3100 con un rendimiento global del 27%.^[10]

2.2. Diseño y análisis retrosintético de monociclamos sustituidos.

Para reducir el número de grupos amino básicos y la redundancia estructural presente en el biciclamo AMD3100 (**24**), en el presente trabajo se ha diseñado una serie de monociclamos sustituidos **70**. Estos compuestos mantienen la estructura de un anillo de ciclamo unido al *linker* (fragmento ciclámico), pero no contienen la segunda unidad de ciclamo, ya que se sustituye por un compuesto polinitrogenado de estructura general **26** del grupo de las aminas utilizadas como sustituyentes R_1 y R_2 en el Capítulo 1 (fragmento no ciclámico), ver Figura 2.5.

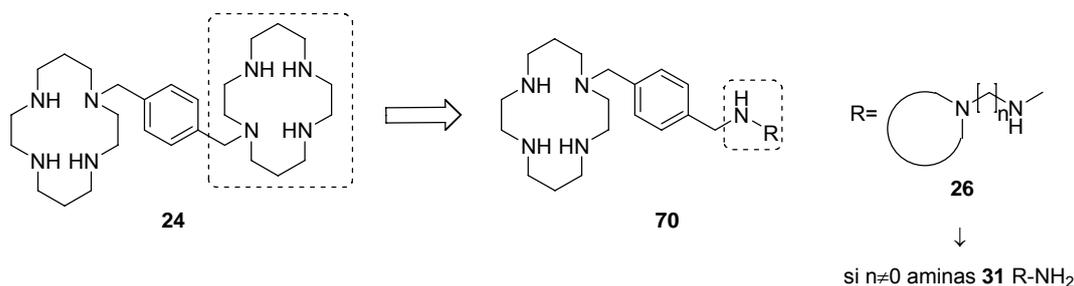


Figura 2.5. Estructura del AMD3100 (**24**) y de los monociclamos análogos **70** donde se sustituye uno de los anillos de ciclamo por un compuesto polinitrogenado **26**.

Los compuestos polinitrogenados que se utilizan para la sustitución del fragmento no ciclámico corresponden a las aminas alifáticas **31**{4-11} y se muestran en la Figura 2.6. Todas ellas están formadas por un heterociclo con mínimo un átomo de nitrógeno, como heteroátomo, unido a una cadena alifática (etileno o trimetileno) que lo separa de un grupo amino terminal. Los heterociclos utilizados son pirrolidina, imidazol, piperidina, pipercolina, piperazina y morfolina.

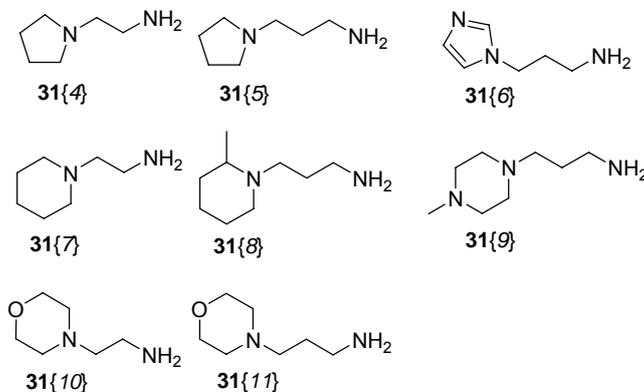


Figura 2.6. Estructura de las aminas **31** que forman el fragmento no ciclámico de los monociclamos **70**.

En el análisis retrosintético de los compuestos objetivo **70** se plantean dos estrategias, que se detallan en la Figura 2.7.

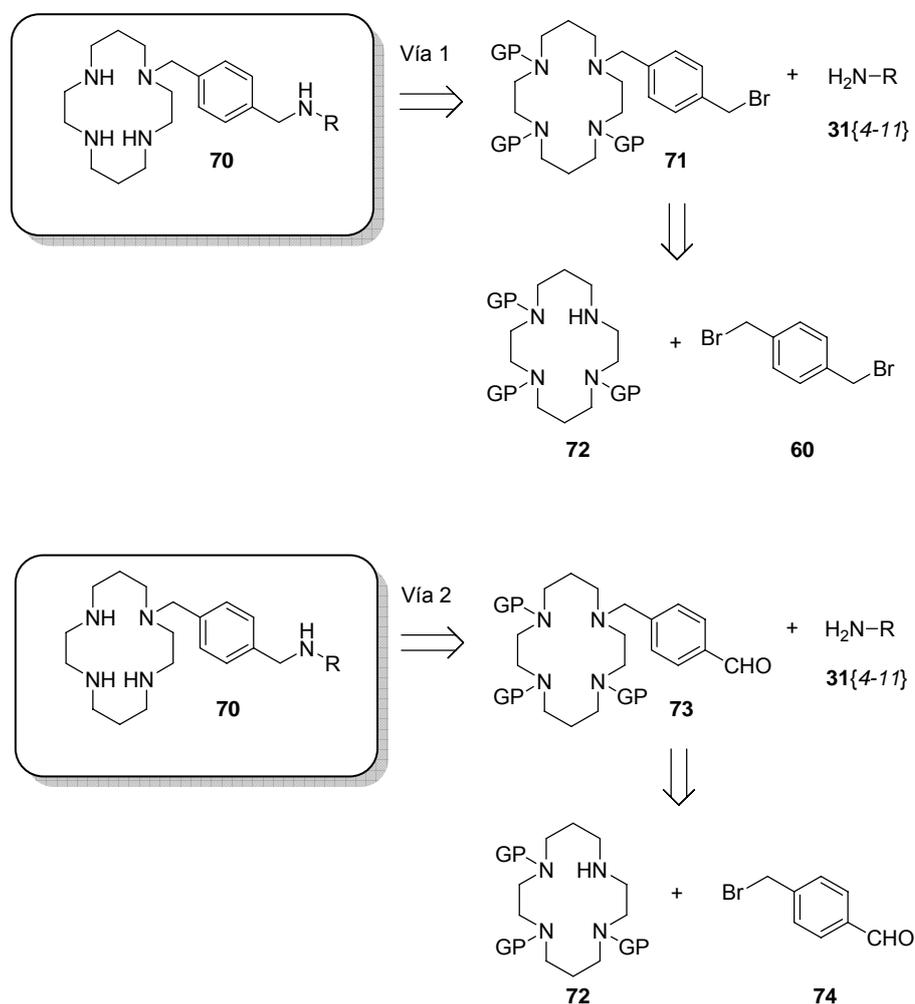


Figura 2.7. Análisis retrosintético de los monociclamos sustituidos con poliaminas **70**.

En el análisis retrosintético de la vía 1, tras la desconexión del enlace C(benzylico)-N(poliámínico) en **70** se obtiene la poliamina **31**, que es comercial, y el bromobencil ciclamo **71**, que no es comercial. En el bromobencil ciclamo **71** se desconecta el enlace N(ciclámico)-C(benzylico) rindiendo α,α' -dibromo-*p*-xileno (**60**), que es comercial, y el ciclamo triprotegido **72**, cuya síntesis está descrita.

La segunda estrategia sintética desconecta también en primer lugar el enlace C(benzylico)-N(poliámínico) de **70** generando el precursor poliamina **31** correspondiente pero el equivalente sintético del otro fragmento, en este caso, es el benzaldehído ciclámico **73**. A partir

de la desconexión del enlace N(ciclámico)-C(bencílico) en este último intermedio **73** se obtiene el ciclamo triprotegido **72** y 4-(bromometil)benzaldehído **74**, ambos productos se hallan descritos en la literatura.

En ambos casos, tres de los cuatro grupos amino del ciclamo **72** deben estar protegidos.

2.3. Protección N-1 del ciclamo.

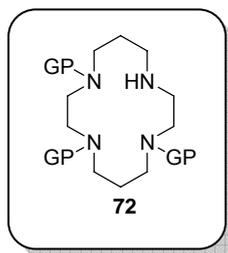


Figura 2.8. Estructura general del ciclamo triprotegido **72**.

La protección N-1 de poliazamacrociclos, donde N-1 se refiere a la protección de todos excepto uno de los grupos amino en la poliamina cíclica, suele ser problemática, ya que la reactividad de dichos grupos amino es idéntica y por tanto el rendimiento es generalmente moderado. Además, aislar el producto de protección N-1 de la mezcla obtenida suele ser difícil.

Se han desarrollado muchos grupos protectores para el grupo amino, entre los que destacan los carbamatos y las amidas. En el caso de poliazamacrociclos se ha optimizado el uso de varios grupos protectores como el *tert*-butoxicarbonilo (Boc), 2,2,2-tricloroetoxicarbonilo (Troc), dietoxifosforilo (Dep), tosilo (Ts) y trifluoroacetilo.

2.3.1. Formación de complejos metálicos.

Las primeras protecciones N-1 del ciclamo se realizaron por la formación de complejos tridentados de tricarbonilcromo con tres de los grupos amino del azamacrociclo, por reacción con hexacarbonilcromo (Figura 2.9). El intermedio complejoado **75** se utiliza para realizar alquilaciones sobre el grupo amino libre y finalmente se descompleja por oxidación al aire en medio ácido; dicho procedimiento presenta un rendimiento global descrito entre un 85% y un 95%.^[11]

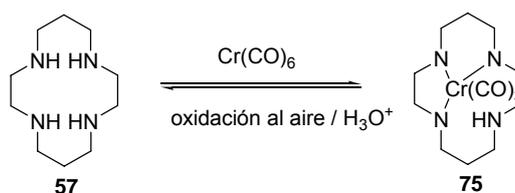


Figura 2.9. Reacción de protección N-1 del ciclamo (**57**) por formación del complejo metálico de tricarbonilcromo **75** y su desprotección por oxidación al aire en medio ácido.

En el mismo grupo de investigación también realizaron la protección N-1 del ciclamo formando un complejo metálico de boro con tris(dimetilamino)borano **76** (Figura 2.10) dejando un grupo amino libre con el que se llevaron a cabo varias alquilaciones *in situ* con rendimientos globales entre un 85% y un 90%. El gran inconveniente de este grupo protector es que el complejo **76** formado es extremadamente sensible a la humedad. Precisamente, el complejo **76** se destruye por hidrólisis.^[12]

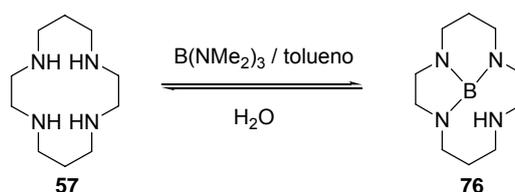


Figura 2.10. Reacción de protección N-1 del ciclamo (**57**) por formación del complejo de boro **76** y su desprotección por hidrólisis.

2.3.2. Formación de carbamatos.

El grupo protector Boc se ha utilizado extensamente en la protección de funcionalidades amino, sin embargo, no fue hasta 1995 que el grupo de Guillard lo utilizó para la protección de los grupos amino presentes en el ciclamo (**57**).^[9] Utilizaron 1.8 moles de dicarbonato de di-*tert*-butilo (Boc_2O) (**65**) por cada mol de ciclamo (**57**) y obtuvieron una mezcla de ciclamos diprottegidos **77-79**, ciclamo triprottegido **66** y ciclamo sin reaccionar **57**, como se muestra en la Figura 2.11, que separaron por cromatografía de columna sobre gel de sílice.

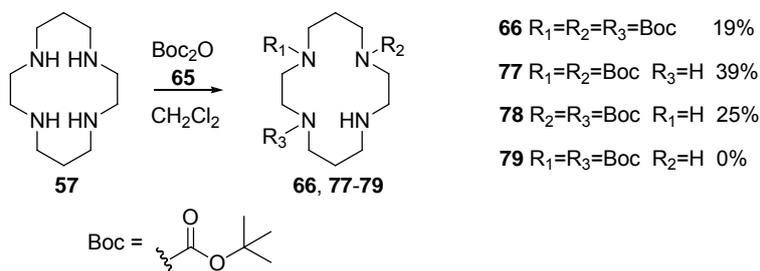


Figura 2.11. Protección de los grupos amino del ciclamo (**57**) con el grupo Boc.

Esta misma metodología fue seguida por Dessolin y colaboradores en la síntesis de nuevos conjugados de mono y biciclamo-AZT con un rendimiento de protección del 24%.^[13]

Fabbrizzi y colaboradores utilizaron esta metodología pero adicionando lentamente el carbamato **65** sobre el ciclamo **57** y les permitió aumentar el rendimiento hasta un 70%.^[14] Por otra parte, la protección con el mismo carbamato **65** pero utilizando como disolvente EtOH en vez de CH₂Cl₂ resultó en el tetraazamacrociclo triprotegido **66** con un rendimiento del 50%.^[15] En 2006 Kitagawa *et al.* usaron la metodología original de Guillard y colaboradores ligeramente modificada para la síntesis del ciclamo triprotegido con Boc **66**, donde aumentaron la proporción de dicarbonato de di-*tert*-butilo (**65**) a 2.5 moles, disueltos en diclorometano, y realizaron una adición lenta sobre una disolución diluida del ciclamo (**57**), también en diclorometano, incluso dejaron que la reacción evolucionase durante la noche. De esta manera consiguieron aumentar el rendimiento hasta un 70%.^[16] El mismo año, otro grupo de investigación realizó un escalado de esta misma reacción (a un orden de magnitud superior) y el rendimiento disminuyó a un 51%.^[17] Cabe destacar que la ventaja principal del grupo Boc es que se desprotege fácilmente en medio ácido y con rendimientos elevados. El Sr. Palomo del Laboratorio de Síntesis del GEM realizó la protección N-1 del ciclamo con Boc utilizando una proporción de ciclamo y Boc₂O de 1 a 2.5, y CH₂Cl₂ como disolvente, tras tres horas de reacción aisló el producto triprotegido **66** con un 33% de rendimiento.^[10]

Otro carbamato, Troc, se ha utilizado en la protección ortogonal de sistemas poliazamacrocíclicos donde los otros grupos protectores eran Boc (no está descrito sobre el ciclamo libre), ya que las condiciones de desprotección de este último carbamato no afectan al primero; el Troc se desprotege por reducción con zinc en ácido acético glacial (Figura 2.12).^[18]

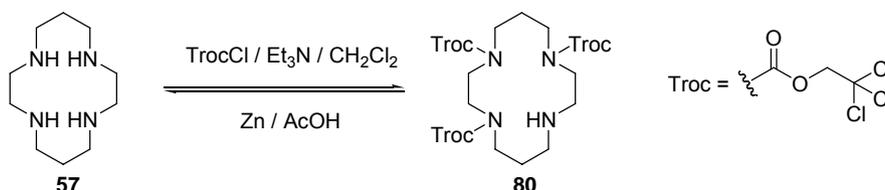


Figura 2.12. Protección N-1 del ciclamo (**57**) con el grupo Troc y su desprotección.

2.3.3. Protección con derivados de fósforo.

En 1998 Guillaume y Marshall realizan la protección N-1 del ciclamo (**57**) con una cantidad equimolar de oxiclouro de fósforo (POCl₃). La ventaja de esta metodología es no tener que utilizar un exceso de ciclamo (**57**). La desprotección de **81** se realiza por hidrólisis ácida (Figura 2.13). Describen el rendimiento global de obtención del AMD3100 (**24**) entre un 62% y un 68%, incluye las etapas de protección del ciclamo (**57**), alquilación con α,α' -dibromo-*p*-xileno (**60**) y desprotección.^[19]

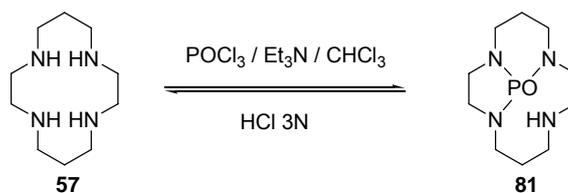


Figura 2.13. Protección N-1 del ciclamo (57) por formación de la triamida fosfórica 81 y su desprotección.

Otros poliazamacrociclos (no se ha descrito sobre el ciclamo libre) se han protegido con el grupo dietoxifosforilo (Dep) por reacción de un macrociclo con clorofosfonato de dietilo (58). Se considera atractivo el uso de este grupo protector en la síntesis de poliaminas ya que se prepara fácilmente a escala multigramo y los productos obtenidos son solubles en la mayoría de disolventes orgánicos. Además la desprotección se puede realizar en condiciones suaves como HCl/dioxano (Figura 2.14).^[6,20]

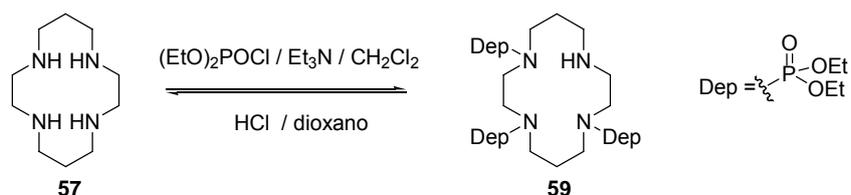


Figura 2.14. Protección N-1 del ciclamo (57) con el grupo dietoxifosforilo y su desprotección.

2.3.4. Protección con amidas.

Las amidas más frecuentemente usadas para la protección de poliazamacrociclos son la trifluoroacetamida y la *p*-toluensulfonamida. Las amidas, en general, son excepcionalmente estables a condiciones ácidas y básicas, y para hidrolizarlas se requiere calefacción en medios fuertemente ácidos o básicos. Sin embargo, los derivados de haloacetilo son más lábiles, por lo que se prefieren como grupos protectores de los nitrógenos del ciclamo, en concreto el grupo trifluoroacetilo, cuya desprotección se realiza en condiciones básicas suaves.

Yang y colaboradores describen la protección N-1 del ciclamo (57) con el grupo trifluoroacetilo con un rendimiento del 92%. Inicialmente llevan a cabo la formación de la amida con cantidades equimolares de trifluoroacetato de etilo (82) y ciclamo (57) en CHCl_3 y llegan a una proporción de 20% de monoprotegido, 30% de diprotegidos, 2% de triprotegido 83 y una pequeña fracción de ciclamo sin reaccionar. Repiten el experimento incrementando 4 veces más los moles de trifluoroacetato de etilo (82) y como disolvente utilizan una mezcla de

$\text{CH}_2\text{Cl}_2/\text{MeOH}$. De esta manera obtienen una proporción de un 1% de monoprotegido, un 60% de diprotegidos y un 32% de triprotegido **83**. Finalmente utilizan MeOH como disolvente para aumentar la velocidad de reacción, 10 moles de trifluoroacetato de etilo (**82**) y trietilamina; obtienen un 92% de ciclamo triprotegido **83** (Figura 2.15).^[21,22] La desprotección la llevan a cabo en medio básico suave, en consecuencia no se puede utilizar este grupo protector en el presente trabajo ya que las reacciones posteriores se realizan en medio básico generado por las aminas utilizadas para generar el fragmento no ciclámico del compuesto final.

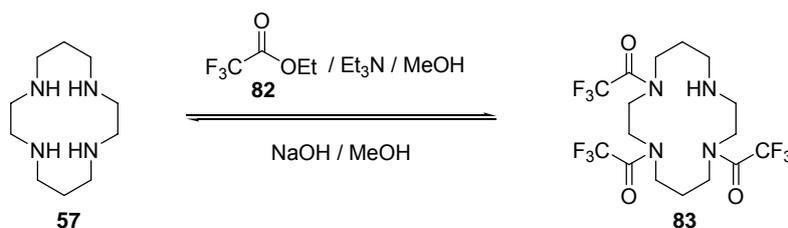


Figura 2.15. Protección N-1 del ciclamo (**57**) con el grupo trifluoroacetilo y su desprotección.

Otra estrategia de protección de los grupos amino es la formación de la *p*-toluensulfonamida correspondiente, en general se utiliza el cloruro de tosilo en un disolvente inerte como CH_2Cl_2 y en presencia de un secuestrador de ácido como trietilamina o piridina. Helps *et al.* sintetizan el ciclamo tritosilado **84** empleando una metodología análoga a la utilizada por Frabrizzi *et al.*^[14] en la síntesis del ciclamo triprotegido con Boc **66**. Realizan la reacción entre el ciclamo (**57**) y el cloruro de tosilo en CH_2Cl_2 y en presencia de trietilamina, con un rendimiento del 51% (Figura 2.16).^[23] El principal inconveniente de este grupo protector es que las condiciones de desprotección no son suaves y pueden afectar a otros grupos funcionales presentes en la molécula.

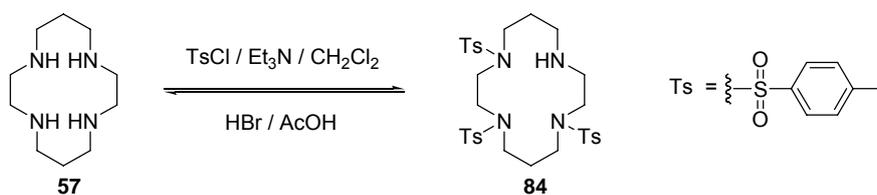


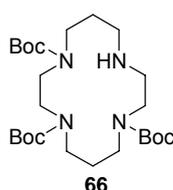
Figura 2.16. Protección N-1 del ciclamo (**57**) con el grupo tosilo y su desprotección.

2.3.5. Síntesis del tri-*tert*-butilcarbamato del ciclamo **66**.

Para la protección de los grupos amino del ciclamo (**57**) en el presente trabajo se decide utilizar el Boc, ya que de todos los protectores descritos en el apartado anterior es el que se ha empleado con mayor frecuencia en la protección de poliazamacrociclos ciclámicos. Posee la característica fundamental de los grupos protectores, es decir, es estable en las condiciones de reacción que se emplearán durante el resto de la vía sintética. Además, tanto la reacción de protección como de desprotección presentan rendimientos elevados y condiciones de reacción suaves.

La protección del ciclamo se lleva a cabo por reacción de 2.3 moles de dicarbonato de di-*tert*-butilo (Boc₂O) (**65**) por cada mol de ciclamo (**57**). El ciclamo (**57**) se disuelve en diclorometano y se adiciona gota a gota una disolución del Boc₂O (**65**) en diclorometano. Se agita a temperatura ambiente durante tiempos diferentes. Se obtiene una mezcla de ciclamos protegidos con Boc y ciclamo sin reaccionar, que se separan por cromatografía de columna. Si se analiza el rendimiento obtenido de **66** se observa cómo aumenta con el tiempo de reacción (Tabla 2.2), y se obtiene el mayor rendimiento, 58%, al dejar evolucionar la reacción durante 6 horas y media.

Tabla 2.2. Reacción de protección con Boc de tres de los grupos amino del ciclamo **57**. Se observa cómo aumenta el rendimiento de **66** al aumentar el tiempo de reacción.



Tiempo de reacción	Rendimiento
3h 45min	29%
5h	41%
5h 30min	46%
5h 45min	53%
6h 30min	58%

El mecanismo de sustitución nucleófila de formación del carbamato **86** ocurre por el ataque nucleófilo de un grupo amino de **62** sobre uno de los grupos carbonilo del Boc₂O (**65**), seguido de la eliminación de CO₂ y *tert*-butanol (Figura 2.17). En el caso del ciclamo triprotegido **66**, el proceso ocurre para tres de los cuatro grupos amino presentes en el ciclamo.^[24]

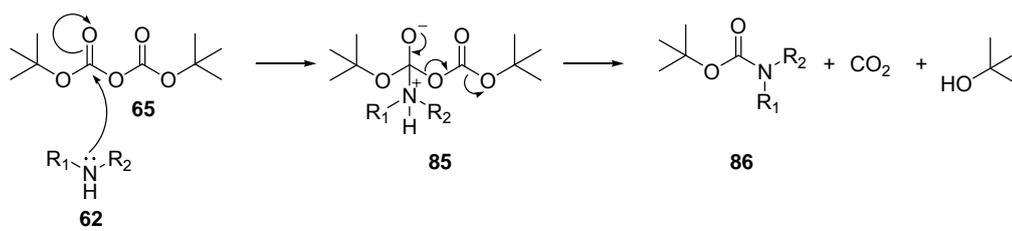


Figura 2.17. Mecanismo de protección de grupos amino con Boc.

2.4. Vía sintética 1.

La primera vía sintética planteada para la obtención de los monociclamos consta de una primera etapa de alquilación del grupo amino libre del ciclamó **66** con α,α' -dibromo-*p*-xileno (**60**) y una segunda etapa de alquilación del grupo amino terminal de una poliamina **31** (fragmento no ciclámico) con el derivado ciclámico bromado **67** (Figura 2.18).

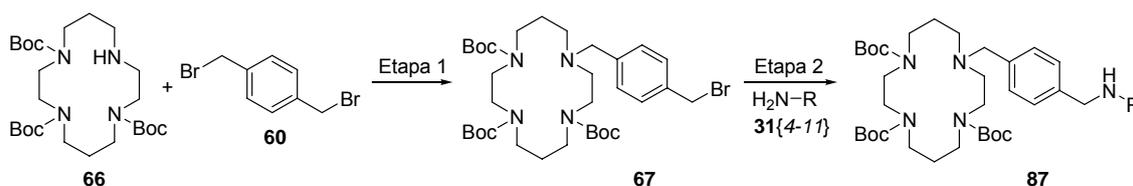


Figura 2.18. Vía sintética 1. La primera etapa es la reacción de alquilación del grupo amino libre del ciclamó triprotegido con Boc **66** con α,α' -dibromo-*p*-xileno (**60**). La segunda etapa es la alquilación del grupo amino terminal de una poliamina **31** (fragmento no ciclámico) con el derivado ciclámico bromado **67**.

2.4.1. Etapa 1: síntesis de 1-[4-(bromometil)fenilmetil]-4,8,11-tris(tert-butoxicarbonil)-1,4,8,11-tetraazaciclotetradecano (**67**).

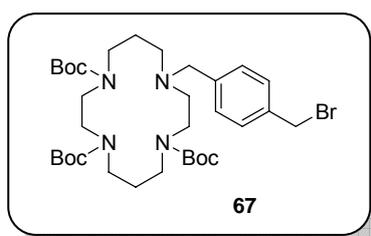


Figura 2.19. Estructura de 1-[4-(bromometil)fenilmetil]-4,8,11-tris(tert-butoxicarbonil)-1,4,8,11-tetraazaciclotetradecano (**67**).

La obtención del derivado ciclámico bromado **67** sigue un mecanismo de sustitución nucleófila S_N2 (Figura 2.20). Se inicia mediante el ataque nucleófilo del grupo amino libre del ciclamó triprotegido con Boc **66** sobre el C electrófilo del derivado halogenado α,α' -dibromo-*p*-xileno (**60**), se desplaza el bromuro y se forma el nuevo enlace C-N. A continuación la base :B desprotona el ión amonio **88** y rinde el producto de alquilación **67**.^[24]

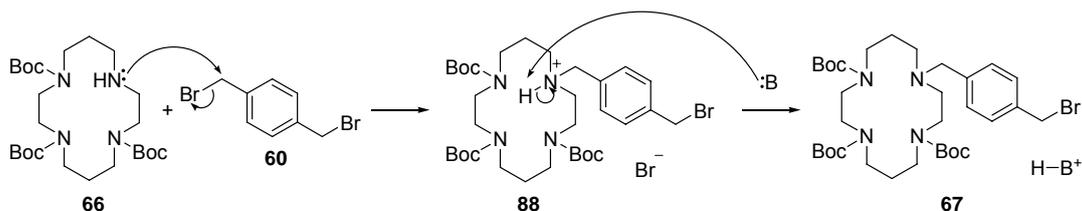


Figura 2.20. Mecanismo de alquilación del ciclamo triprotegido con Boc **66** con α,α' -dibromo-*p*-xileno (**60**).

