

Exploring Molecular Diversity: There is Plenty of Room at Markush's

Leticia Manén Freixa

<http://hdl.handle.net/10803/687373>

Data defensa: 22-12-2022

ADVERTIMENT. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

ADVERTENCIA. El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

WARNING. The access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.

DOCTORAL THESIS

Title	Exploring Molecular Diversity: There is Plenty of Room at Markush's
Presented by	Leticia Manén Freixa
Centre	IQS School of Engineering
Department	Organic and Pharmaceutical Chemistry
Directed by	Dr. Jordi Teixidó i Closa and Dr. Roger Estrada Tejedor

*A mis padres,
que siempre han creído en mí.*

*“A veces sentimos que lo que hacemos es tan solo una gota en el mar,
pero el mar sería menos si le faltara una gota”*

Sta. Teresa de Calcuta

ACKNOWLEDGEMENTS

Quien bien me conoce sabe que siempre me repito en demasía con dos conceptos; “perdón” y “gracias”. Creo que nunca ha de faltar tiempo para disculparse y otro para agradecer todo el bien recibido. El primero quizás no tiene cabida en esta tesis (o sí, si ofendo al lector con ciertas afirmaciones), pero para el segundo, me he dejado el final de tesis como mejor momento para extenderme plácidamente. *Abro hilo*.

Primer de tot, voldria agrair als meus directors de tesi, al Dr. Jordi Teixidó i al Dr. Roger Estrada l’oportunitat d’haver pogut realitzar aquesta tesi doctoral amb ells. Gràcies Jordi, per haver confiat en mi, haver sigut promotor d’aquesta idea i haver-me deixat fer-la meva treballant en dos camps terapèutics que m’interpel·len especialment. Gràcies també per aquest camí llarg on m’has acompanyat com un segon pare (tot i que em sap greu, ja saps que a l’altar m’hi portarà el Sr. Manén). Gràcies també Roger per haver-me acollit i guiat durant aquesta aventura, per la paciència infinita que has tingut amb mi (també amb en Jordi i el *bonus combo* Leti-Jordi), i per tenir la porta sempre oberta per aquesta doctoranda inquieta.

Gracias también al Dr. Francisco-Javier Gamo y la Dra. Sonia Moliner, de la unidad de Global Health de GSK, que accedieron a testear biológicamente nuestras 8-aminoquinolinas y han creído en este proyecto tanto como nosotros.

A mis compis *DisMolers*, que *DisMolan* mazo y aceptan mis *leticismos* continuos. Gracias a todos por los momentos de humor y por el tremendo compañerismo del grupo. Gracias Ángel, por haber hecho camino desde el principio de doctorado, por tu alegría y buen hacer. Gracias Dani, porque contigo siento que mi TOC por la simetría y la correcta combinación de colores se siente realizado. Da gusto pensar que hay alguien más perfeccionista que yo a quien puedo recurrir para consejo. ¡Ah! Me olvidaba, gracias por pipetear de manera maravillosa. Mis *Dacomoleculas* te lo agradecen. Gracias Endika (*aka Pythoniso*) por haberme introducido al gran mundo de Python, por haber sido mentor y compañero en este proyecto, dispuesto a atender todas mis interrupciones. PyLINS no sería lo mismo sin ti. Ni los lunes, sin tus lentejas. Gracias Gema por tu bondad y cariño, por nuestras conversaciones que me llenan de paz en los descansos y por aportar otro toque femenino al grupo (que bien lo necesitan estos *frikis*). I gràcies *Miquel-I*, pels *pitibreaks* necessaris, per sacrificar les teves *lumbers* per guanyar el set de pàdel i per mirar-me incòmodament mentre treballa només per despistar-me. Gracias a ellos, con los que he batallado en mi día a día, y a muchos otros que han ido pasando por aquí (intento ordenarlos cronológicamente); Eli, Raúl O., Pau, Patricia, Irina y Giulia. También gracias a nuestros vecinos de Fotoquímica, especialmente a Mireia, que ha estado siempre como una más en nuestro equipo.