El procedimiento utilizado por Fabrizzi^[25] y Yen^[26] consiste en realizar la reacción con cantidades equimolares de ciclamo triprotegido con Boc **66** y α,α' -dibromo-*p*-xileno (**60**), pero obtienen el derivado bromado **67** con 45% y 70% de rendimiento respectivamente. Experimentalmente se utiliza la metodología descrita por Luengo y colaboradores para la obtención del derivado ciclámico bromado **67**, donde se describe la reacción del ciclamo triprotegido con Boc **66** con un exceso de α,α' -dibromo-*p*-xileno (**60**), en proporción 1:11, en acetonitrilo y en presencia K_2CO_3 . Esta metodología utiliza un exceso de α,α' -dibromo-*p*-xileno (**60**) para favorecer el ataque nucleófilo del grupo amino del ciclamo **66** solamente sobre uno de los extremos del derivado bromado **60** y evitar la formación de un derivado biciclámico. Describen la obtención de este intermedio bromado **67** con un 90% de rendimiento,^[8] sin embargo al reproducir esta metodología se obtiene el producto de interés **67** con un 70% de rendimiento.

2.4.2. Etapa 2: síntesis de monociclamos sustituidos **87** a partir del derivado ciclámico bromado **76**.

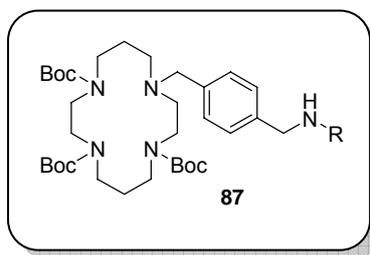


Figura 2.21. Estructura general de los monociclamos sustituidos con poliaminas **87**.

La reacción de obtención de los monociclamos sustituidos **87** es análoga a la etapa 1, también ocurre vía un mecanismo S_N2 . En primer lugar se produce un ataque nucleófilo de la amina primaria **31** sobre el derivado ciclámico bromado **67** (Figura 2.22), que desplaza el bromuro y rinde la amina secundaria objetivo **87**, es decir, el monociclamo sustituido. Sin

embargo, la amina secundaria **87** obtenida es más nucleófila que la amina primaria **31** debido al efecto inductivo +I de los sustituyentes, y por tanto la amina secundaria **87** reacciona preferentemente con el derivado bromado ciclámico **67** presente en el medio dando lugar a una segunda sustitución nucleófila que rinde el término de dialquilación **89** de la amina primaria.

La dificultad de separación de los productos mono **87** y dialquilados **89**, junto con la formación preferente del término de dialquilación **89**, desaconsejan esta vía de obtención de los monociclamos sustituidos con poliaminas **87** y por tanto se exploran otras vías alternativas.

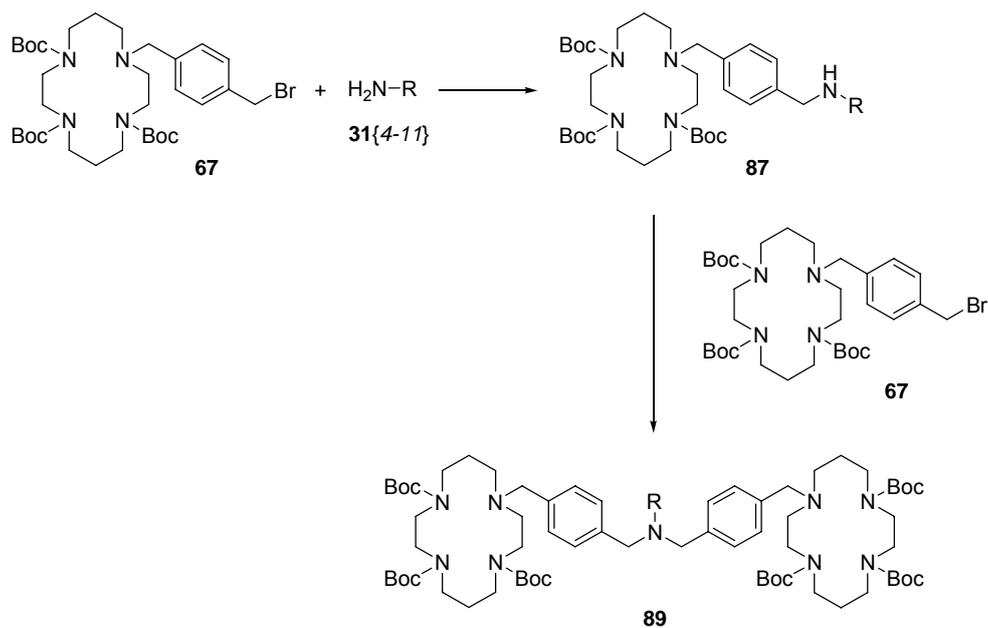


Figura 2.22. La reacción de alquilación de la amina primaria **31** con el derivado ciclámico bromado **67** rinde el monociclamo sustituido **87** objetivo pero la amina secundaria de este compuesto reacciona con el derivado bromado **67** presente en el medio y rinde el término de dialquilación **89**.

2.5. Vía sintética 2.

La segunda vía sintética planteada para la obtención de los monociclamos **87** consta de una primera etapa de alquilación del grupo amino libre del ciclamo triprotegido con Boc **66** con 4-(bromometil)benzaldehído (**74**) y una segunda etapa de aminación reductora del derivado aldehídico del ciclamo **90** con una poliamina **31** (Figura 2.23).

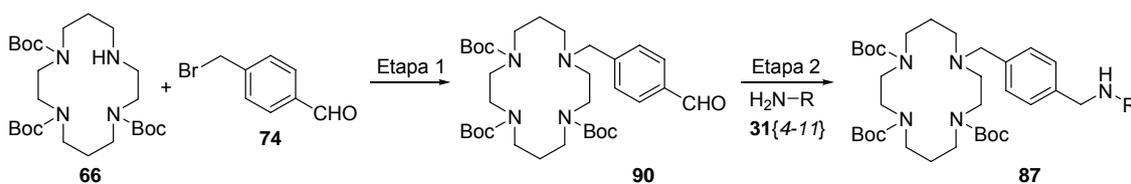


Figura 2.23. Vía sintética 2. La primera etapa es la reacción de alquilación del grupo amino libre del ciclamo triprotegido con Boc **66** con 4-(bromometil)benzaldehído (**74**). La segunda etapa es la aminación reductora del derivado aldehídico del ciclamo **90** con una poliamina **31**.

2.5.1. Síntesis de 4-(bromometil)benzaldehído (**74**).

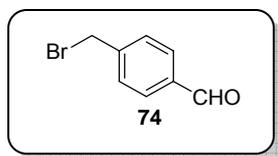


Figura 2.24. Estructura de 4-(bromometil)benzaldehído (**74**).

Se puede obtener el compuesto objetivo **74** por reducción de 4-(bromometil)benzonitrilo (**91**) con hidruro de diisobutilaluminio (DIBAH) con rendimientos descritos del 68%^[27] y del 98%^[28] (Figura 2.25).



Figura 2.25. Reacción de reducción de 4-(bromometil)benzonitrilo (**91**) con DIBAH.

Igualmente se describe su obtención por oxidación de α,α' -dibromo-*p*-xileno (**60**) con óxido de manganeso con un rendimiento del 60% (Figura 2.26).^[29]

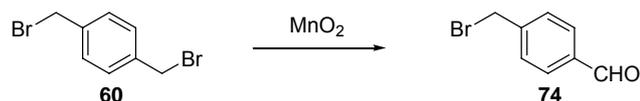


Figura 2.26. Reacción de oxidación de α, α' -dibromo-*p*-xileno (**60**) con MnO_2 .

La síntesis de 4-(bromometil)benzaldehído (**74**) también está descrita^[30] por bromación de 4-(hidroximetil)benzaldehído (**92**) con *N*-bromosuccinimida (NBS) en presencia de trifenilfosfina (PPh_3), con un rendimiento del 73% (Figura 2.27).

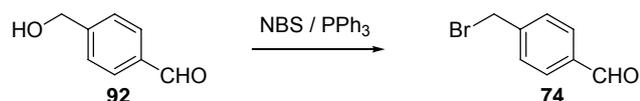


Figura 2.27. Reacción de bromación de 4-(hidroximetil)benzaldehído (**92**) con NBS y PPh_3 .

Se utiliza esta última metodología de obtención del 4-(bromometil)benzaldehído (**74**) debido a la disponibilidad inmediata de los reactivos, la variación de rendimientos descritos para la reducción de 4-(bromometil)benzocitrilo (**91**) con DIBALH, y la presencia de dos grupos de reactividad idéntica en el α, α' -dibromo-*p*-xileno (**60**) para la oxidación con MnO_2 .

El primer paso es la obtención del 4-(hidroximetil)benzaldehído (**92**) mediante la reducción del aldehído libre del monoacetal dietílico del tereftalaldehído (**34**) con $NaBH_4$, seguido de la desprotección del acetal **93** en medio ácido, experimentalmente se obtiene este producto con un rendimiento global del 70%. Finalmente, se lleva a cabo la bromación con 2 moles de NBS y PPh_3 por cada mol de 4-(hidroximetil)benzaldehído (**92**), con un rendimiento del 73% (Figura 2.28). El rendimiento global obtenido al realizar la síntesis de **74** a partir de **34** en el Laboratorio de Síntesis del GEM es del 51%.

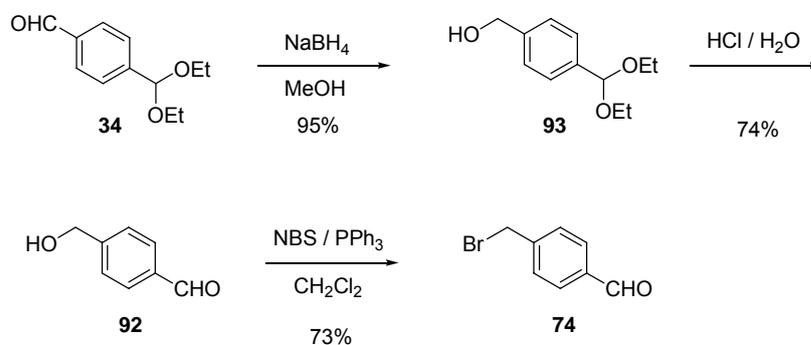


Figura 2.28. Obtención de 4-(bromometil)benzaldehído (**74**) a partir del monoacetal dietílico del tereftalaldehído (**34**), con los rendimientos de cada una de las reacciones. El rendimiento global de obtención de **74** es del 51%.

La bromación descrita por Kawabata y colaboradores rinde un 73% del producto objetivo **74** después de 6 h de reflujo. Al reproducir la metodología descrita, inicialmente se deja que la reacción evolucione durante 6 h, sin embargo, el rendimiento obtenido es del 57% (en comparación con el 73% descrito) y además se obtiene un subproducto que se identifica como α,α,α' -tribromo-*p*-xileno (**94**). Se estudia la evolución de la reacción a diferentes tiempos por cromatografía de capa fina y se observa que en primer lugar se obtiene el producto de interés, 4-(bromometil)benzaldehído (**74**), y posteriormente se forma el subproducto tribromado **94** (sintetizado por S. D. Saraf por bromación de *p*-tolualdehído con bromuro de tionilo^[31,32]) (Figura 2.29). Además, Spaggiari *et al.* describen la formación de dihaluros geminales por reacción de un aldehído con reactivos basados en sistemas (PhO)₃P-halógeno.^[33] Esta reacción explicaría la formación del subproducto α,α,α' -tribromo-*p*-xileno (**94**) a partir del 4-(bromometil)benzaldehído (**74**) obtenido y el exceso de PPh₃ y NBS presentes en el medio de reacción.

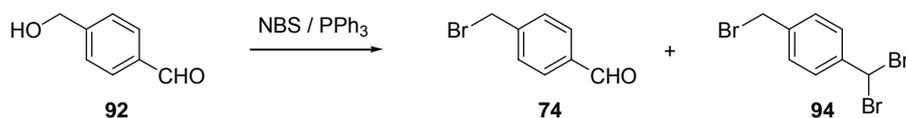


Figura 2.29. La bromación del 4-(hidroximetil)benzaldehído (**92**) rinde 4-(bromometil)benzaldehído (**74**) como producto mayoritario y α,α,α' -tribromo-*p*-xileno (**94**) como producto minoritario.

Se aísla el 4-(bromometil)benzaldehído (**74**) del crudo de reacción por cromatografía de columna y se calcula el rendimiento obtenido según el tiempo de reacción (Tabla 2.3). Se observa que al cabo de 2 h de reacción se obtiene un 73% del 4-(bromometil)benzaldehído (**74**), que coincide con el rendimiento descrito por Kawabata y colaboradores al cabo de 6 h de reacción, mientras que el rendimiento obtenido al dejar la evolucionar la reacción durante 6 h es del 57%, en comparación con el 73% descrito por Kawabata *et al.* Se observa (Tabla 2.3) que el rendimiento de 4-(bromometil)benzaldehído (**74**) disminuye al aumentar el tiempo de reacción debido a la formación de α,α,α' -tribromo-*p*-xileno (**94**).

Tabla 2.3. Rendimiento de obtención de **65** por bromación del 4-(hidroximetil)benzaldehído (**92**) con NBS y PPh₃ a diferentes tiempos de reacción.

	Tiempo de reacción	Rendimiento
<p>74</p>	6h	57%
	4h 30min	61%
	2h	73%

La reacción de sustitución del grupo hidroxilo por un átomo de bromo ocurre por un mecanismo de sustitución nucleófila S_N2 , donde en primer lugar reacciona la trifenilfosfina (**95**) con la NBS (**96**) para formar una sal de fosonio **97**, que reacciona con el alcohol **98** y rinde otro intermedio tipo sal de fosonio **99**. Este intermedio **99** finalmente elimina óxido de trifenilfosfina y rinde el derivado bromado objetivo **100** (Figura 2.30).^[34-36]

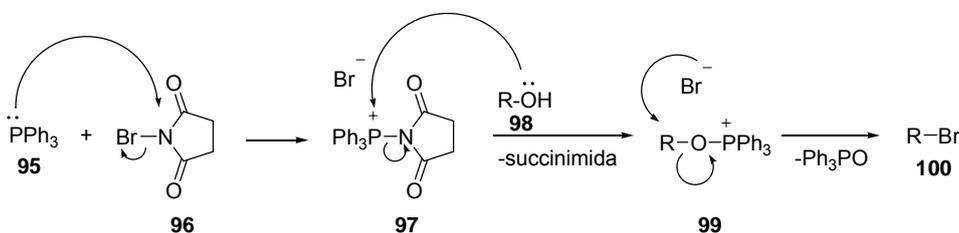


Figura 2.30. Mecanismo de sustitución nucleófila S_N2 de un hidroxilo por un bromo con NBS y PPh_3 .

2.5.2. Etapa 1: síntesis de 1-[4-(carboxaldehído)fenilmetil]-4,8,11-tris(*tert*-butoxicarbonil)-1,4,8,11-tetraazaciclotetradecano (**90**).

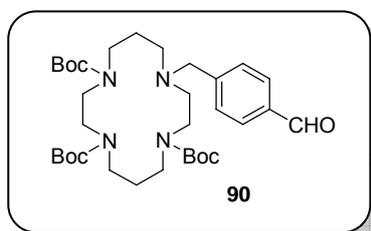


Figura 2.31. Estructura de 1-[4-(carboxaldehído)fenilmetil]-4,8,11-tris(*tert*-butoxicarbonil)-1,4,8,11-tetraazaciclotetradecano (**90**).

La principal novedad de esta vía sintética para la obtención de monociclamos sustituidos con poliaminas es la síntesis del derivado aldehídico del ciclamo **90**. En la literatura, la síntesis de este tipo de monociclamos se realiza a partir del derivado bromado del ciclamo **67**, pero como se ha demostrado en la descripción de la vía sintética 1, ocurren reacciones secundarias de polialquilación cuando se utiliza como nucleófilo una amina primaria.

La obtención del derivado ciclámico **90** a partir del ciclamo triprotegido **66** y el aldehído **74** sigue un mecanismo de sustitución nucleófila S_N2 (análogo al descrito en la síntesis de **67**, ver Figura 2.19). Se inicia mediante el ataque nucleófilo del grupo amino libre del ciclamo triprotegido con Boc **66** sobre el C electrófilo del 4-(bromometil)benzaldehído (**74**), se desplaza

el bromuro y se forma el nuevo enlace C-N. Finalmente, la base desprotona el ión amonio **101** y rinde el producto de alquilación objetivo **90** (Figura 2.32).

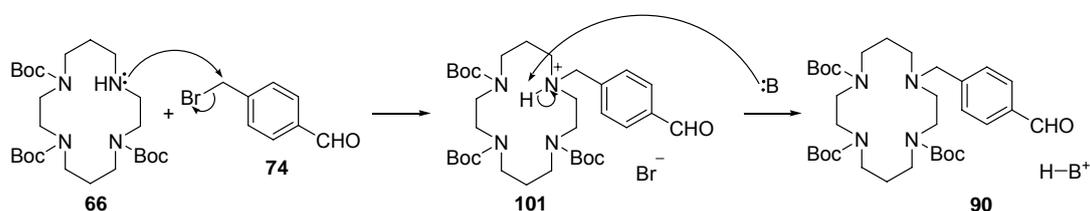


Figura 2.32. Mecanismo de reacción del ciclamo triprotegido con Boc **66** y 4-(bromometil)benzaldehído (**74**).

Experimentalmente se realiza la reacción de obtención del derivado ciclámico **90** por reacción de cantidades equimolares del ciclamo triprotegido con Boc **66** y 4-(bromometil)benzaldehído (**74**) en acetonitrilo y en presencia K_2CO_3 . Se obtiene el compuesto objetivo **90** con un 89% de rendimiento.

2.5.3. Etapa 2: síntesis de monociclamos sustituidos con poliaminas **87** a partir del derivado ciclámico **90**.

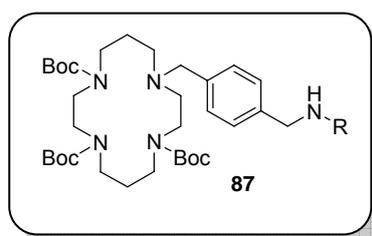


Figura 2.33. Estructura de los derivados monociclámicos sustituidos con poliaminas **87**.

En la literatura se describen aminaciones reductoras de ciclamos con aldehídos utilizando un alcohol como disolvente y a temperatura ambiente^[37] o 65 °C.^[38] Por analogía a las aminaciones reductoras realizadas en el Capítulo 1, se utilizan cantidades equimolares del derivado **90** y de la poliamina **31** en metanol, seguido de la reducción de la imina intermedia **102** con NaBH_4 (Figura 2.34).

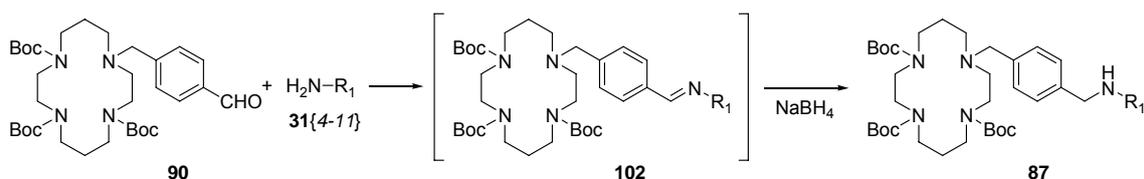


Figura 2.34. Reacción de obtención de los monociclamos **87** a partir del derivado ciclámico **90** y la amina **31** correspondiente.

Se han utilizado métodos computacionales de *docking* sobre el correceptor CXCR4, modelos QSAR y farmacofóricos calculados a partir de los compuestos activos previamente identificados (Capítulos 3 y 4),^[39] para seleccionar los monociclamos **70** posiblemente más activos de la quimioteca diseñada en el apartado 2.2. El resultado de esta selección identifica a los monociclamos **70**{8} y **70**{11} como los candidatos a sintetizar. Por este motivo, la optimización de la reacción de aminación reductora del monociclamo **90** con aminas **31** se realiza concretamente con las aminas **31**{8} y **31**{11}.

El primer experimento de aminación reductora del derivado ciclámico **90** con aminas **31** se realiza con la amina **31**{11} (Figura 2.35) utilizando las mismas condiciones de reacción que en las aminaciones reductoras descritas en el Capítulo 1. Se observa que la reacción no evoluciona, ni siquiera al cabo de dos días en agitación en presencia de un agente deshidratante y a reflujo. Después de aislar el producto **87**{11} del crudo de reacción y de purificarlo por cromatografía de columna, se cuantifica el rendimiento en sólo un 32%. Se repite el experimento, pero en este caso se utiliza una calefacción por microondas a 100 °C durante 2 h, el resultado es el mismo, se obtiene el producto de interés **87**{11} con un 32% de rendimiento.

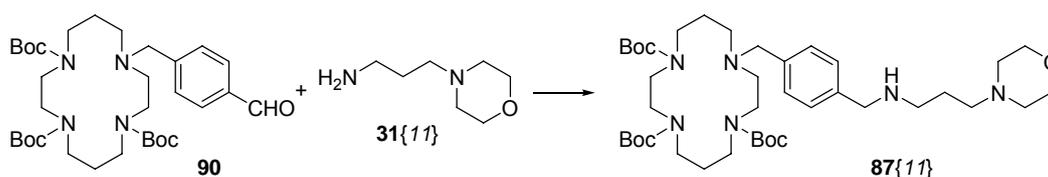


Figura 2.35. Reacción de obtención del monociclamo **87**{11} a partir del ciclamo **90** y la amina **31**{11}.

También se realizan ensayos con la amina **31**{8} (Figura 2.36) utilizando esta metodología, pero en este caso el resultado es incluso peor, no se observa la formación del producto objetivo **87**{8}.

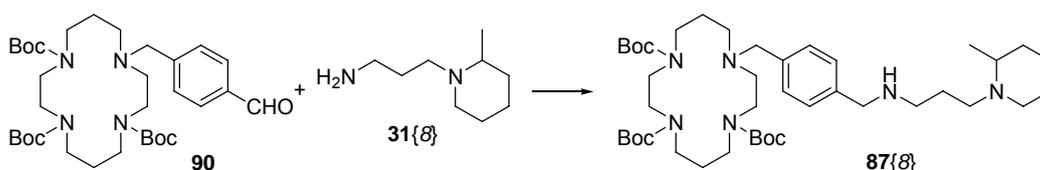


Figura 2.36. Reacción de obtención del monociclamo **87{8}** a partir del ciclamo **90** y la amina **31{8}**.

En la literatura se describe el uso de un ácido de Lewis suave, el isopropóxido de titanio (IV) ($\text{Ti}(\text{O}i\text{Pr})_4$), para favorecer la formación del intermedio imínico en las aminaciones reductoras de aminas y aldehídos o cetonas con NaBH_4 o NaBH_3CN .^[40-44] La principal ventaja de este ácido de Lewis es que se puede utilizar en presencia de grupos funcionales incompatibles con medio ácido, como el Boc. Además, el $\text{Ti}(\text{O}i\text{Pr})_4$ actúa también como agente deshidratante, ya que en presencia de agua forma óxido de titanio e isopropanol.^[45] El mecanismo de la reacción descrito por Mattson *et al.*^[43,46] sugiere la formación de un intermedio de titanio **103**, que se reduce directamente o vía especies imínicas transitorias (Figura 2.37).

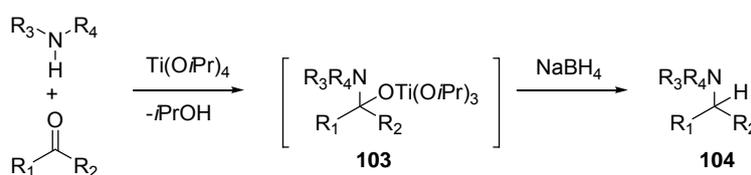


Figura 2.37. Mecanismo de aminación reductora de aminas y aldehídos/cetonas con $\text{Ti}(\text{O}i\text{Pr})_4$.

Esta metodología se ensaya mediante la reacción de cantidades equimolares del derivado ciclámico **90** y la amina **31{8}** con un exceso de $\text{Ti}(\text{O}i\text{Pr})_4$ (dos equivalentes) y a temperatura ambiente. Se obtiene el monociclamo protegido con Boc **87{8}** con un 76% de rendimiento.

En la Tabla 2.4 se resumen los resultados obtenidos en la aminación reductora del ciclamo **90** con las aminas **31{8}** y **31{11}**. Se observa claramente que el uso del catalizador de $\text{Ti}(\text{O}i\text{Pr})_4$ mejora cuantitativamente el rendimiento de esta reacción.

Tabla 2.4. Rendimientos obtenidos en la aminación reductora del ciclamo **90** con las aminas **31{8}** ($X=\text{CH}_2$, $R=\text{Me}$) y **31{11}** ($X=\text{O}$, $R=\text{H}$) con las diferentes metodologías descritas en este apartado.

Compuesto	Metodología	Rendimiento
87{8} $X=\text{CH}_2$, $R=\text{Me}$	MW, 100 °C	–
87{8} $X=\text{CH}_2$, $R=\text{Me}$	$\text{Ti}(\text{O}i\text{Pr})_4$	76%
87{11} $X=\text{O}$, $R=\text{Me}$	reflujo 65 °C	32%
87{11} $X=\text{O}$, $R=\text{Me}$	MW, 100 °C	32%

2.5.4. Desprotección de los grupos Boc de los monociclamos sustituidos 87.

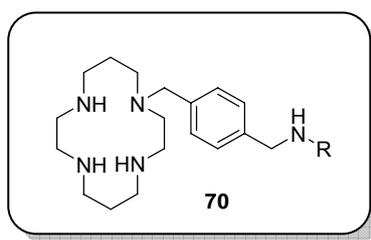


Figura 2.38. Estructura de los monociclamos objetivo **70**.

La desprotección del grupo Boc se realiza generalmente en medio ácido: ácido trifluoroacético, ácido sulfúrico, ácido *p*-toluensulfónico, ácido metanosulfónico y HCl.^[47] Pero también se ha descrito la desprotección térmica, es decir, por calefacción a 150 °C.^[48] Incluso se han llevado a cabo desprotecciones en medio básico,^[49,50] o en condiciones suaves utilizando fluoruro de *N*-tetrabutilamonio.^[51,52] En la síntesis de derivados ciclámicos descritos en la literatura, la desprotección del grupo Boc se realiza preferentemente con HCl/dioxano, HCl/EtOEt con rendimientos que oscilan entre el 34% y el 97%.^[7,8,13]

Experimentalmente se lleva a cabo la desprotección de los monociclamos **87{8}** y **87{11}** con HCl/EtOEt a temperatura ambiente durante 12 h y se obtienen los clorhidratos de los monociclamos desprotegidos **70{8}** y **70{11}** con rendimientos del 90% y 36% respectivamente (Figura 2.39).