Al equipazo de síntesis, que son unos *cracks* organizando *pica-picas*, jugando al *Among Us* y participando en torneos de pádel. Bueno, y haciendo química orgánica, claro, que es “lo nuestro”. Gracias a todos por el buen ambiente, por todas las alegrías compartidas y por la actitud servicial que lidera este grupazo. Gracias Dra. Ana Belén, por todo lo que hemos compartido desde el máster y por estar siempre solícita a echarme una mano en mis dudas sintéticas. Gracias al Dr. Iñaki Borrell, que me abrió las puertas del departamento ya en tiempos del TFG y ha sido un buen consejero dentro de la presente tesis. Gràcies Dr. Albert, Dra. Gemma, Dra. Maia i Dr. Rai (evidentment, en català) per ser referents científics dintre del grup, pels moments de rialles i per compartir la passió *serièfila* tot i que, em sap greu, no heu sumat a ningú a *Anatomia de Grey* (ja fem tard). Gràcies també al Dr. Batllori, per les hores compartides al laboratori d’Orgànica del 3r Curs de grau en Química i Enginyeria Química i per ajudar-me en tots els meus dubtes espectroscòpics i petites incidències amb el RMN. Gracias Ana A., por tu pasión por TODO, por ponerle corazón a todo lo que haces y explicas animando mis tediosas horas delante de la columna. Y, por cierto, magnífica crema de manos ¿Cuándo abres tu línea dermatológica? Gràcies Claudi per fer camí des de el principi amb mi, per tots els cafès (i no cafès, on us he abandonat d’última hora), per la teva la paciència infinita escoltant els meus drames i per sempre posar-hi una rialla als dies foscos. Ets un sol. Tot i que prefereixo no tenir-te com contrincant al pàdel (ja se les tallaràs a una altra). Gracias también Fede por compartir amablemente “EL ROTA” y dejarme la *schlenk* para secar los productos más caprichosos. Gràcies Judith, la *peque*, per fer-ho tot amb alegria i per ser companya de clorur de tionil. Espero que li donis bon ús al material que t’he deixat en herència. Gracias Raúl C., por compartir tu sabiduría y proponer alternativas cada vez que tenía mis *minidramas* sintéticos. Y gracias a Victor porque, aún y nuestra amistad peculiar parecida a una montaña rusa, al final hemos llegado a buen puerto y, oye, creo que hacemos buen equipo. Sobre todo, cuando se trata de ponerle *salseo* al día a día. Y, aunque no esté ya en el equipo, gracias a la Dra. Ana M^a, la *postdoc*, que ha sido muy clave en la parte sintética de esta tesis, mentora, amiga y mejor investigadora. Gracias a todos ellos y los que también han ido entrando y saliendo por síntesis (menos a los que se llevan ordenadores).

Gracias a todo el Grupo de Química Farmacéutica, en general, y otros miembros de IQS que han caminado conmigo durante esta tesis o tiempos anteriores; al Dr. Borrós, Robert T., Iris P., Carlos E., Paula R., Luis B., Ari C., Chema, David, P. Llorenç, Pitti, Oriol Q., Victoria C., Laura F., al equipo de cafetería (Núria, Àlex, Ricard, Cesc) entre otros muchos.

A mis amigos, que me han apoyado y animado durante esta carrera profesional. Gracias a Carlota, que siempre ha estado al pie del cañón, animándome a dar lo mejor de mí y apoyándome aún y nuestra reclusión pandémica (COVID-19) en Santjoanistes 16. A mi *troupe*; Antonio, Carla, Gerar, Jose, Paula y Quique, por apoyarme y ayudarme a airearme algún fin de semana más que necesario. Gracias a mis *haters*, Anna, Betty y Pablo, que no dejaron de sorprenderme durante este proceso y fueron unos grandes *cheerleaders* hasta en los momentos más oscuros.