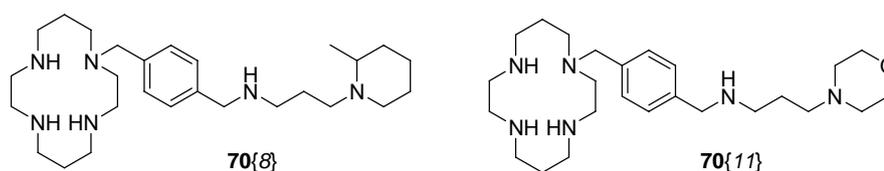


Figura 2.39. Estructura de los monociclamos sustituidos con poliaminas **70{8}** y **70{11}**.

2.6. Resultados y discusión.

Se han utilizado métodos computacionales de *docking* sobre el correceptor CXCR4, modelos QSAR y farmacofóricos calculados a partir de los compuestos activos previamente identificados (Capítulos 3 y 4),^[39] para seleccionar los monociclamos **70** posiblemente más activos de la quimioteca diseñada en el apartado 2.2. El resultado de esta selección identifica a los monociclamos **70{8}** y **70{11}** como los mejores candidatos como inhibidores del correceptor CXCR4. La obtención de estos compuestos se realiza utilizando la vía sintética 2, desarrollada en el apartado anterior, que se basa en la obtención de un intermedio común **90**, que no se halla descrito en la literatura, sobre el que se realizan aminaciones reductoras con las aminas **31** correspondientes y se obtienen, después de la desprotección de los grupos Boc de **87**, los compuestos objetivo **70{8}** y **70{11}** (Figura 2.40).

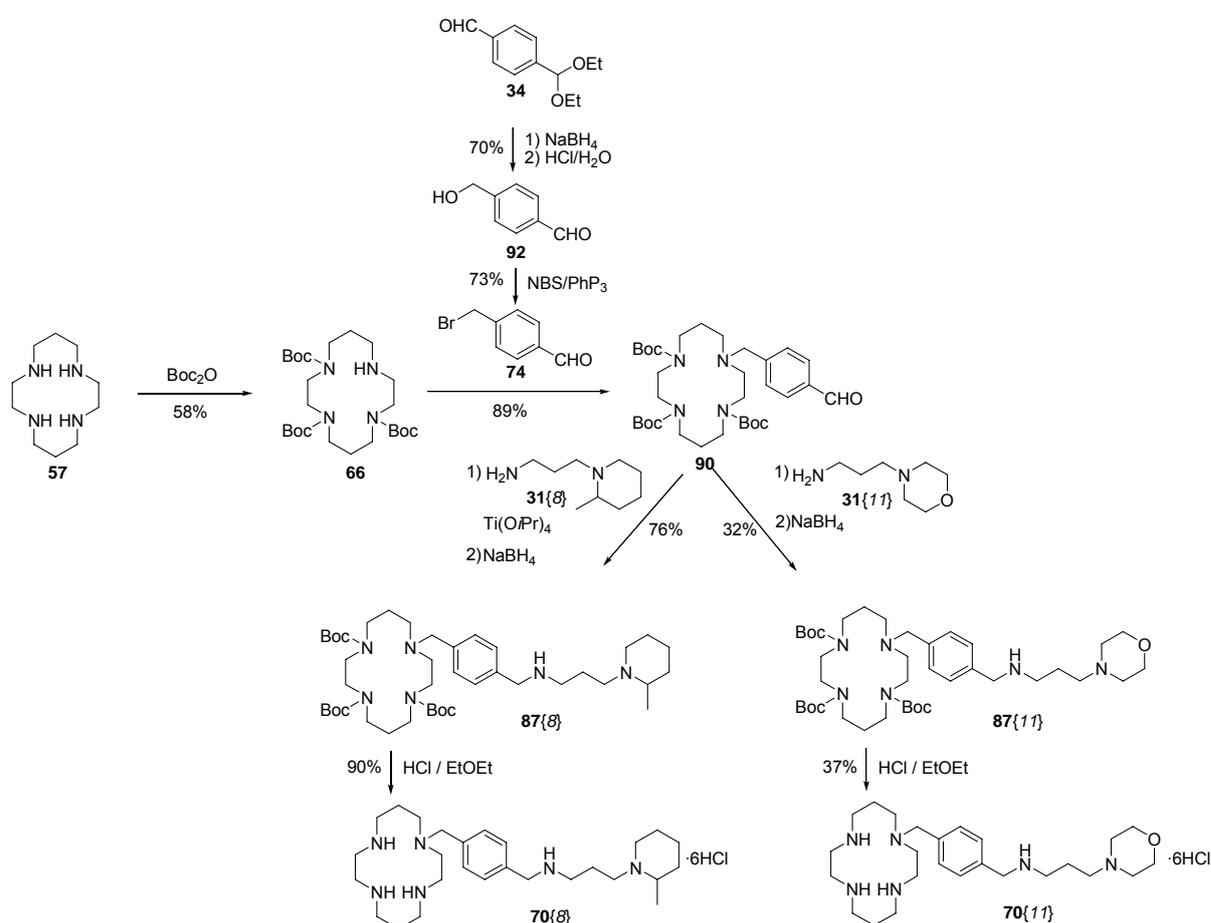


Figura 2.40. Esquema de síntesis de los monociclamos sustituidos con poliaminas **70{8}** y **70{11}**.

La obtención del intermedio **90** a partir del ciclamo (**57**) se realiza en dos etapas, una de protección de los tres de los grupos amino del ciclamo (**57**) con un rendimiento del 58% y una segunda etapa de alquilación del grupo amino libre de **66** con el bromoderivado **74** con un 89% de rendimiento, de manera que el rendimiento global de obtención de este nuevo compuesto **90** es del 52%. A partir de este compuesto **90** se introduce la diversidad de la quimioteca diseñada por reacción de aminación reductora con diferentes aminas **31**.

Se compara (Figura 2.41) el resultado obtenido en la síntesis de este intermedio **90** con el obtenido en la síntesis del intermedio análogo **67** realizado en AnorMED (21% rendimiento a partir del ciclamo (**57**) y el realizado en el presente trabajo (41% de rendimiento a partir del monociclamo **57**). Se determina que el resultado obtenido en la estrategia sintética desarrollada a partir del derivado ciclámico **90** presenta un rendimiento más elevado. Además, la ventaja de este método, frente al utilizado por AnorMED, es que la etapa posterior de reacción con las aminas correspondientes se realiza por aminación reductora a partir de **90**, que presenta rendimientos más elevados que la alquilación a partir del derivado bromado **67** cuando se utilizan aminas primarias, ya que en esta situación se producen polialquilaciones que provocan la disminución drástica del rendimiento de obtención del monociclamo de interés **87** (monoalquilado). Además, el monociclamo **87**, obtenido por alquilación de una amina primaria **31** con el derivado bromado **67**, no se consigue separar de los productos de polialquilación.

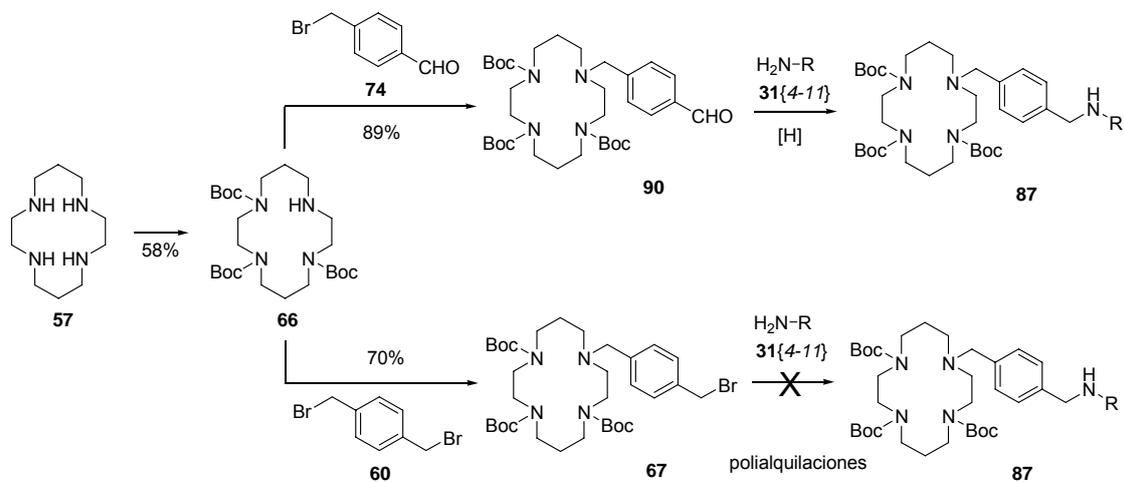


Figura 2.41. Esquema de síntesis de los intermedios monociclámicos **90** y **67** y los monociclamos sustituidos con aminas **87**.

El rendimiento global de la síntesis del monociclamo **70**{8} es del 35%, mientras que el del monociclamo **70**{11} es del 6% (Figura 2.42). Este último valor se puede aumentar si se realiza la etapa de aminación reductora con catálisis de $Ti(OiPr)_4$, como se puede comprobar en los resultados obtenidos en el ensayo realizado con la amina **31**{8}. En la síntesis del derivado ciclámico **87** con esta amina **31**{8} sin el uso del catalizador no se llega a obtener el producto deseado, mientras que el uso del catalizador permite obtener dicho producto con un rendimiento del 76%. Otra etapa que causa una disminución del rendimiento global de

obtención del monociclamo **70**{11} es la desprotección de los grupos Boc de **87**{11}. La diferencia del valor de este rendimiento respecto al 90% obtenido para la **70**{8} se puede explicar por la etapa adicional de recristalización necesaria en el **70**{11}. Asimismo, los resultados descritos en la literatura para la desprotección de monociclamos análogos también presentan variaciones similares del rendimiento.^[7,8,13]

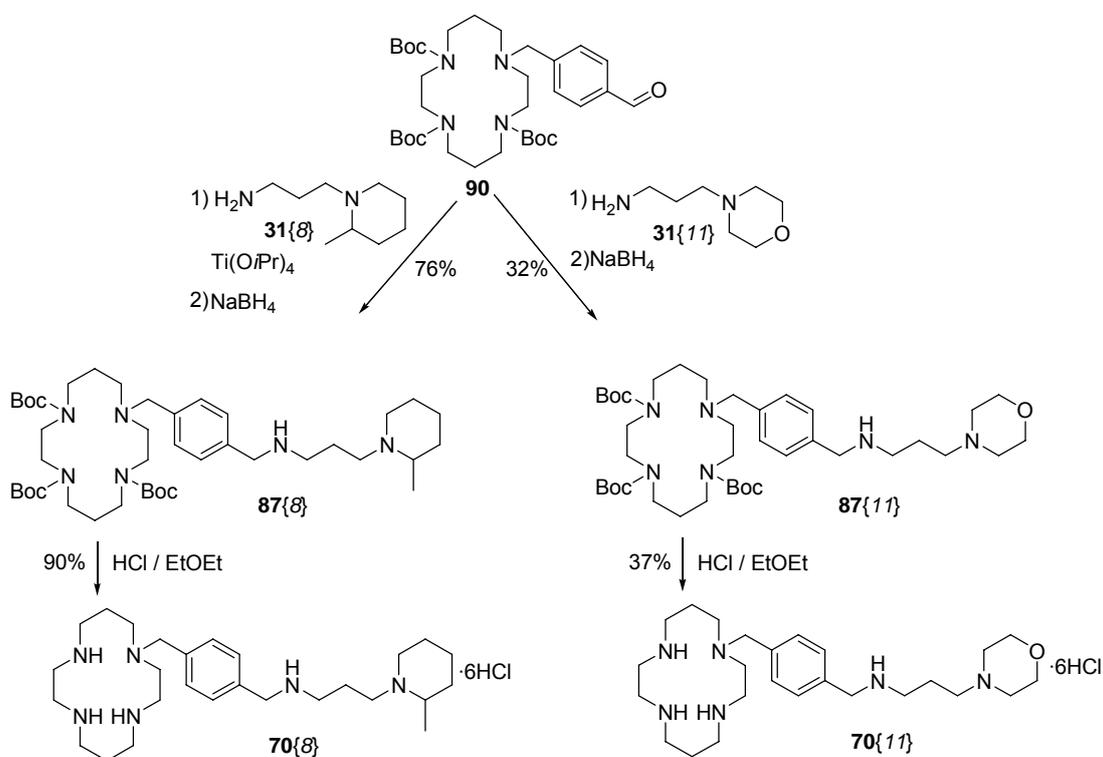
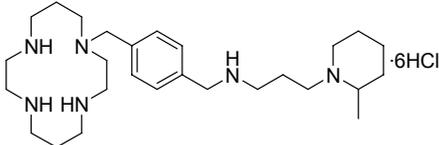
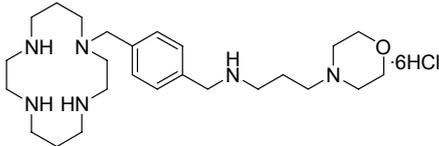
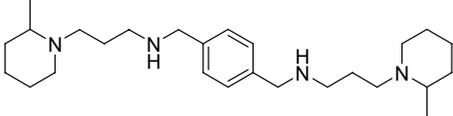
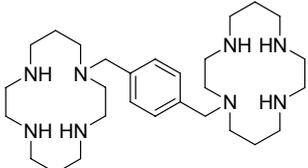


Figura 2.42. Aminación reductora del intermedio ciclámico **90** con las aminas **31**{8} y **31**{11} y posterior desprotección de los grupos Boc para rendir los monociclamos **70**{8} y **70**{11}.

En resumen, se ha desarrollado una metodología que permite obtener monociclamos sustituidos de estructura **70** a partir del intermedio clave aldehído ciclámico **90** y aminas primarias **31**. Este intermedio **90** se obtiene con rendimientos superiores a los descritos para su análogo bromado **67** y además permite utilizar aminas primarias como reactivos en la obtención de los monociclamos **70**.

Se determina la actividad antiviral (EC_{50}) y la citotoxicidad (CC_{50}) de **70{8}** y **70{11}** en el grupo del Dr. Esté del Laboratorio de Retrovirología de la Fundación IrsiCaixa utilizando la metodología detallada en el Capítulo 1 (apartado 1.4). El valor de EC_{50} obtenido para el **70{8}** es de 0.022 $\mu\text{g/mL}$ y para el **70{11}** es 0.058 $\mu\text{g/mL}$; además ambos compuestos presentan un valor de CC_{50} superior a 25 $\mu\text{g/mL}$, es decir, no son tóxicos para las células (Tabla 2.5).

Tabla 2.5. Valores de actividad anti-VIH (EC_{50}) y de citotoxicidad (CC_{50}) expresados en $\mu\text{g/mL}$ de los compuestos sintetizados **70{8}** y **70{11}** y de los compuestos de referencia AMD3100 (inhibidor de la unión del virus a CXCR4) y DS (inhibidor de la unión del virus a la célula) y **32{8,8}** (inhibidor de la unión del virus a CXCR4 diseñado en el Capítulo 1).

Estructura	Nombre	EC_{50} / $\mu\text{g}\cdot\text{mL}^{-1}$	CC_{50} / $\mu\text{g}\cdot\text{mL}^{-1}$
	70{8}	0.022	>25
	70{11}	0.058	>25
	32{8,8}	0.008	>25
	AMD3100	0.001	>5

Cabe destacar que ambos monociclamos, **70{8}** y **70{11}**, presentan mejor actividad antiviral que los monociclamos sintetizados en AnorMED, excepto AMD3645 (Tabla 2.1). Ambos monociclamos, **70{8}** y **70{11}**, presentan valores de actividad en el rango de los ng/mL , sin embargo son menos activos que el biciclamo de referencia AMD3100. Tampoco resultan más activos que la poliamina diseñada en el Capítulo 1, **32{8,8}**. Es un resultado sorprendente en el caso del **70{8}**, ya que si se analiza su estructura se puede comprobar que está formado por el *linker* *p*-fenilbismetilénico presente en el AMD3100 y en **32{8,8}** (ambos simétricos), y sustituido por un ciclamo y por la amina **31{8}**, los fragmentos que están presentes a cada lado del *linker* en los dos compuestos simétricos de referencia. Este resultado podría indicar que

una sustitución no simétrica no es favorable para la unión al correceptor CXCR4, sin embargo, en el Capítulo 1, compuestos de sustitución no simétrica han resultado más activos que sus correspondientes análogos de sustitución simétrica. Un ejemplo sería el compuesto **32**{4,6} cuyo valor de EC₅₀ es de 4.8 µg/mL mientras que sus análogos simétricos **32**{4,4} y **32**{6,6} presentan valores de EC₅₀ de 10.2 µg/mL y >25 µg/mL respectivamente. Por lo tanto, no se puede establecer una relación directa de la estructura con la actividad y se enfatiza la necesidad de disponer de modelos computacionales que permitan predecir la actividad de nuevos monociclamos.

Bibliografía.

- [1] E. de Clercq, N. Yamamoto, R. Pauwels, M. Baba, D. Schols, H. Nakashima, J. Balzarini, Z. Debyser, B. A. Murrer, D. Schwartz, D. Thornton, G. Bridger, S. Fricker, G. Henson, M. Abrams, D. Picker; Potent and selective inhibition of human immunodeficiency virus (HIV)-1 and HIV-2 replication by a class of bicyclams interacting with a viral uncoating event. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 5286-5290.
- [2] K. Princen, S. Hatse, K. Vermeire, S. Aquaro, E. de Clercq, L. O. Gerlach, M. M. Rosenkilde, T. W. Schwartz, R. Skerlj, G. Bridger, D. Schols; Inhibition of human immunodeficiency virus replication by a dual CCR5/CXCR4 antagonist. *J. Virol.* **2004**, *78*, 12996-13006.
- [3] S. Hatse, K. Princen, E. de Clercq, M. M. Rosenkilde, T. W. Schwartz, P. E. Hernandez-Abad, R. T. Skerlj, G. J. Bridger, D. Schols; AMD3465, a monomacrocyclic CXCR4 antagonist and potent HIV entry inhibitor. *Biochem. Pharmacol.* **2005**, *70*, 752-761.
- [4] M. M. Rosenkilde, L. O. Gerlach, S. Hatse, R. Skerlj, D. Schols, G. J. Bridger, T. W. Schwartz; Molecular mechanism of action of monocyclam versus bicyclam non-peptide antagonists in the CXCR4 chemokine receptor. *J. Biol. Chem.* **2007**, *282*, 27354-27365.
- [5] G. J. Bridger, E. M. Boehringer, D. Schols, R. Skerlj, D. E. Bogucki; Antiviral macrocyclic compounds. World Patent WO 00/02870 (filing date: 8-7-1999).
- [6] G. J. Bridger, R. T. Skerlj, D. Thornton, S. Padmanabhan, S. A. Martellucci, G. W. Henson, M. J. Abrams, N. Yamamoto, K. de Vreese, R. Pauwels, E. de Clercq; Synthesis and structure-activity relationships of phenylenebis(methylene)-linked bis-tetraazamacrocycles that inhibit HIV replication. Effects of macrocyclic ring size and substituents on the aromatic linker. *J. Med. Chem.* **1995**, *38*, 366-378.
- [7] S. K. Gupta; Method for inducing chemotaxis in endothelial cells by administering stromal cell derived factor-1alpha. US Patent US 2002/0107195 (filing date: 17-9-2001).
- [8] J. I. Luengo, A. T. Price, A. Shaw, K. Wiggall; CXCR-4 receptor antagonists-thrombopoietin mimetics. World Patent WO 00/66112 (filing date: 3-5-2000).
- [9] B. Boitrel, B. Andrioletti, M. Lachkar, R. Guillard; Characterization and application of a new diprotected cyclam: a novel two-step synthesis of linked tetraazamacrocycles. *Tetrahedron Lett.* **1995**, *36*, 4995-4998.
- [10] A. Palomo. Comunicación personal, **2003**.
- [11] J. J. Yaouanc, N. Le Bris, G. Le Gall, J. C. Clément, H. Handel, H. des Abbayes; Mono N-functionalization of cyclic and linear tetraamines via their tridentate tricarbonylchromium complexes. *J. Chem. Soc. Chem. Commun.* **1991**, 206-207.
- [12] H. Bernard, J. J. Yaouanc, J. C. Clément, H. des Abbayes, H. Handel; General route for the synthesis of mono N-alkylated derivatives of tetraazamacrocycles. *Tetrahedron Lett.* **1991**, *32*, 639-642.
- [13] J. Dessolin, P. Galea, P. Vlieghe, J. C. Chermann, J. L. Kraus; New bicyclam-AZT conjugates: design, synthesis and anti-HIV evaluation, and their interaction with CXCR-4 coreceptor. *J. Med. Chem.* **1999**, *42*, 229-241.

- [14] L. Fabbrizzi, F. Foti, M. Licchelli, P. M. Maccarini, D. Sacchi, M. Zema; Light-emitting molecular machines: pH-induced intramolecular motions in a fluorescent nickel(II) scorpionate complex. *Chem. Eur. J.* **2002**, *8*, 4965-4972.
- [15] F. Riché, A. Moulinet d'Hardemare, S. Sèpe, L. Riou, D. Fagret, M. Vidal; Nitroimidazoles and hypoxia imaging: synthesis of three technetium-99m complexes bearing a nitroimidazole group: biological results. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 71-74.
- [16] T. Kitagawa, A. Dey, P. Lugo-Mas, J. B. Benedict, W. Kaminsky, E. Solomon, J. A. Kovacs; A functional model for the cysteinyl-ligated non-heme iron enzyme superoxide reductase (SOR). *J. Am. Chem. Soc.* **2006**, *128*, 14448-14449.
- [17] C. Schickaneder, F. W. Heinemann, R. Alsfasser; Copper(II) complexes of the tetraazamacrocyclic tertiary amide ligand alanyl-cyclam. *Eur. J. Inorg. Chem.* **2006**, 2357-2363.
- [18] J. D. Chartres, L. F. Lindboy, G. V. Meehan; New heterotopic, linked macrocyclic systems derived from selectively protected macrocycles. *Tetrahedron* **2006**, *62*, 4173-4187.
- [19] D. Guillaume, G. R. Marshall; Efficient one-pot synthesis of JM3100. *Synth. Commun.* **1998**, *28*, 2903-2906.
- [20] M. I. Burguete, L. Lopez-Diago, E. García-España, F. Galindo, S. V. Luis, J. F. Miravet, D. Sroczynski; New efficient procedure for the use of diethoxyphosphoryl as a protecting group in the synthesis of polyazamacrocycles. Preparation of polyazacyclophanes derived from resorcinol. *J. Org. Chem.* **2003**, *68*, 10169-10171.
- [21] C. M. Giandomenico, W. Yang; Process for preparation of N-1 protected N ring nitrogen containing cyclic polyamines and products thereof. World Patent WO 02/26721 A1 (filing date: 28-9-2001).
- [22] W. Yang, C. M. Giandomenico, M. Sartori, D. A. Moore; Facile N-1 protection of cyclam, cyclen and 1,4,7-triazacyclononane. *Tetrahedron Lett.* **2003**, *44*, 2481-2483.
- [23] I. M. Helps, D. Parker, J. R. Morphy, J. Chapman; General routes for the synthesis of mono, di and tri-N-substituted derivatives of cyclam. *Tetrahedron* **1989**, *45*, 219-226.
- [24] M. B. Smith, J. March; Aliphatic nucleophilic substitution. In *March's advanced organic chemistry: reactions, mechanisms, and structure*, 5th edition; Wiley Interscience: New York, **2001**; 389-674.
- [25] L. Fabbrizzi, F. Gatti, P. Pallavicini, E. Zambarbieri; Redox-driven intramolecular anion translocation between transition metal centres. *Chem. Eur. J.* **1999**, *5*, 682-690.
- [26] C. F. Yen, C. H. Wu, Y. H. Huang, C. H. King, S. J. Chen, H. C. Chen, G. H. Hakimelahi; Polyamine compounds. World Patent WO 2008/021788 A2 (filing date: 6-8-2007)
- [27] C. Wessler, A. Homann, U. Fricke, J. Lehmann; NO donors, part 8: synthesis and vasodilating activities of substituted benzylnitrates compared to cyclohexylmethylnitrate and GTN. *Eur. J. Med. Chem.* **2003**, *38*, 581-586.
- [28] X. Z. Wang, T. R. Burke; Enantioselective synthesis of metathesis-derived ureapeptoid macrocycles. *Synlett.* **2004**, 469-472.
- [29] S. Goswami, S. Jana, S. Dey, A. K. Adak; A simple and convenient manganese dioxide oxidation of benzyl halides to aromatic aldehydes under neutral condition. *Chem. Lett.* **2005**, *34*, 194-195.

- [30] E. Kawabata, K. Kikuchi, Y. Urano, H. Kojima, A. Odani, T. Nagano; Design and synthesis of zinc-selective chelators for extracellular applications. *J. Am. Chem. Soc.* **2005**, *127*, 818-819.
- [31] S. D. Saraf; A new synthesis of 4-bromomethylbenzal bromide and 1,4-bis(dibromomethyl)-benzene. *Org. Prep. Proced. Int.* **1981**, *13*, 365-367.
- [32] S. D. Saraf; Reaction of thionyl bromide with aromatic aldehydes. *Synth. Commun.* **1983**, *13*, 7-14.
- [33] A. Spaggiari, D. Vaccari, P. Davoli, G. Torre, F. Prati; A mild synthesis of vinyl halides and gem-dihalides using triphenyl phosphite-halogen-based reagents. *J. Org. Chem.* **2007**, *72*, 2216-2219.
- [34] A. K. Bose, B. Lal; A facile replacement of hydroxyl by halogen with inversion. *Tetrahedron Lett.* **1973**, *40*, 3937-3940.
- [35] S. Hanessian, M. M. Ponpipom, P. Lavallee; Procedures for the direct replacement of primary hydroxyl groups in carbohydrates by halogen. *Carbohydr. Res.* **1972**, *24*, 45-56.
- [36] G. Hodosi, B. Podanyi, J. Kuzmann; The mechanism of the hydroxyl-halogen exchange reaction in the presence of triphenylphosphine, N-bromosuccinimide, and N,N-dimethylformamide: application of a new Vilsmeier-type reagent in carbohydrate chemistry. *Carbohydr. Res.* **1992**, *230*, 327-342.
- [37] R. Tripier, F. Chuburu, M. Baccon, H. Handel; Phenylglyoxal for polyamines modification and cyclam synthesis. *Tetrahedron* **2003**, *59*, 4573-4579.
- [38] H. Luo, R. D. Rogers, M. W. Brechbiel; A convenient and selective route to a trans-difunctionalized macrocyclic hexadentate N_4O_2 ligand. *Can. J. Chem.* **2001**, *79*, 1105-1107.
- [39] V. I. Pérez-Nueno, S. Pettersson, D. W. Ritchie, J. I. Borrell, J. Teixidó; Discovery of novel HIV entry inhibitors for the CXCR4 receptor by prospective virtual screening. *J. Chem. Inf. Model.* **2009**, *49*, 810-823.
- [40] S. Bhattacharyya; Titanium(IV) isopropoxide and sodium borohydride: a reagent of choice for reductive amination. *Tetrahedron Lett.* **1994**, *35*, 2401-2404.
- [41] S. Bhattacharyya; Reductive alkylation of dimethylamine using titanium(IV) isopropoxide and sodium borohydride: an efficient, safe, and convenient method for the synthesis of N,N-dimethylated tertiary amines. *J. Org. Chem.* **1995**, *60*, 4928-4929.
- [42] H. J. Kumpaty, J. S. Williamson, S. Bhattacharyya; Synthesis of N-methyl secondary amines. *Synth. Commun.* **2003**, *33*, 1411-1416.
- [43] R. J. Mattson, K. M. Pham, D. J. Leuck, K. A. Cowen; An improved method for reductive alkylation of amines using titanium(IV) isopropoxide and sodium cyanoborohydride. *J. Org. Chem.* **1990**, *55*, 2552-2554.
- [44] K. A. Neidigh, M. A. Avery, J. S. Williamson, S. Bhattacharyya; Facile preparation of N-methyl secondary amines by titanium(IV) isopropoxide-mediated reductive amination of carbonyl compounds. *J. Chem. Soc. Perkin. Trans. I* **1998**, 2527-2531.
- [45] V. G. Courtecuisse, K. Chhor, J. F. Bocquet, C. Pommier; Kinetics of the titanium isopropoxide decomposition in supercritical isopropyl alcohol. *Ind. Eng. Chem. Res.* **1996**, *35*, 2539-2545.
- [46] H. Takahashi, T. Tsubuki, K. Higashiyama; A new method for the synthesis of α -substituted phenethylamines via titanium amide complexes. *Synthesis* **1988**, 238-239.

- [47] T. W. Greene, P. G. M. Wuts; Protection for the amino group. In *Protective groups in organic synthesis*, 2nd edition; Wiley Interscience: New York, **1991**; 309-405.
- [48] H. H. Wasserman, G. D. Berger, K. R. Cho; Transamination reactions using β -lactams. The synthesis of homaline. *Tetrahedron Lett.* **1982**, *23*, 465-468.
- [49] S. El Kazzouli, J. Koubachi, S. Berteina-Raboin, A. Mouaddib, G. Guillaumet; A mild and selective method for the N-Boc deprotection by sodium carbonate. *Tetrahedron Lett.* **2006**, *47*, 8575-8577.
- [50] N. J. Tom, W. M. Simon, H. N. Frost, M. Ewing; Deprotection of a primary Boc group under basic conditions. *Tetrahedron Lett.* **2004**, *45*, 905-906.
- [51] U. Jacquemard, V. Bénateau, M. Lefoix, S. Routier, J. Y. Mérour, G. Coudert; Mild and selective deprotectino of carbamates with Bu₄NF. *Tetrahedron* **2004**, *60*, 10039-10047.
- [52] S. Routier, L. Saugé, N. Ayerbe, G. Coudert, J. Y. Mérour; A mild and selective method for N-Boc deprotection. *Tetrahedron Lett.* **2002**, *43*, 589-591.

CAPÍTULO 3. QSAR

3.1. QSAR (quantitative structure-activity relationship).

Los modelos QSAR y QSPR (*quantitative structure-property relationship*) cuantifican la correlación de las características derivadas de la estructura de compuestos químicos (descriptores) con su actividad biológica u otra propiedad. En el diseño de quimiotecas, el número de compuestos que se pueden sintetizar en la práctica para poder evaluar su actividad es mucho menor que el número total de compuestos que componen estas quimiotecas virtuales. Este hecho crea una necesidad de desarrollar herramientas que permitan el cribado virtual de las quimiotecas, los modelos QSAR son un ejemplo.

Los primeros trabajos sobre QSAR se remontan al siglo XIX, en el área de la toxicología. En 1863, Crox observa que existe una relación entre la toxicidad de alcoholes primarios alifáticos y su solubilidad en agua. Cinco años más tarde, Crum-Brown y Fraser postulan que la acción fisiológica de una sustancia en un sistema biológico es función de su constitución química. Asimismo, en 1890, Meyer y Overton, en trabajos independientes, relacionan la toxicidad de compuestos orgánicos con su lipofilia. De esta manera se demuestra que existe una relación entre la estructura de los compuestos químicos, las propiedades de sus sustituyentes y toxicidad.^[1] No es hasta bien entrado el siglo XX que Hammett correlaciona las propiedades electrónicas de ácidos y bases orgánicos con sus constantes de equilibrio y reactividad.^[2] En 1952, Taft añade a la ecuación de Hammett los efectos estéricos de los sustituyentes.^[3] El primer modelo QSAR moderno se basa en los trabajos de Hammett y Taft, y fue desarrollado por Hansch y colaboradores en 1962.^[4] Este trabajo describe la relación entre la estructura y la actividad de sustancias reguladoras del crecimiento de plantas con las constantes de Hammett y la hidrofobicidad mediante la Ecuación 3.1, donde C es la concentración tóxica, σ es la constante de Hammett que indica la fuerza relativa de dador o aceptor electrónico de los sustituyentes y π indica la hidrofobicidad relativa de un sustituyente (Ecuación 3.2, donde P_X y P_H son los coeficientes de partición de un derivado y del compuesto de referencia). Hansch mejora el modelo desarrollado en trabajos posteriores donde además incluye nuevas definiciones para las constantes hidrofóbicas. Así se pasa de un modelo lineal a una ecuación parabólica, Ecuación 3.3.^[5,6]

$$\log \frac{1}{C} = a\sigma + b\pi + c \quad [3.1]$$

$$\pi = \log P_X - \log P_H \quad [3.2]$$

$$\log \frac{1}{C} = a \log P - b(\log P)^2 + c\sigma + d \quad [3.3]$$

Estos modelos consideran una plantilla común o referencia, que generalmente es un anillo aromático, con una variación estructural limitada, como cambios de sustituyentes en determinadas posiciones. El principal problema de este enfoque es la gran cantidad de compuestos que se requieren para explorar todas las combinaciones estructurales. Para evitar este problema combinatorio, paralelamente se desarrollaron enfoques alternativos, entre los

que destaca el de Free y Wilson, que relaciona la actividad con la presencia de un grupo funcional en una posición específica en una serie de análogos mediante la Ecuación 3.4, donde BA es la actividad biológica, a_i es la contribución de cada característica estructural, x_i es la presencia ($x_i=1$) o ausencia ($x_i=0$) de un fragmento estructural concreto, y u es la contribución media del compuesto de referencia.^[7]

$$BA = \sum a_i x_i + u \quad [3.4]$$

El alejamiento de los modelos lineales llevó a un desarrollo exponencial en el campo del análisis QSAR. Entre ellos destaca la aproximación de McFarland que relaciona el transporte de un fármaco con su hidrofobicidad en modelos multicompartimentales mediante la Ecuación 3.5.^[8]

$$\log \frac{1}{C} = a \log P - 2a \log(P + 1) + c \quad [3.5]$$

A partir de estos resultados, Kubinyi desarrolla en 1976 el modelo bilineal que mejora los resultados obtenidos con anterioridad. Este modelo está definido mediante la Ecuación 3.6, donde β es la relación de volúmenes de la fase orgánica y la fase acuosa del sistema biológico estudiado.^[9]

$$\log \frac{1}{C} = a \log P - b \log(\beta P + 1) + c \quad [3.6]$$

Posteriormente se desarrollan los métodos QSAR 3D donde se relacionan la forma y propiedades de las moléculas con características moleculares específicas y su relación espacial. Entre ellos destaca el análisis CoMFA (*comparative molecular field analysis*) introducido por Cramer en 1988,^[10] donde se calculan las energías de interacción estérica (van der Waals) y electrostática (Coulomb) entre cada molécula y un átomo o grupo sonda adecuados situados en puntos regularmente espaciados de una malla tridimensional que rodea a todos los compuestos alineados. El análisis CoMSIA (*comparative molecular similarity indices analysis*), desarrollado por Klebe,^[11] es análogo al anterior pero considera cinco campos moleculares: estérico, electrostático, dador y aceptor de hidrógeno, e hidrofóbico (ejemplo en la Figura 3.1).

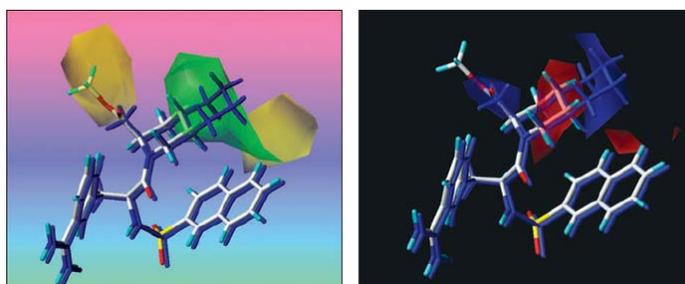


Figura 3.1. Izquierda: modelo CoMSIA donde se indican las interacciones estéricas favorables en verde y desfavorables en amarillo. Derecha: modelo CoMSIA donde se indican las interacciones hidrofóbicas favorables en azul y desfavorables en verde (figura de www.optive.com/data/SYBYL).

El proceso de desarrollo de un modelo QSAR se puede dividir en tres etapas generales: preparación de datos, análisis de datos y validación del modelo. La primera etapa engloba la selección de las bases de datos moleculares para realizar los estudios QSAR, cálculo y selección de descriptores moleculares. La segunda etapa consiste en la aplicación de métodos estadísticos para el desarrollo de los modelos QSAR con los descriptores seleccionados. La mayor parte de ellos están basados en regresión lineal múltiple, regresión por mínimos cuadrados parciales (PLS, *partial least squares*) y métodos no lineales (como algoritmos genéticos, redes neuronales, etc). Todos los métodos utilizan los descriptores como variables independientes y las actividades biológicas como variables dependientes. La última etapa es la validación de los modelos desarrollados ya que un modelo que reproduce los resultados experimentales de la base de datos con que ha sido entrenado (*training set*) no tiene por qué rendir adecuadamente en nuevos compuestos, fenómeno que se conoce con el nombre de *overfitting* o sobreajuste. Por lo tanto esta etapa de validación es tan importante como la etapa de construcción del modelo QSAR.

Además, un modelo QSAR ideal debe cumplir una serie de condiciones. En primer lugar debe considerar un número adecuado de compuestos para ser estadísticamente representativo. Estos compuestos también deben cubrir un rango de actividad de varios órdenes de magnitud para los modelos de regresión, o tener una distribución adecuada de compuestos en cada clase (activo e inactivo) para los modelos de clasificación. Por último, los modelos se deben poder utilizar para realizar predicciones fiables para nuevos compuestos (Figura 3.2).

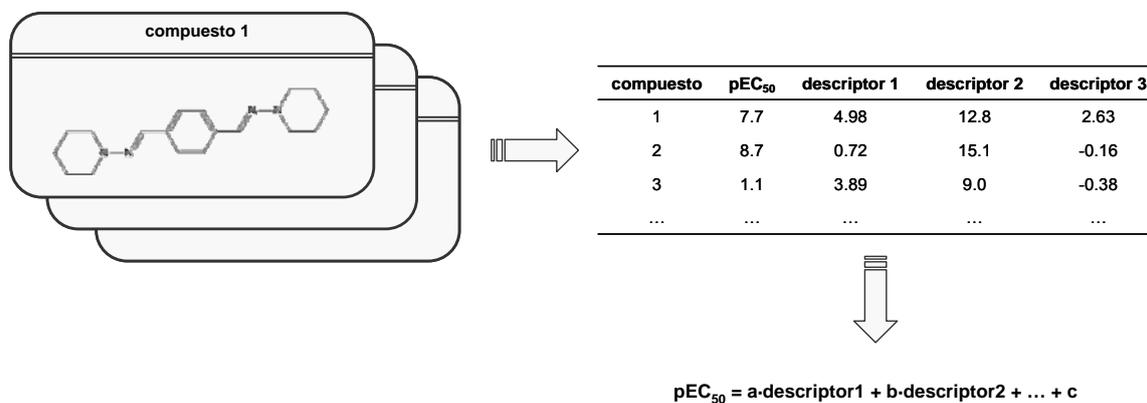


Figura 3.2. Esquema de trabajo para el cálculo de modelos QSAR.

3.2. Parámetros utilizados en el análisis de los modelos QSAR.

Los modelos QSAR se validan mediante dos procedimientos. El primero es la validación cruzada, que es un procedimiento de validación interna mediante el cual porciones del *training set* se dejan fuera de la construcción del modelo y posteriormente se estiman sus valores como si fueran nuevas moléculas. En la validación cruzada *leave one out*, para n compuestos se divide el *training set* en n porciones, se evalúa la actividad de un compuesto creando el modelo con los $n-1$ compuestos restantes, y así sucesivamente para cada una de las porciones. El segundo procedimiento es una validación externa con un *test set*.

El rendimiento de los modelos QSAR obtenidos se valora mediante los siguientes parámetros estadísticos:

- Coeficiente de correlación para el modelo, R^2 , la validación cruzada (LOO), R^2_{LOO} , y la validación con el *test set*, R^2_{test} . Se calculan mediante la Ecuación 3.7 donde x es valor experimental de pEC_{50} (calculado como $-\log EC_{50}$ en unidades molares) e y es el valor de pEC_{50} calculado mediante el modelo.

$$R = \frac{\sum(x - \bar{x}) \cdot (y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \cdot \sum(y - \bar{y})^2}} \quad [3.7]$$

La validación cruzada LOO consiste en eliminar uno de los compuestos del *training set*, generar un modelo a partir del resto de compuestos y realizar la predicción del compuesto eliminado. El valor del coeficiente de correlación de la validación cruzada, R^2_{LOO} , se ha utilizado ampliamente como criterio de robustez y capacidad predictiva de un modelo, con un umbral fijado en 0.5 (y fijado en 0.6 para el modelo). No obstante, un valor elevado de este coeficiente de correlación se considera una condición necesaria para que un modelo tenga una elevada capacidad predictiva, pero no es una condición suficiente. Por lo tanto, los modelos se validan con un *test set* externo para determinar su capacidad predictiva real ($R^2_{test} > 0.5$).^[12,13]

- Error cuadrático medio *RMSE* (*root mean squared error*) para el modelo, $RMSE$, la validación cruzada (LOO), $RMSE_{LOO}$, y la validación con el *test set*, $RMSE_{test}$, definidos por la Ecuación 3.8 donde n es el número de compuestos y SSE se define en la Ecuación 3.13.

$$RMSE = \sqrt{\frac{SSE}{n}} \quad [3.8]$$

- Test de Fisher se utiliza para determinar el grado de bondad con que una variable independiente predice a la variable dependiente, y examinan las medidas de variación. La primera es la suma total de cuadrados (SST), que puede dividirse en dos partes: la variación explicada o suma de cuadrados debida a la regresión (SSR) y la variación no explicada o suma de cuadrados de error (SSE). SST es la

suma de cuadrados de las desviaciones de los valores experimentales de su media y los grados de libertad gdl_{SST} se calculan a partir del número de compuestos n utilizados para crear el modelo (*training set*) (Ecuaciones 3.9 y 3.10).

$$SST = \sum (x - \bar{x})^2 \quad [3.9]$$

$$gdl_{SST} = n - 1 \quad [3.10]$$

SSR representa la diferencia entre el valor promedio de y y el valor promedio de y que sería predicho a partir de la relación de regresión, y los grados de libertad gdl_{SSR} se calculan a partir del número de descriptores d que utiliza el modelo (Ecuaciones 3.11 y 3.12).

$$SSR = \sum (y - \bar{y})^2 \quad [3.11]$$

$$gdl_{SSR} = d \quad [3.12]$$

SSE representa aquella parte de la variación que no es explicada por la regresión, y los grados de libertad gdl_{SSE} se calculan a partir del número de compuestos n utilizados para crear el modelo, y del número de d descriptores que utiliza el modelo, (Ecuaciones 3.13 y 3.14).

$$SSE = \sum (x - y)^2 \quad [3.13]$$

$$gdl_{SSE} = n - d - 1 \quad [3.14]$$

El valor de F del test de Fisher se utiliza para probar hipótesis, de manera que la hipótesis nula H_0 asume que el modelo no predice mejor que el valor medio de la actividad y la hipótesis H_1 asume lo contrario, generalmente se acepta o rechaza H_1 con un cierto nivel de significancia α , que significa que la probabilidad de H_1 de ser cierto o falso es como mínimo α . Para decidir si se acepta H_1 a cierto nivel de significancia, se utiliza la distribución- F . El valor calculado de F se compara con el valor crítico para los correspondientes grados de libertad, es una comparación de la varianza explicada por el modelo y la varianza de los residuales, de manera que cuanto más alejado esté el valor de F del valor crítico más fiable será el modelo. También se calcula p , la probabilidad de error asociada al test de Fisher. Valores elevados de F indican la fiabilidad de la ecuación del modelo QSAR. Se calcula el valor de F con $\alpha = 0.05$ mediante la Ecuación 3.15.

$$F_{gdl_{SSR}, gdl_{SSE}, \alpha} = \frac{SSR/k}{SSE/(n - d - 1)} \quad [3.15]$$

- Se ha descrito que cuando SSE es menor que SST un modelo QSAR es significativo y predice mejor que al azar, además se describe que un modelo QSAR es dependiente si se cumple la condición expresada en la Ecuación 3.16 y cuanto menor sea el valor de este cociente mejor es el modelo.^[14]

$$\frac{SSE}{SST} < 0.4 \quad [3.16]$$

- El test de Grubbs implementado en MOE permite detectar *ouliers* mediante el cálculo de *Z-Score* (Ecuación 3.17), que se puede interpretar como el número de desviaciones estándar de la media, de manera que los compuestos que presentan un valor mayor o igual a 2.5 deben ser cuidadosamente examinados.

$$Z - Score = \frac{|x - y|}{RMSE} \quad [3.17]$$

- Tropsha *et al*^[15] considera un modelo QSAR predictivo si cumple la condición definida por la Ecuación 3.18, donde R^2 y R_0^2 son los coeficientes de correlación para el modelo sin fijar la intercepción y fijando la intercepción a 0 respectivamente, cuyos valores además deben ser parecidos (Figura 3.3).

$$\frac{R^2 - R_0^2}{R^2} < 0.1 \quad [3.18]$$

- Otra de las condiciones descritas por Tropsha *et al*^[15] para asegurar la predictividad de un modelo se presenta en la Ecuación 3.19, donde k es el valor de la pendiente de la regresión cuando la intercepción es 0, cuanto más cercano a 1 mejor (Figura 3.3).

$$0.85 \leq k \leq 1.15 \quad [3.19]$$

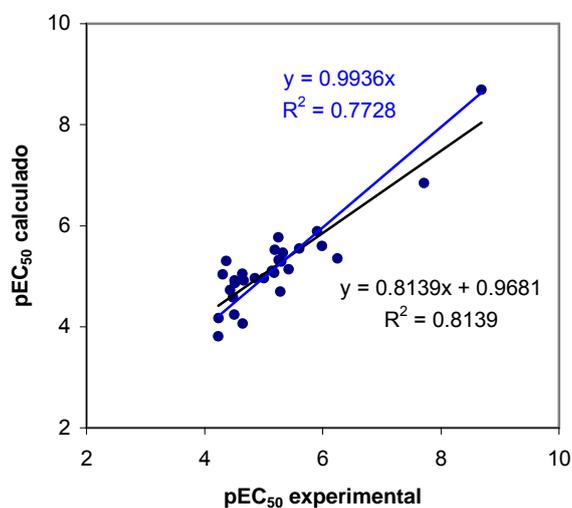


Figura 3.3. Modelo sin fijar la intercepción (negro) y fijando la intercepción a 0 (azul) donde se muestran además el valor de R^2 en cada caso y el valor de la pendiente cuando la intercepción es 0 (ecuación de la recta en azul).

3.3. Selección de *training set* y *test set*.

Se compila una base de datos de 39 compuestos con actividad anti-VIH conocida, 38 procedentes de la quimioteca de análogos al AMD3100 diseñada en el Capítulo 1 y el propio AMD3100. Esta base de datos de compuestos se divide en dos subconjuntos, *training set* (30 compuestos) y *test set* (9 compuestos), tal que ambos subconjuntos presentan el mismo perfil de número de compuestos en los diferentes intervalos del valor de la actividad anti-VIH expresada como pEC₅₀ (Figura 3.4). El *training set* se utiliza para el cálculo de los modelos QSAR mientras que el *test set* se utiliza para evaluar la calidad de los diferentes modelos obtenidos. Se recomienda disponer de entre un 20% y un 30% de la base de datos para construir el *test set*.^[16] Además, es aconsejable que el *test set* esté formado por más de cinco compuestos, cuyos valores de actividad y estructuras deben cubrir el rango de actividades y estructuras de los compuestos del *training set*. Esta condición es necesaria para obtener estadísticas fiables de la comparación de los valores de actividad experimentales y calculados con los modelos QSAR.^[15]

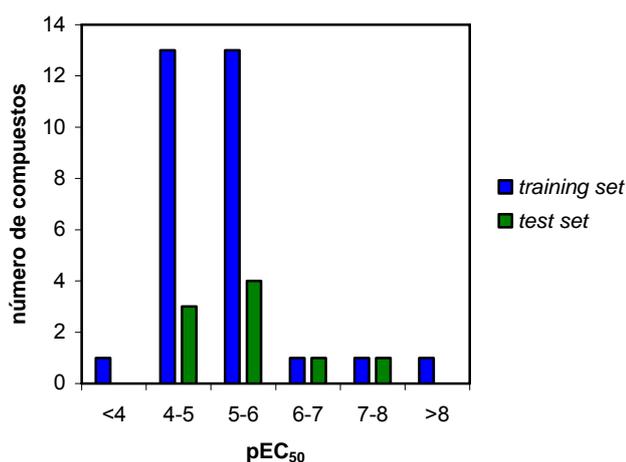


Figura 3.4. Distribución de los compuestos según el valor de su actividad anti-VIH expresada en pEC₅₀ del training set (azul) y del test set (verde)

Después de realizar la división de la base de datos el *training set* queda constituido por las moléculas: **37{1,11}**, **37{2,4}**, **37{2,5}**, **37{2,9}**, **37{3,5}**, **37{3,6}**, **37{3,8}**, **37{3,9}**, **37{3,11}**, **32{4,4}**, **32{4,5}**, **32{4,6}**, **32{4,9}**, **32{5,5}**, **32{5,6}**, **32{5,7}**, **32{5,10}**, **32{6,7}**, **32{6,8}**, **32{6,11}**, **32{7,9}**, **32{7,11}**, **32{8,8}**, **32{8,10}**, **32{8,11}**, **32{9,9}**, **36{1}**, **36{2}**, **35{9}** y AMD3100 (Anexo 2)

Mientras que el *test set* está formado por: **37{1,5}**, **37{2,11}**, **32{5,8}**, **32{5,9}**, **32{5,11}**, **32{8,9}**, **32{9,11}**, **35{5}** y **35{8}** (Anexo 3).

3.4. Alineamiento.

Los compuestos de la base de datos se caracterizan mediante descriptores moleculares, se han calculado 191 descriptores 2D y 3D internos y externos con MOE 2005.06.^[17] Los descriptores 3D, como su nombre indica, dependen de la configuración y disposición en el espacio de las moléculas, por lo tanto es importante escoger una conformación 3D adecuada para los compuestos de la quimioteca. Se ha considerado como conformación adecuada la que adopta el AMD3100 al interactuar con el correceptor CXCR4, determinada por *docking* sobre la estructura modelada del correceptor,^[18] y se ha utilizado como plantilla para alinear los compuestos de la base de datos, tanto del *training set*, como del *test set*. Para el alineamiento se han utilizado estos compuestos protonados a pH fisiológico. Dado que el estado de ionización de los átomos influye en el valor que toman ciertos descriptores, es importante cargar correctamente los compuestos. Se han calculado los pK_a de cada compuesto de la base de datos con el módulo pK_a v8.02 de ChemSketch ACD,^[19] los protones que han presentado un $pK_a \geq 7$ se han ionizado. Además se ha calculado las cargas parciales las moléculas y se ha optimizado la geometría con el *force field* MMFF94.

Cada compuesto se alinea con la plantilla escogida, el AMD3100, utilizando el módulo *FlexAlign* de MOE y fijando los siguientes parámetros de similitud:

- dador de puente de hidrógeno
- aceptor de puente de hidrógeno
- aromaticidad
- hidrógenos polares
- volumen

Los alineamientos se evalúan mediante los parámetros F (similitud), U (energía) y S (score):^[20]

F es el *score* de la similitud, es una medida basada en la superposición de las densidades- P de las dos conformaciones alineadas (AMD3100 fijado en la conformación en el sitio de unión y una conformación generada estocásticamente para cada compuesto de la base de datos), donde P se refiere a los parámetros del alineamiento o propiedades (aromático, dador o aceptor de puente de hidrógeno, etc.) y la densidad es la probabilidad de que un punto en el espacio tenga la propiedad anteriormente definida. La función de densidad de probabilidad utilizada es una función gaussiana definida por la Ecuación 3.20, donde s^2 es la varianza inversa a lo largo de cada eje y x_0 es el centro (y media) de la densidad.

$$f(x) = s^3(2\pi)^{-\frac{2}{3}} \exp\left(-\frac{1}{2}|x-x_0|^2 s^2\right) \quad [3.20]$$

Un átomo con núcleo en x_0 y radio de van der Waals r se representa mediante la Ecuación 3.21, donde a y r determinan el ancho de la densidad.

$$\left(\frac{a}{r}\right)^3 (2\pi)^{\frac{3}{2}} \exp\left(-\frac{1}{2}|x-x_0|^2\left(\frac{a}{r}\right)^2\right) \quad [3.21]$$

La posición 3D de n átomos de una molécula dada una conformación se definen como x_1, \dots, x_n . El grado en que un átomo i presenta una propiedad P se define con el parámetro no negativo w_i . La densidad- P de una conformación se define mediante la Ecuación 3.22, donde se asigna a cada punto del espacio x la probabilidad, expresada como suma de densidades gaussianas, de que x presenta esa propiedad.

$$f_P(x) = \sum_{i=1}^n \frac{w_i}{n} \left(\frac{1}{2\pi} \frac{a^2}{r_i^2}\right)^{\frac{3}{2}} \exp\left(-\frac{a^2}{2} \frac{|x-x_i|^2}{r_i^2}\right) \quad [3.22]$$

El valor del parámetro w_i se define según la propiedad:

- ácido/base: $w_i=1$ si el átomo i es base, -1 si es ácido, y 0 en otros casos
- aromaticidad: $w_i=1$ si el átomo i es aromático, y 0 en caso contrario
- exposición superficial: w_i es la fracción de área superficial de van der Waals expuesta para el átomo i
- aceptor de puente de hidrógeno: $w_i=1$ si el átomo i es aceptor de puente de hidrógeno, y 0 en caso contrario
- dador de puente de hidrógeno: $w_i=1$ si el átomo i es dador de puente de hidrógeno, y 0 en caso contrario
- hidrofobicidad: $w_i=1$ si el átomo i es hidrofóbico, y 0 en caso contrario
- $\log P(o/w)$: w_i es la contribución del átomo i a $\log P(\text{octanol/agua})$ utilizando el método de Wildman y Crippen.^[21]
- refractividad molar: w_i es la contribución del átomo i a la refractividad molar
- carga parcial del *force field*: w_i es la carga parcial del átomo i .
- volumen: w_i es 1 para cada átomo.