Me guardo para el final uno de los agradecimientos más significativos. Gracias, papá y mamá, porque siempre habéis creído en mí y habéis apostado por mi educación. Siempre os habéis implicado en mi crecimiento integral y me habéis ayudado a poner alegría y positivismo en cada etapa vital. Gracias, papá (*el intenso*) por tu continuo humor y por impulsarme en esta carrera final y desestresarme con nuestras escapadas ciclistas. Gracias a mamá, porque sin ti no estaría aquí, porque me has inculcado la cultura del esfuerzo, la inquietud por aprender y siempre me has acompañado desde un inmenso cariño. Y gracias a mi hermano Santiago (*aka Yago*), el primer Dr. Manén Freixa de la familia, por ser referente, por relativizar los malos momentos y acompañarme en esta aventura desde tu magnífica experiencia. *Os quiero.*

Y gracias a José Luis (*aka nutrio*), que es el que más cercanamente ha vivido esta etapa de tesis, con sus luces y sus sombras. Gracias por ser un magnífico compañero y apoyo. Aunque en ocasiones seas un gruñón. Somos un *tándem* perfecto. *Te quierosect.*

SUMMARY

The early Drug Discovery strategy is commonly based on a hit-to-lead process which involves large research on the synthesis of derivatives of an original molecule that had previously shown biological activity against a specific biological target. Therefore, this process implies the synthesis of many analogs leading to the description of a chemical sub-library which generally leads to a highly focused study on the chemical space nearby the hit compound. However, when this drug is finally patented, a wider chemical space derived from a Markush structure is described, theorizing that some analogs within may present biological activity. Nevertheless, this claim involving the Markush structure does not imply the proven synthesis of all the chemical library but just a small population of it.

We hypothesize that there is a great part of the chemical space of these libraries that is unexplored and can hide potential lead candidates which may even surpass the activity of the original hit. Through this project, an alternative is proposed claiming that a rational selection of a short sample of small molecules – founded on similarity-based clustering – can represent more significantly the stated chemical space offering the possibility to explore the unknown space that could hide more potential biological activity.

After a review on the latest approved drugs by the FDA in the period from 2008 to 2020 and the ChEMBL database of bioactive molecules, an exploration of the resulting wide chemical space of small molecules with drug-like properties has been assessed in order to define accessible spots that might hide biological activity. The obtained results from seven real cases of study have proven that random and rationally selected molecules represent more significantly the combinatorial libraries stated in the patents rather than the reported molecules until date.

Furthermore, two practical studies implementing our suggested methodology have been developed to better describe the chemical space of the antimalarial drug Tafenoquine and Dacomitinib, a second-generation tyrosine kinase inhibitor for non-small-cell lung cancer treatment. The assessment driven by a better chemical space exploration of these two families have led to the rational synthesis of seven antimalarial analogs and eight kinase inhibitors which have shown interesting inhibitory activities.

Our results evince that the application of cheminformatics for library selection may improve the ability to better inspect chemical datasets in order to identify new potential hits and represent large libraries for further reprofiling purposes.

SUMARIO

La estrategia de las etapas iniciales del descubrimiento de fármacos está normalmente basada en un proceso denominado *hit-to-lead* que implica un extenso estudio entorno a la síntesis de derivados de una molécula original que previamente haya expresado cierta actividad biológica frente a una diana concreta. Por ende, este proceso conlleva la síntesis de muchos análogos que describirían una sublibrería química, la cual generalmente evidencia que estos estudios están muy focalizados alrededor del espacio químico del compuesto original. Aún y así, cuando esta molécula es finalmente patentada, se describe un espacio químico mucho más vasto por medio de estructuras Markush teorizando que algunos de sus derivados puedan presentar también actividad biológica. Sin embargo, la presencia de estas estructuras no implica la síntesis comprobada de toda la biblioteca molecular sino solo una pequeña muestra de la misma.

Nuestra hipótesis es que hay una gran parte del espacio químico de estas bibliotecas que está sin explorar y puede ocultar posibles candidatos que pueden hasta superar la actividad del *hit* original. A través de este proyecto, se propone una alternativa que sostiene que una selección racional de pocas moléculas – fundada en el agrupamiento según su similitud química – puede representar de manera más significativa el espacio químico establecido, ofreciendo la posibilidad de explorar regiones desconocidas que podrían ocultar más potencial biológico.

Después de revisar los últimos fármacos aprobados por la FDA en el período de 2008 a 2020 y la base de datos de moléculas bioactivas de ChEMBL, se ha llevado a cabo una exploración del amplio espacio químico resultante de moléculas pequeñas con propiedades similares a las de los medicamentos para definir nuevos espacios accesible que podrían ocultar actividad. Los resultados obtenidos de siete casos de estudios reales han demostrado que tanto la selección racional como la aleatoria representan más significativamente las bibliotecas combinatorias declaradas en las patentes que las moléculas descritas hasta la fecha.