Dadas dos densidades- P para dos moléculas, la superposición de las dos densidades es una suma de densidades gaussianas en las distancias interatómicas expresada mediante la Ecuación 3.23.

$$F_P = \sum_{i=1}^n \sum_{j=1}^{n'} \frac{w_i}{n} \frac{w'_j}{n'} \left(\frac{a^2}{2\pi(r_i^2 + s_j^2)}\right)^{\frac{3}{2}} \exp\left(-\frac{a^2}{2} \frac{|x_i - x'_j|^2}{r_i^2 + r_j'^2}\right) \quad [3.23]$$

Esta ecuación de superposición se puede generalizar a más de una propiedad mediante la suma de las ecuaciones individuales de superposición. Dadas tres propiedades definidas por los pesos de las propiedades atómicas u , v y w , se expresa la similitud de dos

conformaciones moleculares mediante la Ecuación 3.24, donde C_u , C_v y C_w son los pesos positivos de enfatización o no enfatización de las densidades- P particulares.

$$F = \sum_{i=1}^n \sum_{j=1}^{n'} \frac{C_u u_i u'_j + C_v v_i v'_j + C_w w_i w'_j}{nn'} \left(\frac{a^2}{2\pi(r_i^2 + r_j'^2)} \right)^{\frac{3}{2}} \exp \left(-\frac{a^2}{2} \frac{|x_i - x'_j|^2}{r_i^2 + r_j'^2} \right) \quad [3.24]$$

Finalmente, el score de F se calcula con la Ecuación 3.25, de manera que cuanto menor es su valor mayor es la similitud entre los dos compuestos alineados.

$$F - \text{Score} = -kT \log F \quad [3.25]$$

U es la energía media de tensión de los compuestos alineados, expresada en kcal/mol. Se calcula como la suma de las energías potenciales individuales, E , dividida por el número de compuestos. La energía potencial del *force field* MMFF94 está compuesta por la suma de los diferentes términos energéticos expresados en las Ecuaciones 3.26-3.32, que corresponden al estiramiento de enlaces, doblamiento de ángulos, interacciones no enlazantes (van der Waals y electrostáticas), torsión de ángulos diedros y términos de interacciones cruzadas.^[22]

La energía de enlace se calcula mediante la Ecuación 3.26, donde k_{enlace} es la constante de fuerza, $r-r_0$ es la longitud del enlace y c_s es la constante cúbica de estiramiento.

$$E_{enlace} = k_{enlace}(r - r_0)^2 + \left(1 + c_s(r - r_0) + \frac{7}{12}(c_s^2(r - r_0)^2) \right) \quad [3.26]$$

La energía de los ángulos se calcula mediante la Ecuación 3.27, donde k_θ es la constante de fuerza, $\theta-\theta_0$ es el ángulo considerado y c_b es la constante cúbica de flexión.

$$E_{angulo} = k_\theta(\theta - \theta_0)^2(1 + c_b(\theta - \theta_0)) \quad [3.27]$$

La energía de torsión se calcula mediante la Ecuación 3.28, donde V son las constantes de fuerza para los términos en la serie de Fourier y θ es el ángulo diedro.

$$E_{torsion} = 0.5(V_1(1 + \cos \theta)) + V_2(1 + \cos 2\theta) + V_3(1 + \cos 3\theta) \quad [3.28]$$

La energía de las interacciones no enlazantes de van der Waals se calcula con la Ecuación 3.29, donde r es la distancia entre los dos átomos considerados, r^* es la distancia de interacción de mínima energía entre los mismos átomos basada en las polarizabilidades atómicas parametrizadas y ε_{ij} es la profundidad del pozo de potencial.

$$E_{vdW} = \varepsilon_{ij} \left(\frac{1.07r^*}{r + 0.07r^*} \right)^7 \left(\frac{1.12r^{*7}}{r^7 + 0.07r^{*7}} - 2 \right) \quad [3.29]$$

La energía de las interacciones no enlazantes de tipo electrostático se expresa mediante la Ecuación 3.30, donde D es la constante dieléctrica, q es la carga del átomo, r es la distancia y δ es la constante electrostática.

$$E_{\text{electrostatica}} = \frac{q_i q_j}{D(r_{ij} + \delta)^n} \quad [3.30]$$

La energía de interacciones cruzadas de extensión-flexión se calcula mediante la Ecuación 3.31.

$$E_{\text{extension-flexion}} = (k_r(r - r_0) + k_r'(r' - r'_0)) \cdot (\theta - \theta_0) \quad [3.31]$$

La energía de flexión fuera del plano sigue la Ecuación 3.32, donde χ es el ángulo entre el enlace y el plano considerados.

$$E_{\text{oop}} = k_{\text{oop}} \chi^2 \quad [3.32]$$

Finalmente, para evaluar un alineamiento se utiliza el parámetro S o *score* del alineamiento, y se calcula mediante la Ecuación 3.33. Cuanto menor es su valor mejor es el alineamiento de las moléculas.

$$S = -kT \log F + U \quad [3.33]$$

Una vez generados los diferentes alineamientos para cada molécula de la base de datos con el AMD3100 se selecciona el que presenta un menor valor de S . En la figura 3.5 se puede ver el alineamiento de varios compuestos de la base de datos frente a la plantilla utilizada, el AMD3100.

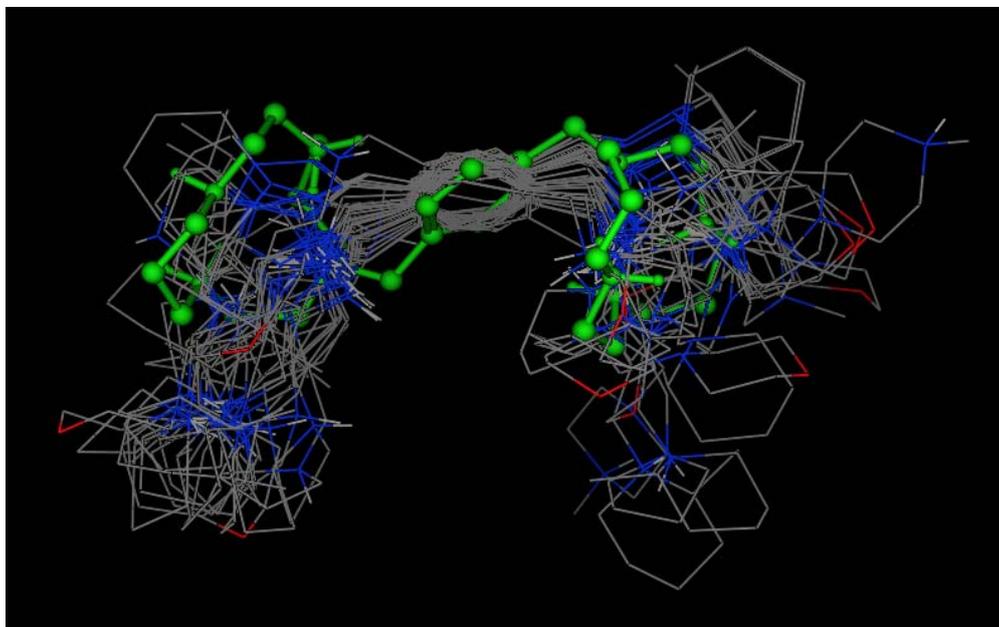


Figura 3.5. Alineamiento de algunos compuestos de la base de datos frente a la conformación del AMD3100 en el sitio de unión (en verde).

3.5. Selección de descriptores.

Los compuestos de la base de datos se han caracterizado mediante descriptores moleculares, se han calculado 191 descriptores 2D y 3D internos y externos con MOE 2005.06 (Anexo 1).

Los descriptores 2D incluyen propiedades físicas como número de átomos, número de enlaces, coeficiente de partición octanol/agua ($\log P$), refractividad molecular, área de van der Waals, etc.;^[21] áreas superficiales subdivididas que se basan en el área aproximada de la superficie accesible de van der Waals para cada átomo junto con alguna propiedad atómica; los índices de Kier y Hall;^[23,24] descriptores de matrices de distancia y adyacencia;^[25-27] descriptores de características farmacofóricas (dadores y aceptores de puente de hidrógeno, grupos polares, cargas positivas y negativas, hidrofobicidad, etc.); y cargas parciales calculadas mediante el método PEOE (*partial equalization of orbital electronegativities*).^[28,29]

Los descriptores 3D internos utilizan información de la geometría pero es independiente de la orientación absoluta de una conformación, mientras que los externos sí son dependientes de esta orientación. Los descriptores 3D incluyen los de energía potencial como la energía potencial (E), la componente electrostática de la energía potencial (E_{ele}), la energía de solvatación (E_{sol}), etc. que se calculan utilizando el *force field* MMFF94; área superficial, volumen y forma como el área de la superficie accesible al agua (ASA), el volumen de van der Waals (vol), el área de la superficie de van der Waals (VSA), etc. que dependen de la conectividad de la estructura y su conformación; y los descriptores de carga, dependientes de la conformación, como el área de la superficie accesible al agua de todos los átomos con carga positiva ($ASA+$) o con carga negativa ($ASA-$), el área de la superficie accesible al agua de todos los átomos hidrofóbicos (ASA_H), etc. que dependen de las cargas parciales y de la conformación que adopta el compuesto. El área de la superficie accesible al agua se calcula considerando un radio de 1.4 Å para la molécula de agua.

La selección de descriptores para los diferentes modelos se hace mediante matrices de correlación de los descriptores entre sí y con la actividad biológica. Se seleccionan descriptores que tengan la máxima correlación con la actividad y la mínima correlación entre ellos (Tabla 3.1). Estos descriptores son:

- $VAdjEq$ que es el índice de contenido medio de información de igualdad de adyacencia.
- Q_VSA_HYD , que es el área total hidrofóbica de la superficie de van der Waals, calculada como la suma del área de la superficie de van der Waals de los átomos cuyo valor absoluto de carga parcial es menor o igual a 2.
- E_{oop} , que es la energía potencial de un compuesto fuera del plano.
- *dipole* Y , que es la componente y del momento dipolar, depende de las coordenadas externas de la molécula.

- *SlogP_VSA8*, que es la suma del área aproximada de la superficie accesible de van der Waals para los átomos cuya contribución al $\log P(o/w)$ está entre 0.3 y 0.4.
- *SlogP_VSA9*, que es igual que el anterior pero se suman las áreas de la superficie accesible de van der Waals de los átomos cuya contribución al $\log P(o/w)$ es superior a 0.4.
- *SMR_VSA5*, que es igual que el anterior pero se suman las áreas de la superficie accesible de van der Waals de los átomos cuya contribución a la refractividad molar está entre 0.44 y 0.485.
- *FASA+*, que es el cociente entre el área de la superficie accesible al agua de todos los átomos con carga parcial positiva (*ASA+*) y el área de la superficie accesible al agua (*ASA*).

Cabe destacar que el descriptor que presenta una mayor correlación con la actividad es *Q_VSA_HYD*, con un valor de 84.

Tabla 3.1. Matriz de correlación de los descriptores más relevantes de la base de datos de compuestos y su actividad anti-VIH, donde pEC_{50} es la actividad expresada como $-\log(EC_{50})$.

	<i>pEC₅₀</i>	<i>VAdjEq</i>	<i>Q_VSA_HYD</i>	<i>E_oop</i>	<i>dipoleY</i>	<i>SlogP_VSA8</i>	<i>SlogP_VSA9</i>	<i>SMR_VSA5</i>	<i>FASA+</i>
<i>pEC₅₀</i>	100								
<i>VAdjEq</i>	-60	100							
<i>Q_VSA_HYD</i>	84	-68	100						
<i>E_oop</i>	11	-14	14	100					
<i>dipoleY</i>	41	-14	20	-25	100				
<i>SlogP_VSA8</i>	56	-26	41	-39	42	100			
<i>SlogP_VSA9</i>	41	-35	36	-26	18	59	100		
<i>SMR_VSA5</i>	54	-39	37	-24	49	68	60	100	
<i>FASA+</i>	38	-16	59	47	-6	-34	-32	-12	100

3.6. Modelos QSAR: ecuación, validación y parámetros estadísticos.

3.6.1. Modelos QSAR de 3 descriptores.

Se han construido diferentes modelos de 3 descriptores, de los 191 inicialmente calculados con MOE, combinando los ocho descriptores seleccionados en el apartado anterior, sin embargo, sólo se presentan aquellos modelos que han obtenido los valores más elevados del coeficiente de correlación R^2 , tanto para el modelo como para la validación cruzada LOO y la validación con el *test set*, y los mejores valores en los parámetros del análisis estadístico.

3.6.1.1. Modelo 1.

El modelo 1 se calcula utilizando los descriptores Q_VSA_HYD , $SlogP_VSA8$ y $dipoleY$, y su relación con la actividad viene definida por la Ecuación 3.34.

$$pEC_{50} = 2.45282 + 0.00958 \cdot Q_VSA_HYD + 0.00601 \cdot SlogP_VSA8 + 0.10366 \cdot dipoleY \quad [3.34]$$

Los coeficientes de correlación y parámetros estadísticos que definen este modelo son los siguientes:

$$N=30, d=3, R^2=0.79, RMSE=0.47, Z\text{-Score}>2.5 \{6-8\}, F=31.73 (2.98), p=7.64 \cdot 10^{-9}$$

$$R^2_{LOO}=0.72, RMSE_{LOO}=0.56, R^2_{test}=0.69, RMSE_{test}=0.56, n=9$$

$$SSE=6.74, SST=31.43, R_0^2=0.73, k=0.99$$

En la Figura 3.6 se representa el gráfico de la actividad experimental frente a la calculada utilizando el modelo 1.

Este modelo tiene un compuesto, **32**{6,8}, que presenta un valor de *Z-Score* de 2.55 (Figura 3.8). El test de Grubbs define que compuestos que presenten este parámetro con un valor superior a 2.5 deben ser eliminados del *training set* y recalcularse el modelo porque pueden representar un punto estadísticamente erróneo o *outlier*.

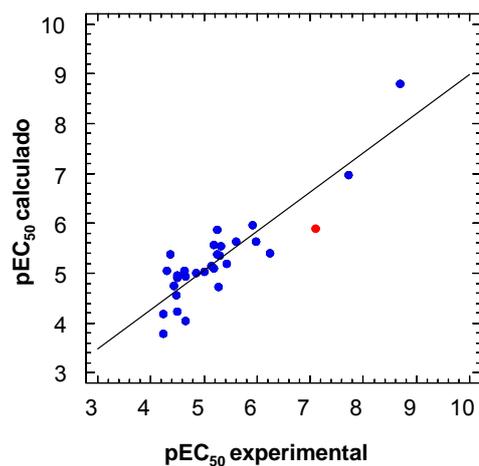


Figura 3.6. Gráfico de los valores experimentales de pEC_{50} frente a los valores calculados con el modelo 1 para el training set. El punto rojo señala la posición el outlier.

3.6.1.2. Modelo 2.

Se elimina de la base de datos del *training set* el *outlier* detectado en el modelo anterior y se recalcula la ecuación utilizando los mismos descriptores, Q_VSA_HYD , $SlogP_VSA8$ y $dipoleY$ (Ecuación 3.35).

$$pEC_{50} = 2.52586 + 0.00940 \cdot Q_VSA_HYD + 0.00507 \cdot SlogP_VSA8 + 0.10611 \cdot dipoleY \quad [3.35]$$

Los coeficientes de correlación y parámetros estadísticos que definen este modelo son los siguientes:

$$N=29, d=3, R^2=0.81, RMSE=0.42, Z-Score < 2.5, F=36.45 (2.99), p=2.78 \cdot 10^{-9}$$

$$R^2_{LOO}=0.75, RMSE_{LOO}=0.49, R^2_{test}=0.69, RMSE_{test}=0.57, n=9$$

$$SSE=5.20, SST=27.93, R_0^2=0.77, k=0.99$$

En la Figura 3.7 se representa el gráfico de la actividad experimental frente a la calculada utilizando el modelo 2.

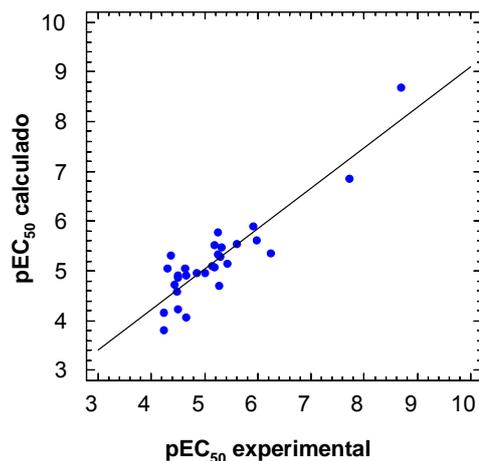


Figura 3.7. Gráfico de los valores experimentales de pEC₅₀ frente a los valores calculados con el modelo 2 para el training set.

Si se comparan los resultados obtenidos en el modelo anterior con los de este modelo, se observa que los coeficientes de correlación son más elevados en el último, el valor de F es ligeramente más elevado y los errores cuadráticos medios son menores. La eliminación de **32{6,8}** ha mejorado la ecuación de la regresión y por lo tanto se confirma su posición como *outlier*. En resumen, el modelo 2 es más predictivo que el modelo 1 para los sistemas polinitrogenados de la quimioteca diseñada. Además, en el modelo 2 no hay ningún compuesto con un valor de $Z\text{-Score} > 2.5$, por lo tanto se considera que no hay ningún *outlier* en el *training set*. Se comparan los valores de los residuales (diferencia entre la actividad experimental y la actividad calculada mediante un modelo QSAR) y del $Z\text{-Score}$ de ambos modelos (Figura 3.8). Ambos parámetros son prácticamente iguales para cada uno de los compuestos del *training set* en los dos modelos, exceptuando el compuesto **32{6,8}** ya que no está incluido en el *training set* del último modelo.

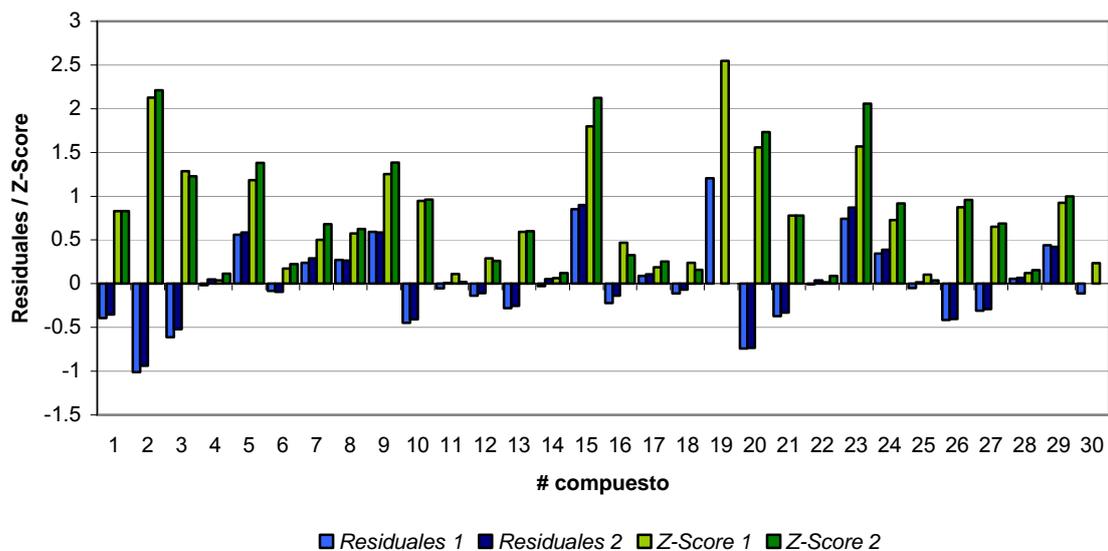


Figura 3.8. Gráfico donde se representan los valores de los residuales del modelo 1 (azul claro) y del modelo 2 (azul oscuro) junto con los valores del Z-Score para el modelo 1 (verde claro) y el modelo 2 (verde oscuro) para los compuestos del training set. El compuesto 19 representa a **32{6,8}**.

3.6.1.3. Modelo 3.

En el siguiente modelo de tres descriptores se sustituye *SlogP_VSA8* de los dos modelos anteriores por *SMR_VSA5* y se obtiene la relación de la Ecuación 3.36.

$$pEC_{50} = 2.05241 + 0.00971 \cdot Q_VSA_HYD + 0.09417 \cdot dipoleY + 0.00429 \cdot SMR_VSA5 \quad [3.36]$$

Los coeficientes de correlación y parámetros estadísticos que definen el modelo 3 son los siguientes:

$$N=30, d=3, R^2=0.79, RMSE=0.47, Z-Score < 2.5, F=31.72 (2.98), p=7.66 \cdot 10^{-9}$$

$$R^2_{LOO}=0.75, RMSE_{LOO}=0.49, R^2_{test}=0.69, RMSE_{test}=0.57, n=9$$

$$SSE=6.75, SST=31.43, R_0^2=0.73, k=0.99$$

En la Figura 3.9 se representa el gráfico de los valores de pEC_{50} experimental frente al calculado utilizando la Ecuación 3.36.

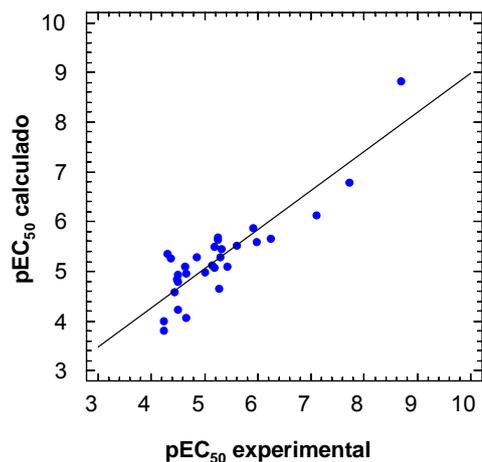


Figura 3.9. Gráfico de los valores experimentales de pEC_{50} frente a los valores calculados con el modelo 3 para el training set.

También se representan los valores de los residuales y de Z-Score obtenidos con este modelo QSAR (Figura 3.10). Se observa que en ningún caso el valor de Z-Score es superior a 2.5.

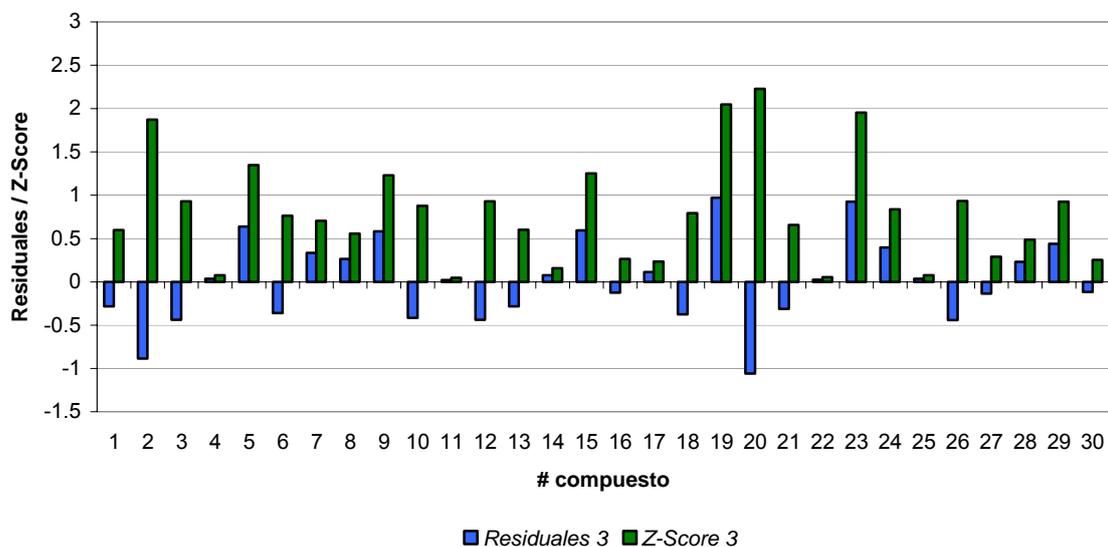


Figura 3.10. Gráfico donde se representan los valores de los residuales (azul) y de Z-Score (verde) del modelo 3 para los compuestos del training set.

3.6.2. Modelos QSAR de 4 descriptores.

3.6.2.1. Modelo 4.

La Ecuación 3.37 utiliza como variables independientes los cuatro descriptores presentes en los tres modelos anteriores y como variable dependiente pEC_{50} .

$$pEC_{50} = 2.12212 + 0.00942 \cdot Q_VSA_HYD + 0.00390 \cdot SlogP_VSA8 + 0.00278 \cdot SMR_VSA5 + 0.08712 \cdot dipoleY$$

[3.37]

Los coeficientes de correlación y parámetros estadísticos que definen el modelo 3 son los siguientes:

$$N=30, d=4, R^2=0.79, RMSE=0.47, Z\text{-Score}<2.5, F=23.85 (2.76), p=3.19 \cdot 10^{-8}$$

$$R^2_{LOO}=0.70, RMSE_{LOO}=0.57, R^2_{test}=0.68, RMSE_{test}=0.56, n=9$$

$$SSE=6.53, SST=31.43, R_0^2=0.74, k=0.99$$

En la Figura 3.11 se representa el gráfico del pEC_{50} experimental frente al calculado utilizando el modelo 4.

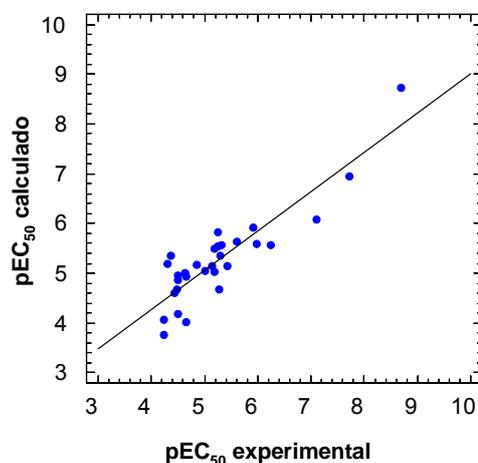


Figura 3.11. Gráfico de los valores experimentales de pEC_{50} frente a los valores calculados con el modelo 4 para el training set.

Se representan los valores de los residuales y de Z -Score obtenidos con el modelo 4 (Figura 3.12). Todos los compuestos del *training set* presentan un valor de Z -Score inferior a 2.5.

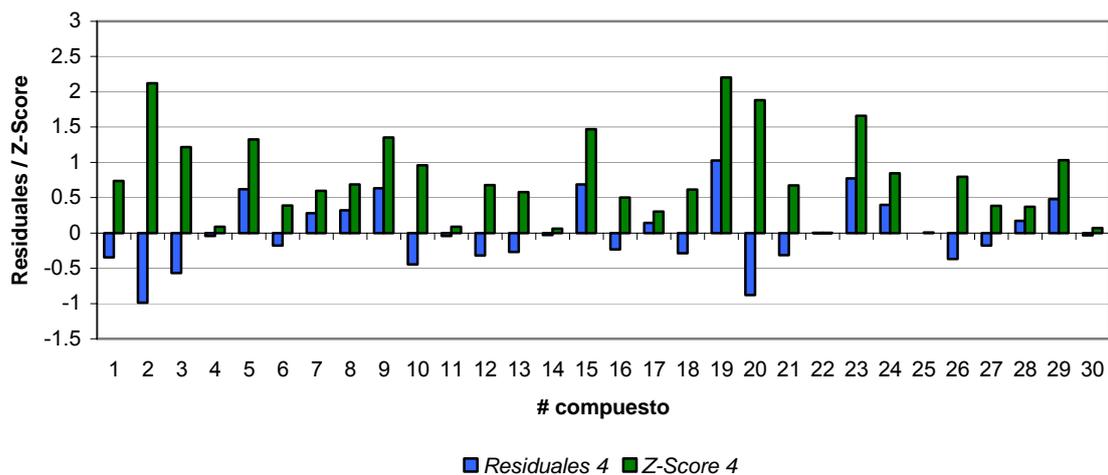


Figura 3.12. Gráfico donde se representan los valores de los residuales (azul) y de Z-Score (verde) del modelo 4 para los compuestos del training set.

3.6.2.2. Modelo 5.

A partir de los modelos 1 y 2, de tres descriptores, se obtiene la relación de la actividad con los descriptores utilizados en esos dos modelos y el descriptor *FASA+* (Ecuación 3.38).

$$pEC_{50} = 2.45265 + 0.00958 \cdot Q_VSA_HYD + 0.10367 \cdot dipoleY + 0.00602 \cdot SlogP_VSA8 + 0.00132 \cdot FASA+$$

[3.38]

Los coeficientes de correlación y parámetros estadísticos que definen este modelo se muestran a continuación:

$$N=30, d=4, R^2=0.79, RMSE=0.47, Z\text{-Score}>2.5 \{6-8\}, F=22.88 (2.76), p=4.77 \cdot 10^{-8}$$

$$R^2_{LOO}=0.72, RMSE_{LOO}=0.56, R^2_{test}=0.69, RMSE_{test}=0.56, n=9$$

$$SSE=6.74, SST=31.43, R^2_0=0.73, k=0.99$$

De forma análoga al modelo 1, este modelo también presenta un *outlier*, el compuesto **32**{6,8} con un valor de $Z\text{-Score}>2.5$.

En la Figura 3.13 se representa el gráfico de la actividad experimental frente a la calculada utilizando el modelo 5.

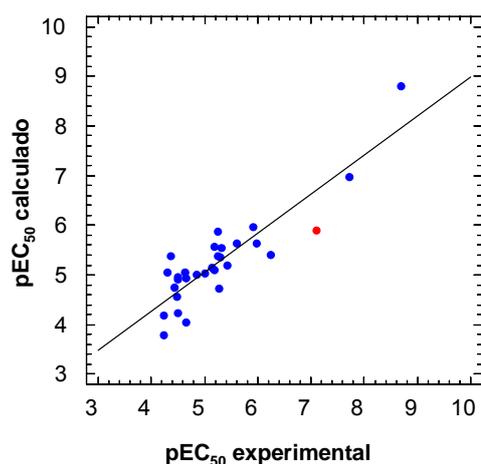


Figura 3.13. Gráfico de los valores experimentales de pEC_{50} frente a los valores calculados con el modelo 5 para el training set. El punto rojo es el outlier.

3.6.2.3. Modelo 6.

Se elimina de la base de datos del *training set* el *outlier* detectado en el modelo anterior y se recalcula la ecuación utilizando los mismos descriptores, Q_VSA_HYD , $SlogP_VSA8$, $dipoleY$ y $FASA+$ (Ecuación 3.39).

$$pEC_{50} = 2.52568 + 0.00940 \cdot Q_VSA_HYD + 0.10611 \cdot dipoleY + 0.00507 \cdot SlogP_VSA8 + 0.00130 \cdot FASA+$$

[3.39]

Los coeficientes de correlación y parámetros estadísticos que definen este modelo son los siguientes:

$$N=29, d=4, R^2=0.81, RMSE=0.42, Z\text{-Score}<2.5, F=26.24 (2.78), p=1.86 \cdot 10^{-8}$$

$$R^2_{LOO}=0.75, RMSE_{LOO}=0.49, R^2_{test}=0.69, RMSE_{test}=0.57, n=9$$

$$SSE=5.20, SST=27.93, R_0^2=0.77, k=0.99$$

En la Figura 3.14 se representa el gráfico de la actividad experimental frente a la calculada utilizando el modelo 6.

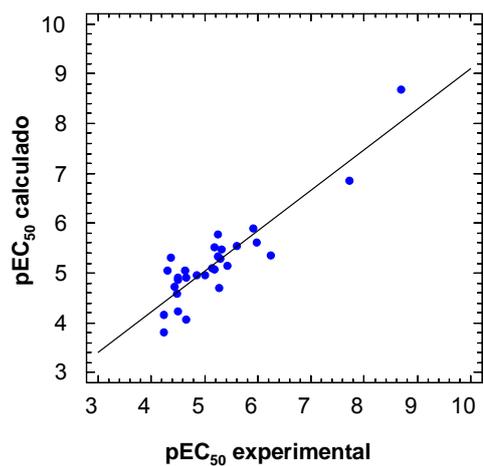


Figura 3.14. Gráfico de los valores experimentales de pEC₅₀ frente a los valores calculados con el modelo 6 para el training set.

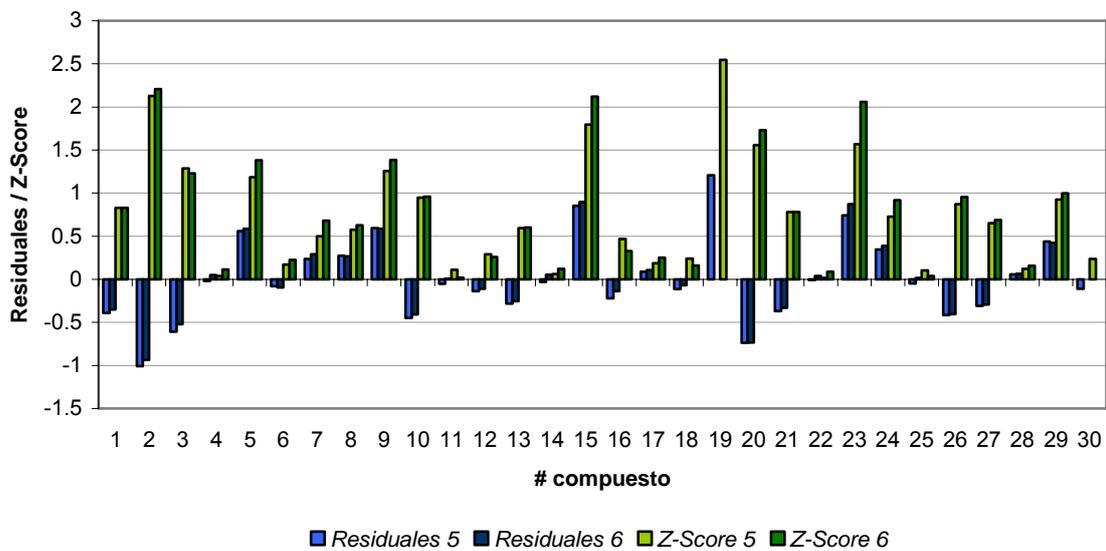


Figura 3.15. Gráfico donde se representan los valores de los residuales del modelo 5 (azul claro) y del modelo 6 (azul oscuro) junto con los valores del Z-Score para el modelo 5 (verde claro) y el modelo 6 (verde oscuro) para los compuestos del training set. El compuesto 19 representa a **32{6,8}**.

Si se comparan los resultados obtenidos en el modelo anterior con los de este modelo, se observa que los coeficientes de correlación son más elevados en el último, el valor de F es ligeramente más elevado y los errores cuadráticos medios del ajuste y de la validación cruzada LOO son menores. En resumen, la eliminación de **32{6,8}** ha mejorado la ecuación de la regresión y por lo tanto se confirma su posición como *outlier*. Además, en el modelo 6 no hay ningún compuesto con un valor de $Z\text{-Score} > 2.5$, es decir, no hay ningún *outlier*. Se comparan los valores de los residuales (diferencia entre la actividad experimental y la actividad calculada mediante un modelo QSAR) y del $Z\text{-Score}$ de ambos modelos (Figura 3.15). Ambos parámetros son prácticamente iguales para cada uno de los compuestos del *training set* en los dos modelos, exceptuando el compuesto **32{6,8}** ya que no está incluido en el *training set* del último modelo.

3.6.2.4. Modelo 7.

A partir del modelo 3, de tres descriptores, se obtiene una nueva correlación de la actividad con los descriptores utilizados además del descriptor $FASA+$ (Ecuación 3.40).

$$pEC_{50} = 2.13398 + 0.01017 \cdot Q_VSA_HYD + 0.09382 \cdot dipoleY + 0.00386 \cdot SMR_VSA5 - 0.40499 \cdot FASA+$$

[3.40]

Los coeficientes de correlación y parámetros estadísticos que definen este modelo se muestran a continuación:

$$N=30, d=4, R^2=0.79, RMSE=0.47, Z\text{-Score} < 2.5, F=23.03 (2.76), p=4.47 \cdot 10^{-8}$$

$$R^2_{LOO}=0.69, RMSE_{LOO}=0.59, R^2_{test}=0.68, RMSE_{test}=0.57, n=9$$

$$SSE=6.71, SST=31.43, R_0^2=0.73, k=0.99$$

La Figura 3.16 muestra el gráfico de la actividad experimental frente a la calculada utilizando el modelo 7.

También se representan los valores de los residuales y de $Z\text{-Score}$ obtenidos con este modelo (Figura 3.17). Todos los compuestos del *training set* presentan un valor de $Z\text{-Score}$ inferior a 2.5.

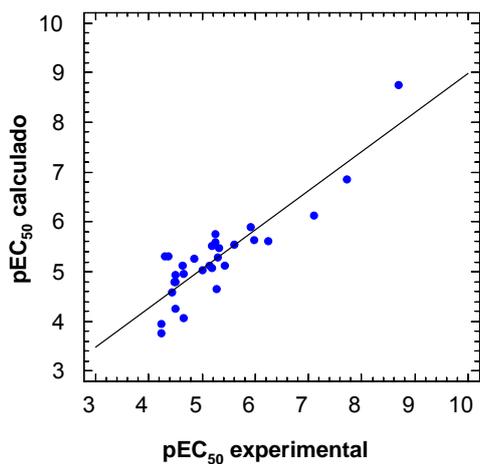


Figura 3.16. Gráfico de los valores experimentales de pEC_{50} frente a los valores calculados con el modelo 7 para el training set.

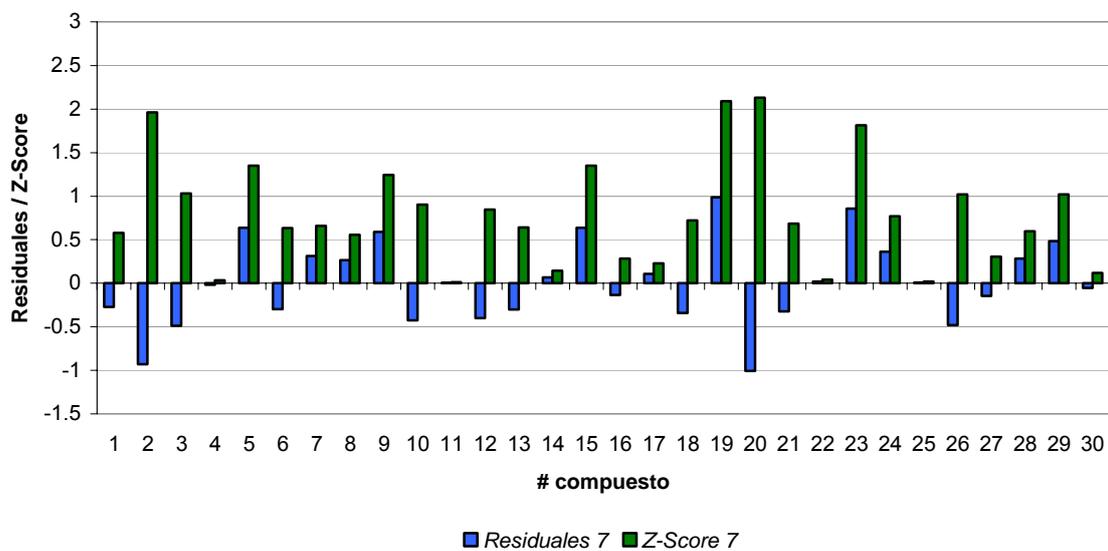


Figura 3.17. Gráfico donde se representan los valores de los residuales (azul) y de Z-Score (verde) del modelo 7 para los compuestos del training set.

3.6.2.5. Modelo 8.

A partir de los descriptores utilizados en los modelos 1 y 2, y añadiendo el descriptor $VAdjEq$, se obtiene una relación con la actividad biológica según la Ecuación 3.41.

$$pEC_{50} = 2.45309 + 0.00958 \cdot Q_VSA_HYD + 0.10367 \cdot dipoleY + 0.00601 \cdot SlogP_VSA8 - 0.00051 \cdot VAdjEq$$

[3.41]

Los coeficientes de correlación y parámetros estadísticos que definen este modelo QSAR son los siguientes:

$$N=30, d=4, R^2=0.79, RMSE=0.47, Z\text{-Score}>2.5 \{6-8\}, F=22.88 (2.76), p=4.47 \cdot 10^{-8}$$

$$R^2_{LOO}=0.72, RMSE_{LOO}=0.56, R^2_{test}=0.69, RMSE_{test}=0.56, n=9$$

$$SSE=6.74, SST=31.43, R^2_0=0.73, k=0.99$$

De forma análoga al modelo 1, este modelo también presenta un *outlier*, el compuesto **30{6,8}** con un valor de $Z\text{-Score}>2.5$.

En la Figura 3.18 se representa el gráfico de la actividad experimental frente a la calculada utilizando el modelo 8.

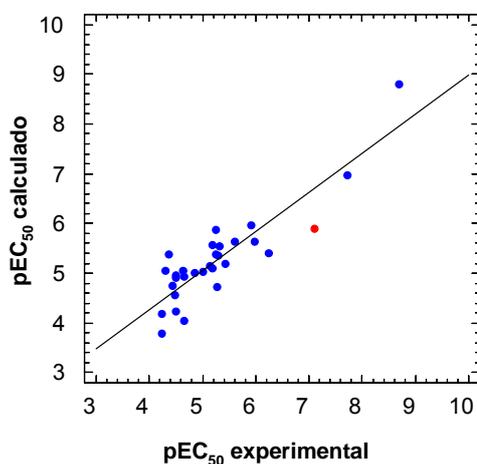


Figura 3.18. Gráfico de los valores experimentales de pEC_{50} frente a los valores calculados con el modelo 8 para el training set. El punto rojo es el outlier.

3.6.2.6. Modelo 9.

Se procede de la misma manera que en los otros casos donde se ha detectado un *outlier*, se elimina este compuesto de la base de datos del *training set* y se recalcula la ecuación del modelo utilizando los mismos descriptores, Q_VSA_HYD , $SlogP_VSA8$, $dipoleY$ y $VAdjEq$ (Ecuación 3.42).

$$pEC_{50} = 2.52606 + 0.00940 \cdot Q_VSA_HYD + 0.10611 \cdot dipoleY + 0.00507 \cdot SlogP_VSA8 - 0.00040 \cdot VAdjEq$$

[3.42]

Los coeficientes de correlación y parámetros estadísticos que definen este modelo son los siguientes:

$$N=29, d=4, R^2=0.81, RMSE=0.42, Z\text{-Score}<2.5, F=26.24 (2.76), p=1.86 \cdot 10^{-8}$$

$$R^2_{LOO}=0.75, RMSE_{LOO}=0.49, R^2_{test}=0.69, RMSE_{test}=0.57, n=9$$

$$SSE=5.20, SST=27.93, R_0^2=0.77, k=0.99$$

En la Figura 3.19 se representa el gráfico de la actividad experimental frente a la calculada utilizando este modelo.

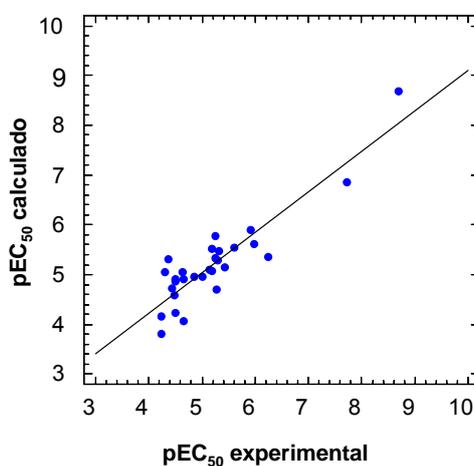


Figura 3.19. Gráfico de los valores experimentales de pEC_{50} frente a los valores calculados con el modelo 9 para el *training set*.

Si se comparan los resultados obtenidos en el modelo anterior con los de este modelo, se observa que los coeficientes de correlación son más elevados en este último, el valor de F es ligeramente más elevado y los errores cuadráticos medios son menores. Por consiguiente, la eliminación del *outlier* ha mejorado el modelo QSAR. Se comparan los valores de los residuales (diferencia entre la actividad experimental y la actividad calculada mediante un modelo QSAR) y del Z -Score de ambos modelos (Figura 3.20). Ambos parámetros son prácticamente iguales para cada uno de los compuestos del *training set* en los dos modelos, exceptuando el compuesto **32**{6,8} ya que no está incluido en el *training set* del último modelo.

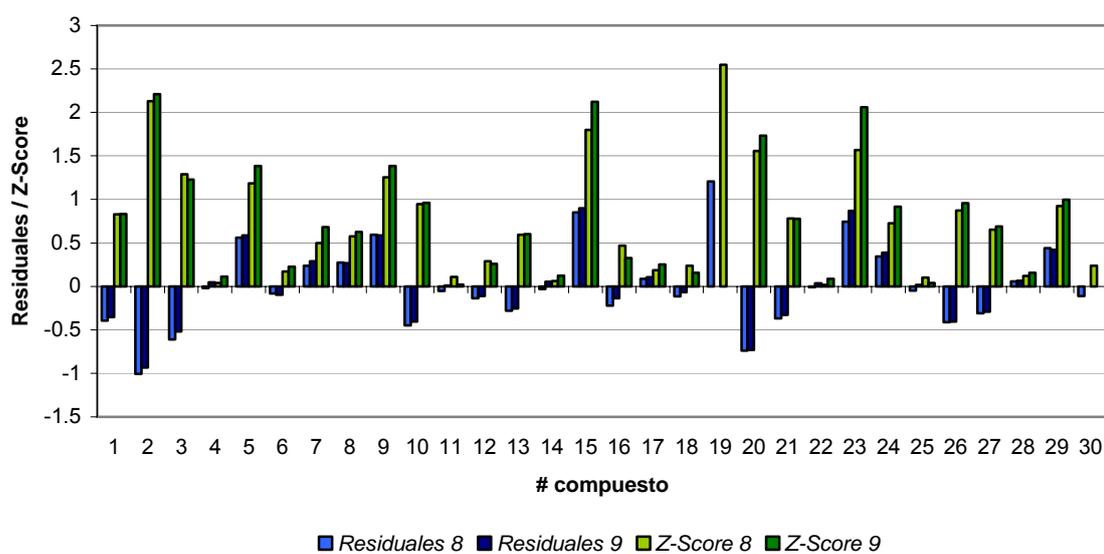


Figura 3.20. Gráfico donde se representan los valores de los residuales del modelo 8 (azul claro) y del modelo 9 (azul oscuro) junto con los valores del Z -Score para el modelo 8 (verde claro) y el modelo 9 (verde oscuro) para los compuestos del *training set*. El compuesto 19 representa a **32**{6,8}.

3.6.2.7. Modelo 10.

A partir del modelo 3, de tres descriptores, se obtiene una nueva correlación de la actividad con los descriptores utilizados incluyendo el descriptor $VAdjEq$ (Ecuación 3.43).

$$pEC_{50} = 2.05232 + 0.00971 \cdot Q_VSA_HYD + 0.09417 \cdot dipoleY + 0.00429 \cdot SMR_VSA5 + 0.00017 VAdjEq$$

[3.43]

Los coeficientes de correlación y parámetros estadísticos que definen este modelo son los siguientes:

$$N=30, d=4, R^2=0.79, RMSE=0.47, Z\text{-Score}<2.5, F=22.88 (2.76), p=4.78 \cdot 10^{-8}$$

$$R^2_{LOO}=0.72, RMSE_{LOO}=0.56, R^2_{test}=0.69, RMSE_{test}=0.58, n=9$$

$$SSE=6.75, SST=31.43, R^2_0=0.73, k=0.99$$

En la Figura 3.21 se representa el gráfico de la actividad experimental frente a la calculada utilizando este modelo.

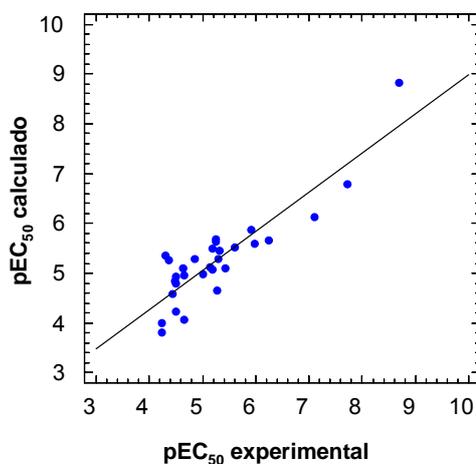


Figura 3.21. Gráfico de los valores experimentales de pEC_{50} frente a los valores calculados con el modelo 10 para el training set.

También se representan los valores de los residuales y de Z -Score obtenidos con este modelo (Figura 3.22); todos los compuestos del *training set* presentan un valor de Z -Score inferior a 2.5.

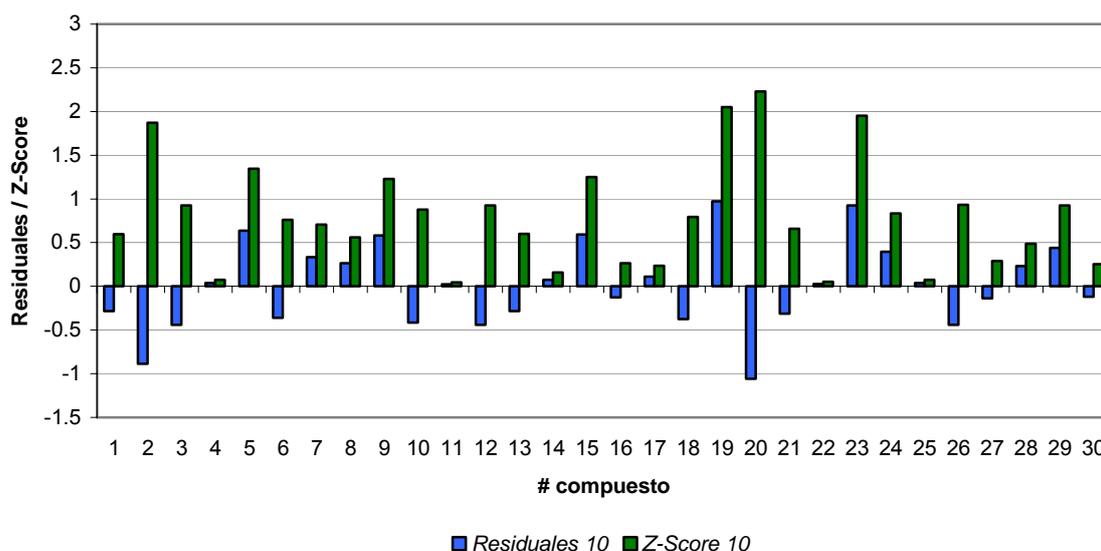


Figura 3.22. Gráfico donde se representan los valores de los residuales (azul) y de Z-Score (verde) del modelo 10 para los compuestos del training set.

3.6.2.8. Modelo 11.

Se calcula un nuevo modelo con los descriptores Q_VSA_HYD , $dipoleY$, $FASA+$ y E_{oop} , cuya relación con la actividad viene determinada por la Ecuación 3.44.

$$pEC_{50} = 2.74420 + 0.01169 \cdot Q_VSA_HYD + 0.12745 \cdot dipoleY + 0.59407 \cdot E_{oop} - 1.43021 \cdot FASA+ \quad [3.44]$$

Los coeficientes de correlación y parámetros estadísticos que definen este modelo son los siguientes:

$$N=30, d=4, R^2=0.78, RMSE=0.48, Z\text{-Score}>2.5 \{6,8\}, F=21.72 (2.76), p=7.85 \cdot 10^{-8}$$

$$R^2_{LOO}=0.66, RMSE_{LOO}=0.78, R^2_{test}=0.62, RMSE_{test}=0.59, n=9$$

$$SSE=7.02, SST=31.43, R^2_0=0.71, k=0.99$$

En la validación del modelo se detecta la presencia de un *outlier*, **32**{6,8}, ya que su valor de Z-Score es 2.60.

En la Figura 3.23 se representa el gráfico de la actividad experimental frente a la calculada utilizando este modelo.

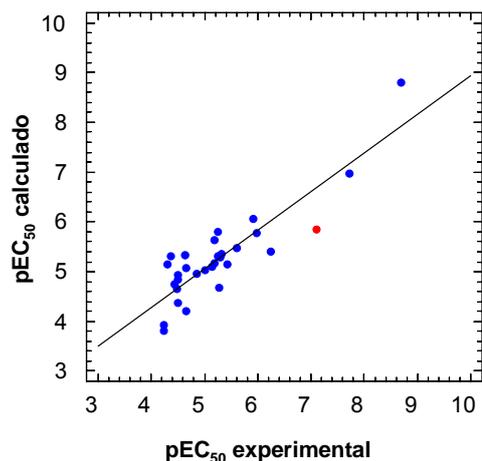


Figura 3.23. Gráfico de los valores experimentales de pEC_{50} frente a los valores calculados con el modelo 11 para el training set. El outlier está marcado como punto rojo.

3.6.2.9. Modelo 12.

Se recalcula el modelo anterior eliminando el punto erróneo detectado en la base de datos *training set* (Ecuación 3.45).

$$pEC_{50} = 2.77754 + 0.01141 \cdot Q_VSA_HYD + 0.12791 \cdot dipoleY + 1.12606 \cdot E_oop - 1.55813 \cdot FASA +$$

[3.45]

Los coeficientes de correlación y parámetros estadísticos que definen este modelo son los siguientes:

$$N=29, d=4, R^2=0.81, RMSE=0.42, Z\text{-Score}<2.5, F=26.23 (2.78), p=1.86 \cdot 10^{-8}$$

$$R^2_{LOO}=0.71, RMSE_{LOO}=0.60, R^2_{test}=0.59, RMSE_{test}=0.62, n=9$$

$$SSE=5.20, SST=27.93, R^2_0=0.77, k=0.99$$

En la Figura 3.24 se representa el gráfico de la actividad experimental frente a la calculada utilizando el modelo 12.

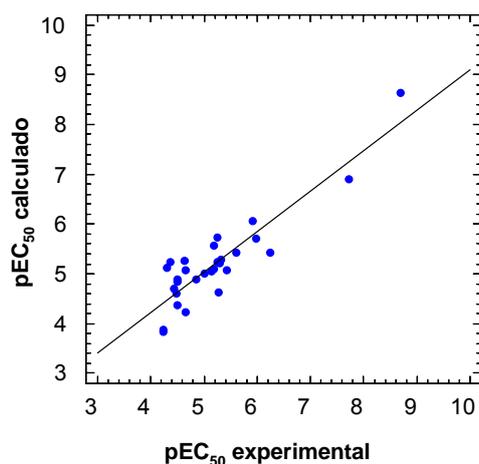


Figura 3.24. Gráfico de los valores experimentales de pEC_{50} frente a los valores calculados con el modelo 12 para el training set.

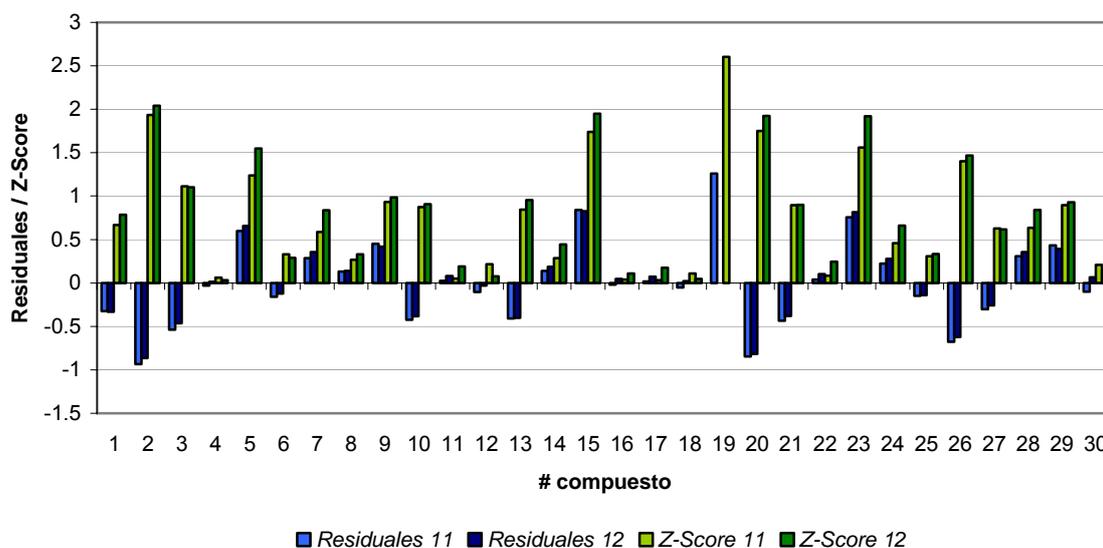


Figura 3.25. Gráfico donde se representan los valores de los residuales del modelo 11 (azul claro) y del modelo 12 (azul oscuro) junto con los valores del Z-Score para el modelo 11 (verde claro) y el modelo 12 (verde oscuro) para los compuestos del training set. El compuesto 19 representa a **32**{6,8}.

Si se comparan los resultados obtenidos en el modelo anterior con los de este modelo, se observa que los coeficientes de correlación son más elevados en este último, el valor de F es ligeramente más elevado y los errores cuadráticos medios son menores para el modelo y la validación cruzada LOO. Por consiguiente, la eliminación del *outlier* ha mejorado el modelo

QSAR. Se comparan los valores de los residuales (diferencia entre la actividad experimental y la actividad calculada mediante un modelo QSAR) y del *Z-Score* de ambos modelos (Figura 3.25). Ambos parámetros son prácticamente iguales para cada uno de los compuestos del *training set* en los dos modelos, exceptuando el compuesto **32**{6,8} ya que no está incluido en el *training set* del último modelo.

3.6.2.10. Modelo 13.

El último modelo de cuatro descriptores se genera a partir de *Q_VSA_HYD*, *SlogP_VSA8*, *SlogP_VSA9* y *SMR_VSA5*, cuya relación con la actividad viene determinada por la Ecuación 3.46.

$$pEC_{50} = 1.70723 + 0.00950 \cdot Q_VSA_HYD + 0.00546 \cdot SlogP_VSA8 - 0.00281 \cdot SlogP_VSA9 + 0.00474 \cdot SMR_VSA5$$

[3.46]

Los coeficientes de correlación y parámetros estadísticos que definen este modelo son los siguientes:

$$N=30, d=4, R^2=0.78, RMSE=0.48, Z-Score < 2.5, F=21.71 (2.76), p=7.88 \cdot 10^{-8}$$

$$R^2_{LOO}=0.67, RMSE_{LOO}=0.62, R^2_{test}=0.58, RMSE_{test}=0.60, n=9$$

$$SSE=7.03, SST=31.43, R^2_0=0.71, k=0.99$$

En la Figura 3.26 se representa el gráfico de la actividad experimental frente a la calculada utilizando este modelo.

También se representan los valores de los residuales y de *Z-Score* obtenidos con este modelo (Figura 3.27); todos los compuestos del *training set* presentan un valor de *Z-Score* inferior a 2.5.

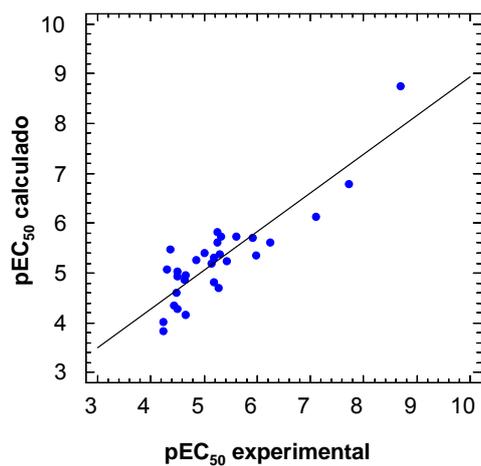


Figura 3.26. Gráfico de los valores experimentales de pEC₅₀ frente a los valores calculados con el modelo 13 para el training set.

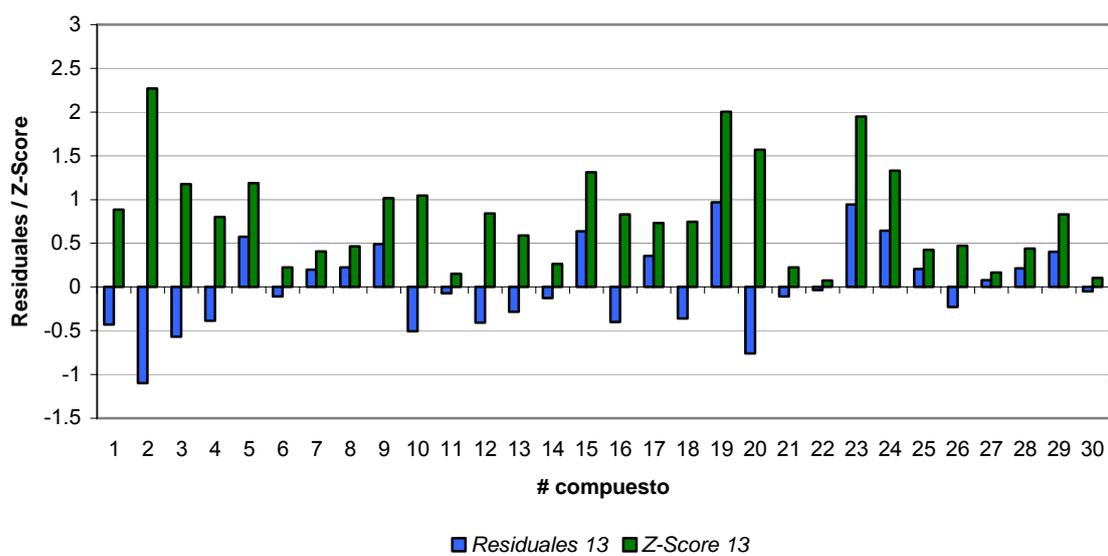


Figura 3.27. Gráfico donde se representan los valores de los residuales (azul) y de Z-Score (verde) del modelo 13 para los compuestos del training set.

3.6.3. Modelo QSAR de 5 descriptores.

3.6.3.1. Modelo 14.

Se combinan los descriptores Q_VSA_HYD , $SlogP_VSA8$, SMR_VSA5 , $dipoleY$ y $VAdjEq$ según la Ecuación 3.47.

$$pEC_{50} = 2.12216 + 0.00942 \cdot Q_VSA_HYD + 0.00390 \cdot SlogP_VSA8 + 0.00278 \cdot SMR_VSA5 + 0.08712 \cdot dipoleY - 0.00008 \cdot VAdjEq$$

[3.47]

Los coeficientes de correlación y parámetros estadísticos que definen este modelo son los siguientes:

$$N=30, d=5, R^2=0.79, RMSE=0.47, Z\text{-Score}<2.5, F=18.32 (2.62), p=1.69 \cdot 10^{-7}$$

$$R^2_{LOO}=0.70, RMSE_{LOO}=0.57, R^2_{test}=0.68, RMSE_{test}=0.56, n=9$$

$$SSE=6.53, SST=31.43, R^2_0=0.74, k=0.99$$

En la Figura 3.28 se representa el gráfico de la actividad experimental frente a la calculada utilizando este modelo.

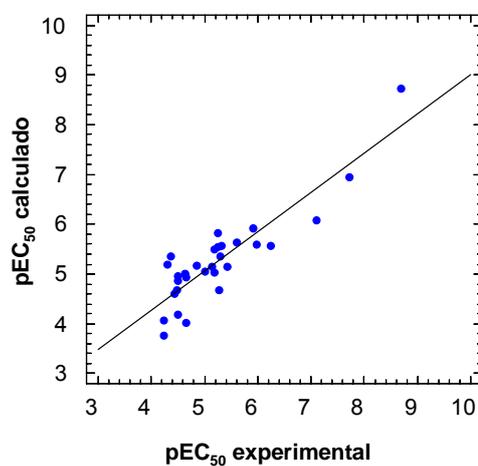


Figura 3.28. Gráfico de los valores experimentales de pEC_{50} frente a los valores calculados con el modelo 15 para el training set.

También se representan los valores de los residuales y de *Z-Score* obtenidos con el modelo 14 (Figura 3.29) y se comprueba que todos los compuestos del *training set* presentan un valor de *Z-Score* inferior a 2.5.

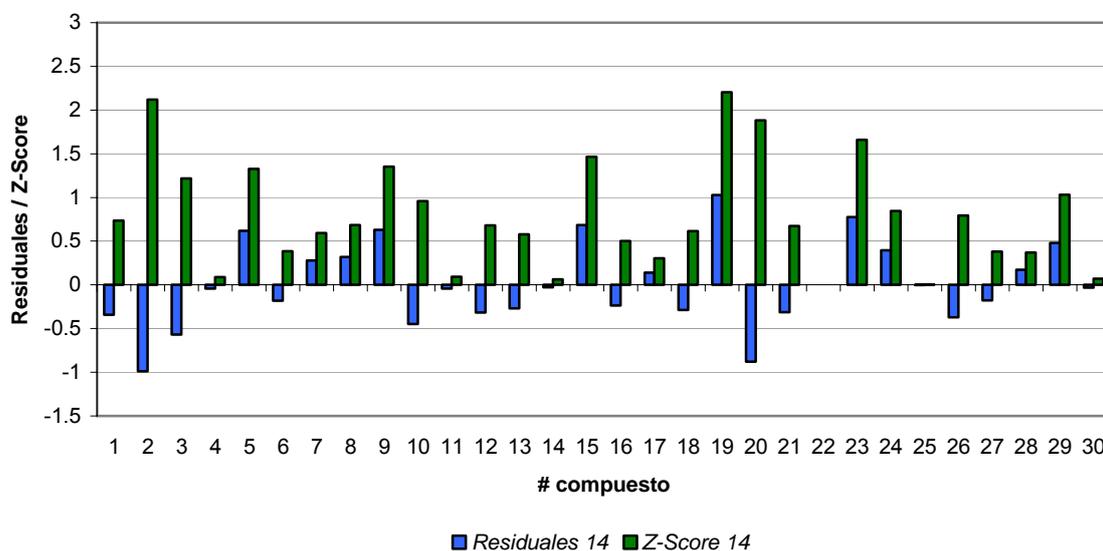


Figura 3.29. Gráfico donde se representan los valores de los residuales (azul) y de *Z-Score* (verde) del modelo 14 para los compuestos del *training set*.

3.6.4. Comparación de los modelos QSAR obtenidos.

El análisis de los modelos QSAR obtenidos se puede resumir en una serie de condiciones. En primer lugar, un modelo predictivo debe tener un elevado valor del coeficiente de correlación R^2 , superior a 0.6, pero también lo deben ser los coeficientes de correlación de la validación cruzada LOO y con el *test set*, superiores a 0.5. Además, el valor del coeficiente de correlación cuando la intercepción de la línea de regresión se fija en el origen, R_0^2 , debe ser lo más parecido posible a R^2 , de esta manera el modelo obtenido se asemejará al máximo al modelo ideal (con intercepción en 0 y pendiente 1). Teniendo en cuenta esta condición, un modelo será predictivo cuando el valor de la pendiente de la línea de regresión con intercepción en el origen sea cercano a 1. Por consiguiente un modelo con elevada capacidad de predicción debe satisfacer las condiciones descritas en la Ecuación 3.18 y la Ecuación 3.19. Se describe que un modelo es dependiente cuando se cumple la condición expresada en la Ecuación 3.16, incluso cuanto menor sea el valor de esta relación mejor es el modelo. Por último, el valor de F debe ser mayor que el valor crítico de F para el nivel de α fijado.

$$\frac{R^2 - R_0^2}{R^2} < 0.1 \tag{3.18}$$

$$0.85 \leq k \leq 1.15 \tag{3.19}$$

$$\frac{SSE}{SST} < 0.4 \tag{3.16}$$

A partir de los umbrales definidos se comparan los parámetros estadísticos obtenidos para cada uno de los modelos QSAR del apartado anterior (Tabla 3.2).

Tabla 3.2. *Parámetros estadísticos obtenidos para los modelos QSAR obtenidos (modelos 1-3 de tres descriptores, modelos 4-13 de cuatro descriptores y modelo 14 de cinco descriptores). Los modelos que han presentado un mejor rendimiento se destacan en negrita (modelos 2, 6 y 9).*

Modelo	R^2 (RMSE)	R^2_{LOO} (RMSE _{LOO})	R^2_{test} (RMSE _{test})	R_0^2	$\frac{R^2 - R_0^2}{R^2}$	k	SSE/SST	F
1	0.79 (0.47)	0.72 (0.56)	0.69 (0.56)	0.73	0.07	0.99	0.22	31.73
2	0.81 (0.42)	0.75 (0.49)	0.69 (0.58)	0.77	0.05	0.99	0.19	36.45
3	0.79 (0.47)	0.72 (0.56)	0.69 (0.58)	0.73	0.07	0.99	0.22	31.72
4	0.79 (0.47)	0.70 (0.57)	0.68 (0.56)	0.74	0.07	0.99	0.21	23.85
5	0.79 (0.47)	0.72 (0.56)	0.69 (0.56)	0.73	0.07	0.99	0.22	22.88
6	0.81 (0.42)	0.75 (0.49)	0.69 (0.58)	0.77	0.05	0.99	0.19	26.24
7	0.79 (0.47)	0.69 (0.59)	0.68 (0.57)	0.73	0.07	0.99	0.21	23.03
8	0.79 (0.47)	0.72 (0.56)	0.69 (0.56)	0.73	0.07	0.99	0.22	22.88
9	0.81 (0.42)	0.75 (0.49)	0.69 (0.58)	0.77	0.05	0.99	0.19	26.24
10	0.79 (0.47)	0.72 (0.56)	0.69 (0.58)	0.73	0.07	0.99	0.22	22.88
11	0.78 (0.48)	0.66 (0.78)	0.62 (0.59)	0.71	0.08	0.99	0.22	21.72
12	0.81 (0.42)	0.71 (0.60)	0.59 (0.62)	0.77	0.05	0.99	0.19	26.23
13	0.78 (0.48)	0.67 (0.62)	0.58 (0.60)	0.71	0.08	0.99	0.22	21.71
14	0.79 (0.47)	0.70 (0.57)	0.68 (0.56)	0.74	0.07	0.99	0.21	18.32

Todos los modelos descritos presentan valores de los coeficientes de correlación del modelo, de la validación cruzada LOO y de la validación con el *test set* por encima de los umbrales definidos. No obstante, de todos ellos destacan los modelos 2, 6, 9 y 12, que

presentan los máximos valores. Sin embargo, el modelo 12, pese a tener una correlación elevada para el modelo, en la validación (especialmente con el *test set*) no demuestra una capacidad predictiva tan buena como los otros tres. Además, estos modelos son los que más se asemejan al modelo QSAR ideal, ya que sus valores de R^2 y R_0^2 presentan la mayor similitud dentro del grupo; y se observa que el valor de k es 0.99 en todos los casos, prácticamente 1. Asimismo, en estos modelos SSE es menor que SST , y se cumple la relación mínima que deben mantener estos parámetros; cabe destacar que los modelos 2, 6, 9 y 12 presentan el menor valor de los que se han obtenido. Para todos los modelos descritos el valor de F es superior al valor de F crítico.

En general, todas las regresiones cumplen los requisitos mínimos que se han definido para que un modelo sea predictivo, pero destacan los modelos 2, 6 y 9. De todos ellos, se ha considerado el modelo 2 como el mejor, ya que está definido mediante tres descriptores mientras que los otros dos se han definido con cuatro descriptores, y por lo tanto presenta mayor F . Estos tres modelos presentan descriptores 3D en su relación con la actividad, hecho que pone de manifiesto la importancia de un alineamiento correcto de los compuestos de la base de datos con la conformación del AMD3100 en el sitio de unión. Este alineamiento también se debe realizar con los compuestos cuya actividad se desee predecir.

En la Figura 3.30 se presenta la relación entre la actividad experimental y la obtenida por predicción con el modelo 2 para las validación cruzada LOO y la validación con los compuestos del *test set*.

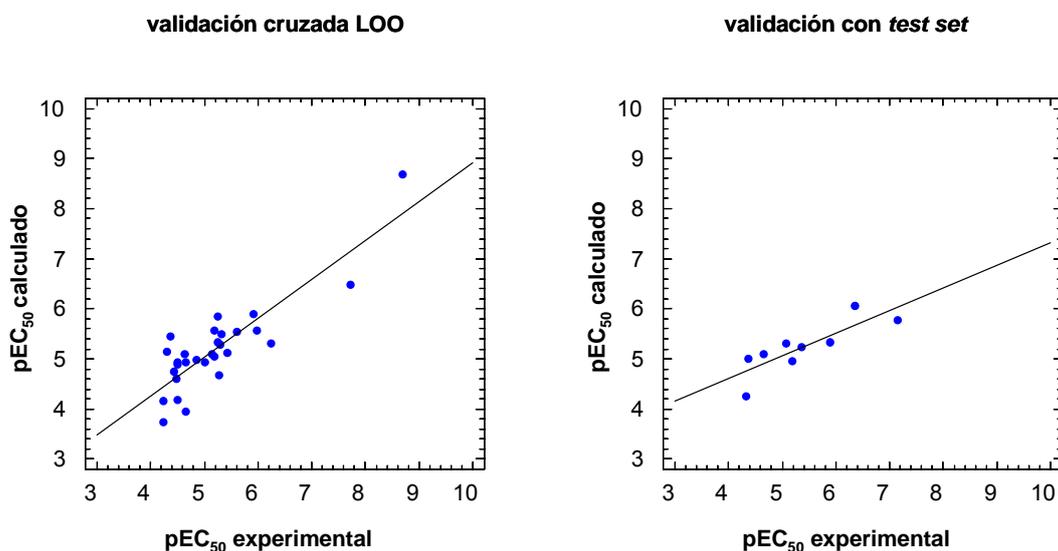


Figura 3.30. Gráfico de los valores experimentales de pEC_{50} frente a los valores calculados con el modelo 2 para la validación cruzada LOO (izquierda) y para el test set (derecha).

3.7. Árbol de clasificación.

3.7.1. Introducción.

Los modelos QSAR se han calculado a partir de una base de datos de compuestos activos con el fin de predecir la actividad anti-VIH de nuevos compuestos. Al realizar predicciones de actividad de compuestos inactivos el modelo debe realizar una extrapolación para obtener un resultado y por tanto se reduce su fiabilidad. La dificultad de incluir compuestos inactivos en el *training set* radica en que no se dispone de un valor concreto de EC_{50} sino de un umbral de concentración por debajo del cual el compuesto no ha resultado activo, es la concentración máxima utilizada al realizar los ensayos de actividad anti-VIH. Por lo tanto, se necesitan modelos que permitan discernir entre compuestos activos e inactivos, como los basados en la técnica de partición recursiva. Esta partición recursiva se traduce en una organización jerárquica del espacio de representación que puede modelarse mediante una estructura de tipo árbol de clasificación. La partición recursiva se utiliza en varios campos entre los que destacan el diagnóstico,^[30] y el cribado virtual de alto rendimiento (*high-throughput virtual screening*).^[31-35]

Un árbol de clasificación se utiliza para clasificar datos (moléculas) en clases según sus propiedades. Los árboles de clasificación se construyen a partir de un *training set* donde la clase a la que pertenece cada compuesto se conoce *a priori*, por ejemplo activo o inactivo. En cada nodo o punto de ramificación se plantea una pregunta y en función de la respuesta, el compuesto pasa a una rama u otra del árbol. Así sucesivamente hasta llegar a una hoja, o nodo final, que tiene una clase asignada. Un caso particular de los árboles de clasificación son los binarios, donde los puntos de ramificación corresponden a decisiones del tipo sí-no que dividen cada nodo en dos nodos hijos que se excluyen mutuamente. Cada nodo hijo, a su vez se divide en otros nodos hasta llegar a los nodos finales donde se asigna una clase (Figura 3.31).

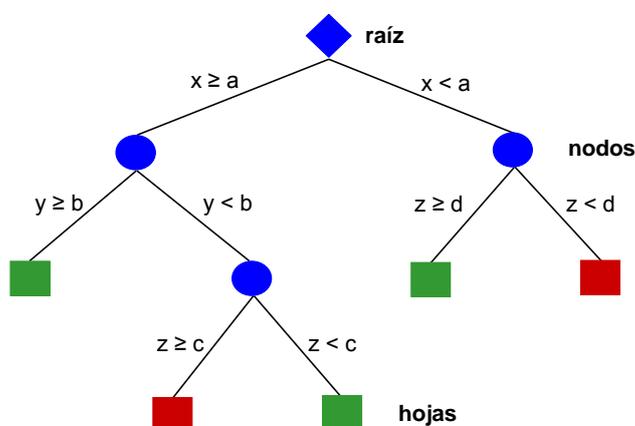


Figura 3.31. Esquema de un árbol de clasificación. El rombo representa la raíz, los círculos representan los nodos y los cuadrados representan las hojas coloreadas en verde o rojo según la clase asignada.

Los árboles de clasificación binarios de compuestos de la quimioteca diseñada se construyen a partir de un *training set*, que se utiliza para crear un modelo inicial, generalmente sobredimensionado, que se poda hasta adquirir un tamaño óptimo mediante la validación cruzada. Se puede generalizar que un árbol de clasificación es mejor cuanto menores son su tamaño y el valor del índice de clasificación incorrecta.

3.7.2. Análisis de los árboles de clasificación

Cada hoja de un árbol tiene una clase asignada, que está representada de manera mayoritaria por los compuestos que están en esa hoja. Un compuesto que no pertenece a la clase asignada en esa hoja está considerado como incorrectamente clasificado. Cada hoja t tiene un índice de clasificación incorrecta del nodo, $r(t)$, calculado como la proporción de casos incorrectamente clasificados para aquella hoja. Se define mediante la Ecuación 3.48, donde n_j es el número de casos de la clase j en la hoja t , y n_t es el número total de casos en la hoja t .

$$r(t) = 1 - \frac{n_j}{n_t} \quad [3.48]$$

La precisión de la clasificación de un árbol se expresa mediante el índice de clasificación incorrecta, $R(T)$, que cuantifica la proporción de casos clasificados incorrectamente por el árbol T y se calcula mediante la Ecuación 3.49, donde $N_{misclassified}$ es el número total de casos incorrectamente clasificados y N_{total} es el número total de casos en el *training set*.

$$R(T) = \frac{N_{misclassified}}{N_{total}} \quad [3.49]$$

Los árboles construidos se evalúan mediante una validación cruzada. En este procedimiento el *training set* se divide en V subconjuntos S_1-S_v , se construyen V árboles de clasificación $T_j (j=1, \dots, V)$ donde para cada árbol T_j se utilizan todos los compuestos excepto el subconjunto S_j , que se utiliza para validar el árbol T_j . Finalmente se selecciona el mejor modelo. En los árboles construidos se ha utilizado 5 subconjuntos para la validación cruzada. Adicionalmente se realiza una validación mediante un *test set* externo.

Para comparar el rendimiento de los árboles de clasificación construidos, además de $R(T)$, se utilizan otros parámetros.^[36]

- Exactitud, X , corresponde a la proporción de compuestos correctamente clasificados, y se calcula mediante la Ecuación 3.50, donde TP es el número de compuestos activos correctamente clasificados, TN es el número de compuestos inactivos correctamente clasificados, FP es el número de compuestos incorrectamente clasificados como activos, y FN es el número de compuestos

incorrectamente clasificados como inactivos. Su valor oscila entre 0 y 1, cuanto mayor es el valor más exacta es la clasificación.^[37]

$$X = \frac{TP + TN}{TP + TN + FP + FN} \quad [3.50]$$

- Sensibilidad, S , corresponde a la proporción de compuestos activos correctamente clasificados, y se calcula a partir de la Ecuación 3.51. Su valor oscila entre 0 y 1, cuanto mayor es su valor más sensible es la clasificación.

$$S = \frac{TP}{TP + FN} \quad [3.51]$$

- Especificidad, E , corresponde a la proporción de compuestos inactivos correctamente clasificados, y se calcula mediante la Ecuación 3.52. Su valor oscila entre 0 y 1, cuanto mayor es su valor más específica es la clasificación.

$$E = \frac{TN}{TN + FP} \quad [3.52]$$

- Precisión, P , es una medida de la habilidad de un árbol de clasificación de predecir una clase específica, y se calcula a partir de la Ecuación 3.53. Su valor oscila entre 0 y 1, cuanto mayor es su valor más precisa es la clasificación.

$$P = \frac{TP}{TP + FP} \quad [3.53]$$

- Coeficiente de correlación de Matthews, MCC , es una medida de la calidad de una clasificación binaria y se calcula mediante la Ecuación 3.53.^[38] Su valor oscila entre 1, clasificación perfecta, y -1, clasificación inversa, un valor de 0 indica una clasificación aleatoria.

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FN)(TP + FP)(TN + FN)(TN + FP)}} \quad [3.54]$$

3.7.3. Selección de *training set* y *test set*.

Se compila una base de datos de 55 compuestos con actividad anti-VIH determinada, tanto activos como inactivos. 54 moléculas proceden de la quimioteca de análogos al AMD3100 diseñada en el Capítulo 1 y además se incluye el propio AMD3100. Esta base de datos de compuestos se divide en dos subconjuntos, *training set* (44 compuestos) y *test set* (11 compuestos), ambos con aproximadamente un 30% de compuestos inactivos. Para realizar esta división, se calculan 191 descriptores 2D y 3D con MOE 2005.06 a partir de los compuestos de la base de datos alineados con el AMD3100 en la conformación que adopta en el sitio de unión con el correceptor CXCR4, se calculan las componentes principales que expliquen un 98% de la varianza y éstas se utilizan para agrupar los compuestos en 11 *clusters*. Se selecciona un compuesto de cada *cluster* para formar el *test set*, ya que se recomienda utilizar entre un 20% y un 30% de la base de datos para construir este subconjunto,^[16] mientras que el *training set* se compone del resto de moléculas de la base de datos. Un compuesto se considera activo cuando su EC_{50} es inferior a 25 $\mu\text{g/mL}$ y se le asigna la clase 1. En cambio, cuando el EC_{50} es superior a 25 $\mu\text{g/mL}$ (que corresponde a la concentración máxima utilizada en los ensayos de actividad anti-VIH) se considera que el compuesto es inactivo y se le asigna la clase 0 (Figura 3.32).

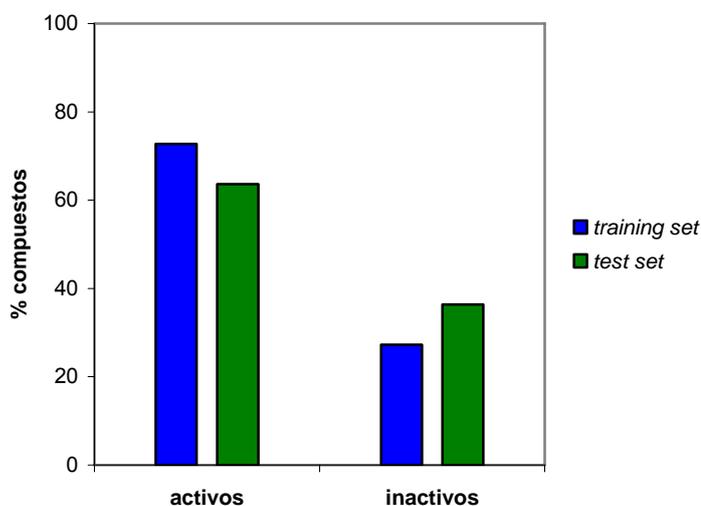


Figura 3.32. Distribución de los compuestos de la base de datos según la clase a la que pertenecen (activos o inactivos) del *training set* (azul) y del *test set* (verde).

Después de realizar la división de la base de datos el *training set* queda constituido por las moléculas: **33**{1,1}, **37**{1,5}, **37**{1,6}, **37**{1,11}, **37**{2,4}, **37**{2,5}, **37**{2,6}, **37**{2,9}, **33**{3,3}, **37**{3,6}, **37**{3,8}, **37**{3,11}, **32**{4,4}, **32**{4,5}, **32**{4,6}, **32**{4,9}, **32**{4,11}, **32**{5,5}, **32**{5,8}, **32**{5,9}, **32**{5,10}, **32**{5,11}, **32**{6,7}, **32**{6,8}, **32**{6,9}, **32**{6,11}, **32**{7,9}, **32**{7,11}, **32**{8,8}, **32**{8,9},

32{8,10}, 32{9,9}, 32{9,11}, 32{10,11}, 32{11,11}, 36{1}, 36{2}, 35{4}, 35{7}, 35{8}, 35{9}, 35{10}, 35{11} y AMD3100 (Anexo 4).

Mientras, el *test set* está formado por: **33{2,2}, 37{2,11}, 37{3,5}, 37{3,9}, 32{5,6}, 32{5,7}, 32{6,10}, 32{8,11}, 32{9,10}, 35{5} y 35{6}** (Anexo 5).

3.7.4. Modelos de clasificación.

Para construir los diferentes modelos se utiliza el módulo *QuaSAR-Classify* de MOE2005.06,^[17] y calculan 191 descriptores 2D y 3D (Anexo 1), se seleccionan y eliminan los descriptores mediante un proceso iterativo de construcción de modelos y evaluación del índice de clasificación incorrecta para el *training set* y para el *test set*. Los mejores modelos obtenidos se describen en la Tabla 3.3. El árbol de clasificación 7 rinde el menor valor de clasificación incorrecta del *training set*, mientras que el árbol 16 rinde el menor valor de clasificación incorrecta del *test set*. Además, se observa que el modelo que clasifica correctamente la mayor proporción de compuestos es el árbol 7, con un valor de la exactitud de 0.98. Por otra parte, el mayor número de compuestos activos correctamente clasificados se obtiene en los árboles 5, 6 y 16, cuya sensibilidad es 1. El análisis de la especificidad muestra que el valor máximo, 1, se obtiene para los modelos 3, 7 y 14 que clasifican correctamente el mayor número de compuestos inactivos. El valor más elevado de la precisión, 1, se obtiene para modelos 3, 7 y 14. Por último, la comparación de los modelos con el coeficiente de correlación de Matthews, indica que la clasificación del *training set* realizada con el árbol 7 presenta la mejor calidad (con un valor de 0.94), seguida de las clasificaciones realizadas con los árboles 5, 6 y 16 (con un valor de 0.90). Globalmente, el modelo que rinde el mejor resultado es el árbol de clasificación 7, sin embargo, su índice de clasificación incorrecta de los compuestos del *test set* es de 0.36, mayor que en otros modelos. El análisis del modelo 8 muestra que este último presenta un comportamiento análogo al modelo 7. Finalmente se selecciona el árbol de clasificación 16, ya que es el que presenta el menor índice de clasificación incorrecta tanto para el *training set* como para el *test set*, además de ser uno de los modelos que clasifica correctamente una mayor proporción de compuestos, especialmente los compuestos activos, además de presentar una calidad elevada según el coeficiente de correlación de Matthews.

Tabla 3.3. Árboles de clasificación, sus descriptores, índices de clasificación incorrecta, $R(T)$, para el modelo y para el test set, y parámetros de exactitud, X , selectividad, S , especificidad, E , precisión, P , y el coeficiente de correlación de Matthews. Los mejores valores de cada parámetro se resaltan en negra.

Modelo	Descriptores	$R(T)$	X	S	E	P	MCC
1	<i>a_ICM</i> <i>PEOE_VSA_POL</i> <i>E_nb</i>	$R(T)=0.068$ $R(T)_{test}=\mathbf{0.182}$	0.93	0.97	0.85	0.94	0.83
2	<i>petitjeanSC</i> <i>chi1v_C</i> <i>PEOE_VSA-6</i>	$R(T)=0.091$ $R(T)_{test}=0.455$	0.91	0.91	0.90	0.97	0.76
3	<i>chi1v_C</i> <i>PEOE_VSA-6</i> <i>SlogP_VSA1</i>	$R(T)=\mathbf{0.045}$ $R(T)_{test}=0.273$	0.95	0.94	1.00	1.00	0.89
4	<i>petitjeanSC</i> <i>chi1v_C</i> <i>glob</i>	$R(T)=0.091$ $R(T)_{test}=0.273$	0.91	0.97	0.79	0.91	0.79
5	<i>chi1v_C</i> <i>PEOE_VSA_POL</i> <i>glob</i>	$R(T)=\mathbf{0.045}$ $R(T)_{test}=0.273$	0.95	1.00	0.86	0.94	0.90
6	<i>chi1v_C</i> <i>PEOE_VSA-6</i> <i>PEOE_VSA_POL</i> <i>glob</i>	$R(T)=\mathbf{0.045}$ $R(T)_{test}=0.273$	0.95	1.00	0.86	0.94	0.90
7	<i>chi1v_C</i> <i>PEOE_VSA-6</i> <i>ASA_H</i>	$R(T)=\mathbf{0.023}$ $R(T)_{test}=0.364$	0.98	0.97	1.00	1.00	0.94
8	<i>chi1v_C</i> <i>PEOE_VSA-6</i> <i>ASA_H</i>	$R(T)=\mathbf{0.045}$ $R(T)_{test}=0.364$	0.95	0.97	0.92	0.97	0.89
9	<i>petitjeanSC</i> <i>chi1v_C</i> <i>ASA_H</i>	$R(T)=0.091$ $R(T)_{test}=0.364$	0.91	0.91	0.90	0.97	0.76
10	<i>a_aro</i> <i>ASA_H</i>	$R(T)=0.091$ $R(T)_{test}=0.455$	0.91	0.94	0.83	0.94	0.77
11	<i>ASA_H</i> <i>b_rotR</i>	$R(T)=0.091$ $R(T)_{test}=0.364$	0.91	0.91	0.90	0.97	0.76
12	<i>ASA_H</i> <i>PEOE_VSA-4</i>	$R(T)=0.091$ $R(T)_{test}=0.273$	0.91	1.00	0.75	0.88	0.81
13	<i>ASA_H</i> <i>PEOE_VSA-6</i>	$R(T)=\mathbf{0.045}$ $R(T)_{test}=0.364$	0.95	0.97	0.92	0.97	0.89
14	<i>E_oop</i> <i>ASA_H</i>	$R(T)=0.068$ $R(T)_{test}=0.455$	0.93	0.91	1.00	1.00	0.83

Modelo	Descriptoros	$R(T)$	X	S	E	P	MCC
15	ASA_H PEOE_VSA-4 PEOE_VSA-6	$R(T)=0.068$ $R(T)_{test}=0.182$	0.93	0.97	0.85	0.94	0.83
16	PEOE_VSA-4 PEOE_VSA-6 SlogP_VSA1 ASA_H	$R(T)=0.045$ $R(T)_{test}=0.091$	0.95	1.00	0.86	0.94	0.90

El árbol de clasificación 16 está compuesto por 13 nodos, una profundidad de 4 niveles (número máximo de decisiones que se han de tomar para clasificar un compuesto) y utiliza los descriptoros *PEOE_VSA-4*, *PEOE_VSA-5*, *SlogP_VSA1* y *ASA_H* (Figura 3.33):

- *PEOE_VSA-4* es la suma del área aproximada de la superficie accesible de van der Waals para los átomos cuya carga parcial calculada mediante el método PEOE^[28,29] está entre -0.25 y -0.20.
- *PEOE_VSA-6* es igual que el anterior pero para los átomos cuya carga sea menor que -0.30.
- *SlogP_VSA1* es la suma del área aproximada de la superficie accesible de van der Waals para los átomos cuya contribución al $\log P(o/w)$ está entre -0.4 y -0.2.
- *ASA_H* es el área de la superficie accesible al agua de los átomos hidrofóbicos cuyo valor absoluto de la carga parcial es menor que 0.2.

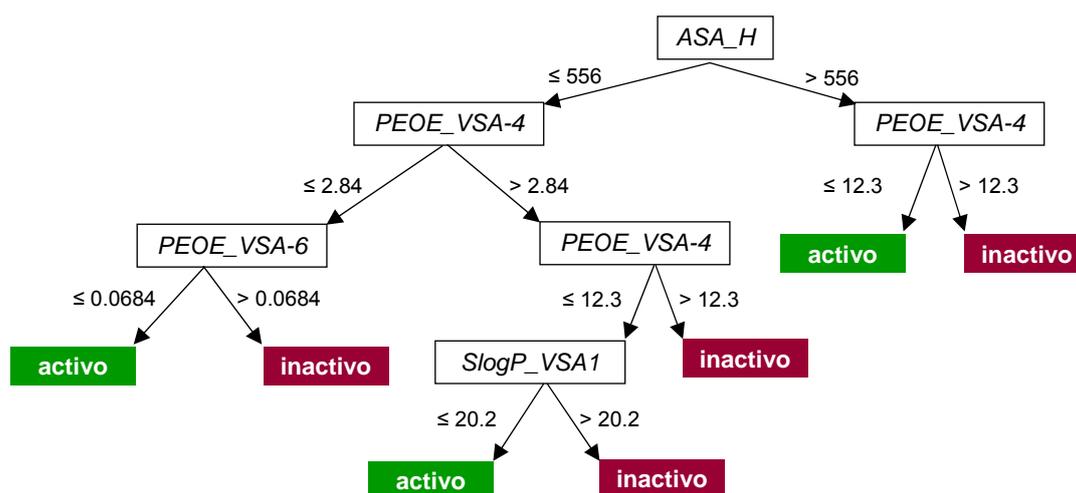


Figura 3.33. Esquema del árbol de clasificación 16, que ha rendido el menor índice de clasificación incorrecta, $R(T)$, tanto para el training set como para el test set.

El modelo de clasificación 16 seleccionado solamente asigna de manera incorrecta la clase de dos compuestos activos del *training set*, **37**{3,6} y **32**{5,10}, cuyos valores de EC_{50} son 11.2 y 2.4 $\mu\text{g/mL}$ respectivamente, y de un compuesto inactivo del *test set*, **35**{6} (Figura 3.34). Por lo tanto, la precisión del árbol de clasificación es muy elevado para ambas bases de datos, concretamente 0.94 en el caso del *training set* y 1 en el caso del *test set* (Tabla 3.4).

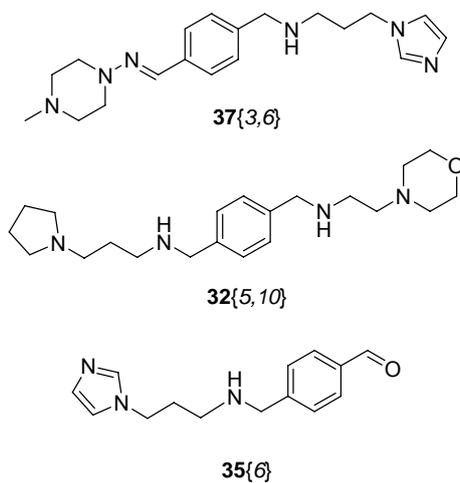


Figura 3.34. Estructura de los compuestos clasificados incorrectamente con el árbol 16.

3.8. Predicciones.

Se predice la actividad anti-VIH de los compuestos no sintetizados incluidos en el diseño de la quimioteca de los Capítulos 1 y 2.

En primer lugar se debe preparar la base de datos de compuestos para los que se quiere realizar las predicciones de actividad. El primer paso es protonar estos compuestos a pH fisiológico, optimizar la geometría y calcular las cargas parciales con el *force field* MMFF94. Después se alinean cada uno de estos compuestos con el AMD3100 en la conformación que adopta en el sitio de unión con el correceptor de entrada CXCR4 y se escoge aquella que presenta el mejor alineamiento, es decir, el menor valor de S (Ecuación 3.33). Finalmente se calculan los descriptores a partir de la conformación seleccionada.

A continuación se realizan las predicciones de actividad. En primer lugar se utiliza el árbol de clasificación 16 para discernir los compuestos activos de los inactivos y seguidamente se utiliza el modelo QSAR 2 para predecir el valor de la actividad anti-VIH (Tabla 3.4). Sólo se debería predecir el valor de la actividad con los modelos QSAR desarrollados en el caso de los compuestos que han resultado activos en la predicción realizada con el árbol de clasificación, ya que este último se ha construido a partir de compuestos activos e inactivos, mientras que los modelos QSAR se han construido solamente a partir de compuestos activos. Por lo tanto la fiabilidad en la predicción de inactivos es mayor en el modelo de clasificación que en el modelo QSAR. Sin embargo, las predicciones del valor de la actividad también se han realizado para los compuestos clasificados como inactivos para poder comparar los resultados de ambos modelos. De esta manera se han observado resultados diferentes para los compuestos **37**{1,4}, predicho como inactivo con el árbol de clasificación mientras que el modelo QSAR lo predice como activo con un valor de EC₅₀ de 2.94 µg/mL; **33**{2,3}, también predicho como inactivo con el árbol de clasificación mientras que el modelo QSAR predice un valor de EC₅₀ de 3.60 µg/mL; **32**{7,10} cuya predicción del valor de EC₅₀ utilizando el modelo QSAR es de 3.61 µg/mL mientras que el árbol lo clasifica como inactivo; y **37**{3,10} que es identificado como activo con el árbol de clasificación mientras que el modelo QSAR predice un EC₅₀ superior a 20 µg/mL.

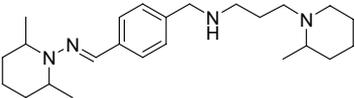
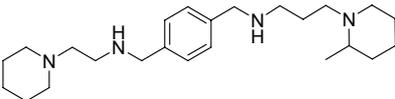
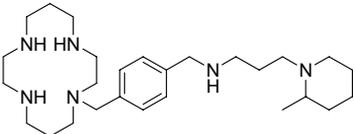
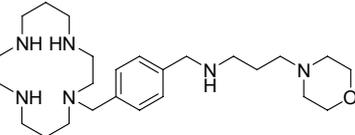
Tabla 3.4. Predicción de la actividad anti-VIH de los compuestos no sintetizados de la quimioteca definida en los Capítulos 1 y 2 utilizando el árbol de clasificación 16 (distinción entre activos e inactivos) y el modelo QSAR 2 (cálculo del EC_{50}). El EC_{50} de los compuestos clasificados como inactivos se muestra en cursiva. Los tres compuestos que han resultado más activos en la predicción se resaltan en negrita.

Compuesto	Clase	EC_{50} / μM	EC_{50} / $\mu\text{g}\cdot\text{mL}^{-1}$
33{1,2}	1	2.24	0.73
33{1,3}	0	48.09	15.12
37{1,4}	0	9.28	2.94
37{1,7}	1	5.98	1.97
37{1,8}	1	3.70	1.33
37{1,10}	0	30.44	10.09
33{2,3}	0	10.52	3.60
37{2,7}	1	1.88	0.67
37{2,8}	1	1.58	0.61
37{2,10}	1	10.13	3.64
37{3,7}	0	18.96	6.55
37{3,10}	1	58.34	20.27
32{4,7}	1	5.94	2.06
32{4,8}	1	2.16	0.81
32{4,10}	0	17.34	6.05
32{7,8}	1	0.66	0.26
32{7,10}	0	9.95	3.61
70{1}	1	5.90	2.38
70{2}	1	3.85	1.66
70{3}	1	24.95	10.45
70{4}	1	4.75	2.00
70{5}	1	5.23	2.72
70{6}	1	2.87	1.24
70{7}	1	2.70	1.17
70{8}	1	0.87	0.40
70{9}	1	4.61	2.14
70{10}	1	19.70	8.58
70{11}	1	7.30	3.29

En general, los compuestos clasificados como inactivos son los que han presentado los mayores valores de EC_{50} , en todos los casos han sido superiores a 9 μM . También, se puede observar que los tres compuestos más activos en la predicción de la actividad anti-VIH son 37{2,8}, 32{7,8} y 70{8}, cuyos valores de EC_{50} son de 1.58, 0.66 y 0.87 μM respectivamente.

La síntesis de estos compuestos ha permitido determinar la actividad anti-VIH real, cuyo resultado expresado como EC_{50} es 1.56 μM para **37**{2,8}, 1.03 μM para **32**{7,8} y 0.03 μM para **70**{8}. El monociclamo **70**{11} también se ha sintetizado y se ha determinado su actividad anti-VIH, de manera que se ha podido comparar el valor obtenido experimentalmente, 0.08 μM , con el calculado mediante el modelo QSAR 2, 7.30 μM (Tabla 3.5). La gran diferencia entre el valor experimental y el calculado para los dos monociclamos probablemente se debe a que el *training set* utilizado en el cálculo de los modelos QSAR sólo incluye un compuesto de la familia de los ciclamos, el AMD3100. Por lo tanto, se demuestra la necesidad de desarrollar otros modelos o utilizar otras técnicas de cribado virtual que incluyan esta familia de inhibidores de CXCR4 en el *training set*. Los resultados obtenidos para los compuestos **37**{2,8} y **32**{7,8} demuestran la utilidad de estos modelos ya que los valores de EC_{50} obtenidos experimentalmente y con el modelo son del mismo orden de magnitud, cabe destacar que la diferencia entre la actividad experimental y la calculada con el modelo QSAR 2 para el primer compuesto es sólo de 0.02 μM .

Tabla 3.5. Comparación del valor experimental y calculado con el modelo QSAR 2 de EC_{50} en μM para los tres compuestos más activos en la predicción con dicho modelo, **37**{2,8}, **32**{7,8}, **70**{8}, y el monociclamo **70**{11}.

Compuesto	EC_{50} experimental	EC_{50} calculado
37 {2,8} 	1.56	1.58
32 {7,8} 	1.03	0.66
70 {8} 	0.03	0.87
70 {11} 	0.08	7.30

Bibliografía.

- [1] C. D. Selassie; History of quantitative structure-activity relationships. In *Burger's medicinal chemistry and drug discovery*, 6th edition; D. J. Abraham Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, **2003**; Volume 1: Drug discovery, 1-48.
- [2] L. P. Hammett; Some relations between reaction rates and equilibrium constants. *Chem. Rev.* **1935**, *17*, 125-136.
- [3] R. W. Taft; Polar and steric substituent constants for aliphatic and o-benzoate groups from rates of esterification and hydrolysis of esters. *J. Am. Chem. Soc.* **1952**, *74*, 3120-3128.
- [4] C. Hansch, P. P. Maloney, T. Fujita, R. M. Muir; Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. *Nature* **1962**, *194*, 178-180.
- [5] T. Fujita, J. Iwasa, C. Hansch; A new substituent constant, π , derived from partition coefficients. *J. Am. Chem. Soc.* **1964**, *86*, 5175-5180.
- [6] C. Hansch; Quantitative approach to biochemical structure-activity relationships. *Acc. Chem. Res.* **1969**, *2*, 232-239.
- [7] S. M. Free, J. W. Wilson; A mathematical contribution to structure-activity studies. *J. Med. Chem.* **1964**, *7*, 395-399.
- [8] J. W. McFarland; Parabolic relation between drug potency and hydrophobicity. *J. Med. Chem.* **1970**, *13*, 1192-1196.
- [9] H. Kubinyi; Quantitative structure-activity relations. 7. The bilinear model, a new model for nonlinear dependence of biological activity on hydrophobic character. *J. Med. Chem.* **1977**, *20*, 625-629.
- [10] R. D. Cramer, D. E. Patterson, J. D. Bunce; Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959-5967.
- [11] G. Klebe, U. Abraham, T. Mietzner; Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **1994**, *37*, 4130-4146.
- [12] A. Golbraikh, M. Shen, Z. Xiao, Y. Xiao, K. Lee; Rational selection of training and test sets for the development of validated QSAR models. *J. Comput. Aided Mol. Des.* **2003**, *17*, 241-253.
- [13] S. Zhang, L. Wei, K. Bastow, W. Zheng, A. Brossi, K. Lee, A. Tropsha; Antitumor agents 252. Application of validated QSAR models to database mining: discovery of novel tylophorine derivatives as potential anticancer agents. *J. Comput. Aided Mol. Des.* **2007**, *21*, 97-112.
- [14] V. K. Agrawal, J. Singh, M. Gupta, Y. A. Jaliwala, P. V. Khadikar, C. T. Supuran; QSAR studies on bezopyran potassium channel activators. *Eur. J. Med. Chem.* **2006**, *41*, 360-366.
- [15] A. Golbraikh, A. Tropsha; Beware of q²! *J. Mol. Graph. Model.* **2002**, *20*, 269-276.

- [16] A. Yasri, D. Hartsough; Toward an optimal procedure for variable selection and QSAR model building. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1218-1227.
- [17] MOE (Molecular Operating Environment) 2006.08 Release; Chemical Computing Group, Inc.: Montreal, Canada, **2004**.
- [18] V. I. Pérez-Nueno, D. W. Ritchie, O. Rabal, R. Pascual, J. I. Borrell, J. Teixidó; Comparison of ligand-based and receptor-based virtual screening of HIV entry inhibitors for the CXCR4 and CCR5 receptors using 3D ligand shape matching and ligand-receptor docking. *J. Chem. Inf. Model.* **2008**, *48*, 509-533.
- [19] ChemSketch ACD pK_a 8.02; Advanced Chemistry Development, Inc.: Toronto, Canada, **2006**.
- [20] P. Labute, C. Williams, M. Feher, E. Sourial, J. M. Schmidt; Flexible alignment of small molecules. *J. Med. Chem.* **2001**, *44*, 1483-1490.
- [21] S. A. Wildman, G. M. Crippen; Prediction of physiochemical parameters by atomic contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868-873.
- [22] T. A. Halgren; Merck molecular force field. I. Basis, form, scope, parametrization, and performance of MMRR94. *J. Comp. Chem.* **1998**, *17*, 490-519.
- [23] L. H. Hall, L. B. Kier; The nature of structure-activity relationships and their relation to molecular connectivity. *Eur. J. Med. Chem.* **1977**, *12*, 307-312.
- [24] L. B. Kier; A shape index from molecular graphs. *Quant. Struct. -Act. Relat.* **1985**, *4*, 109-116.
- [25] A. T. Balaban; Five new topological indices for the branching of tree-like graphs. *Theor. Chim. Acta* **1979**, *53*, 355-375.
- [26] A. T. Balaban; Highly discriminating distance-based topological index. *Chem. Phys. Lett.* **1982**, *89*, 399-404.
- [27] M. Petitjean; Applications of the radius-diameter diagram to the classification of topological and geometrical shapes of chemical compounds. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 331-337.
- [28] J. Gasteiger, M. Marsili; Iterative partial equalization of orbital electronegativity. A rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219-3228.
- [29] J. Gasteiger, H. Saller; Calculation of the charge distribution in conjugated systems by a quantification of the resonance concept. *Angew. Chem. Int. Ed. Engl.* **1985**, *24*, 687-689.
- [30] J. A. Koziol, J. Y. Zhang, C. A. Casiano, X. X. Peng, F. D. Shi, A. C. Feng, E. K. L. Chan, E. M. Tan; Recursive partitioning as an approach to selection of immune markers for tumor diagnosis. *Clin. Cancer Res.* **2003**, *9*, 5120-5126.
- [31] N. Baurin, J. C. Mozziconacci, E. Arnoult, P. Chavatte, C. Marot, L. Morin-Allory; 2D QSAR consensus prediction for high-throughput virtual screening. An application to COX-2 inhibition modeling and screening of the NCI database. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 276-285.
- [32] J. W. Godden, J. R. Furr, J. Bajorath; Recursive median partitioning for virtual screening of large databases. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 182-188.

-
- [33] D. K. Jones-Hertzog, P. Mukhopadhyay, C. E. Keefer, S. S. Young; Use of recursive partitioning in the sequential screening of G-protein-coupled receptors. *J. Pharm. Toxicol.* **1999**, *42*, 207-215.
- [34] A. Rusinko, M. W. Farmen, C. G. Lambert, P. L. Brown, S. S. Young; Analysis of a large structure/biological activity data set using recursive partitioning. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1017-1026.
- [35] A. M. van Rhee; Use of recursive forests in the sequential screening process: consensus selection by multiple recursion trees. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 941-948.
- [36] J. Burton, I. Ijjaali, O. Barberan, F. Petitet, D. P. Vercauteren, A. Michel; Recursive partitioning for the prediction of cytochromes P450 2D6 and 1A2 inhibition: importance of the quality of the dataset. *J. Med. Chem.* **2006**, *49*, 6231-6240.
- [37] M. Jacobsson, P. Lidén, E. Stjernschantz, H. Boström, U. Norinder; Improving structure-based virtual screening by multivariate analysis of scoring data. *J. Med. Chem.* **2003**, *46*, 5781-5789.
- [38] P. Baldi, S. Brunak, Y. Chauvin, C. A. F. Andersen, H. Nielsen; Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics* **2000**, *16*, 412-424.