Se han desarrollado dos estudios prácticos que implementan esta metodología sugerida para describir mejor el espacio químico del fármaco antipalúdico Tafenoquina y Dacomitinib, un inhibidor de la tirosina quinasa de segunda generación para el tratamiento del cáncer de pulmón de células no pequeñas. La exploración del espacio químico de estas dos familias ha llevado a la síntesis racional de siete análogos antipalúdicos y ocho inhibidores de quinasas que han mostrado interesantes actividades inhibitorias.

Estos resultados demuestran que la aplicación de la quimioinformática para la selección de bibliotecas puede mejorar la capacidad de inspeccionar mejor los conjuntos de datos químicos para identificar nuevos potenciales *hits* y representar grandes bibliotecas para fines de reposicionamiento.

SUMARI

L'estratègia de les etapes inicials del descobriment de fàrmacs està normalment basada en un procés anomenat *hit-to-lead* que implica un extens estudi entorn de la síntesi de derivats d'una molècula original que prèviament hagi mostrat certa activitat biològica davant d'una diana concreta. Per tant, aquest procés comporta la síntesi de molts anàlegs que descriurien una subquimioteca, que generalment evidencia que aquests estudis estan molt focalitzats al voltant de l'espai químic del compost original. Així i tot, quan aquesta molècula és finalment patentada, es descriu un espai químic molt més vast per mitjà d'estructures Markush donant per suposat que alguns dels seus derivats puguin presentar també activitat biològica. Tot i això, la presència d'aquestes estructures no implica la síntesi comprovada de tota la biblioteca molecular sinó només una petita mostra de la mateixa.

La nostra hipòtesi és que hi ha una gran part de l'espai químic d'aquestes biblioteques que està sense explorar i pot amagar possibles candidats que poden fins i tot superar l'activitat del *hit* original. A través d'aquest projecte, es proposa una alternativa que sosté que una selecció racional de poques molècules – basat en l'agrupament segons semblança molecular – pot representar de manera més significativa l'espai químic establert, oferint la possibilitat d'explorar regions desconegudes que podrien amagar més potencial biològic.

Després de revisar els darrers fàrmacs aprovats per la FDA en el període del 2008 al 2020 i la base de dades de molècules bioactives de ChEMBL, s'ha dut a terme una exploració de l'ampli espai químic resultant de molècules petites amb propietats similars a les dels medicaments per definir nous espais accessibles que podrien ocultar activitat. Els resultats obtinguts de set casos d'estudis reals han demostrat que tant la selecció racional com l'aleatòria representen més significativament les biblioteques combinatòries declarades a les patents, que les molècules descrites fins ara.

S'han realitzat dos estudis pràctics que implementen aquesta metodologia suggerida per descriure millor l'espai químic del fàrmac antipalúdic Tafenoquina i del Dacomitinib, un inhibidor de tirosina cinases de segona generació per al tractament del càncer de pulmó de cèl·lules no petites. L'exploració de l'espai químic d'aquestes dues famílies ha portat a la síntesi racional de set anàlegs antipalúdics i vuit inhibidors de cinases que han mostrat interessants activitats inhibidores.

Aquests resultats demostren que l'aplicació de la quimioinformàtica per a la selecció de biblioteques pot millorar la capacitat d'inspeccionar millor els conjunts de dades químiques per identificar nous compostos precandidats i representar grans biblioteques per a posteriors campanyes de reposicionament.

LIST OF ABBREVIATIONS

ATR	Attenuated Total Reflection
API	Active Principal Ingredient
BCL	Bibliographic Combinatorial Library
BD	Bibliographic Data
Bibliogr.	Bibliographic
Comb.	Combinatorial
Da	Daltons
EGFR	Epidermal Growth Factor Receptor
EI	Electro ionization
eq.	Equivalents
eV	Electronvolt
FDA	Food and Drug Administration
FW	Free-Wilson
HMBC	Heteronuclear Multiple Bond Correlation
HRC	Hierarchical Agglomerative Clustering
HRMS	High Resolution Mass Spectrometry
HSQC	Heteronuclear Single Quantum Correlation
HTS	High Throughput Screening
IR	Infrared Spectroscopy
J	Coupling constant
k	Number of clusters
KMN	k-means
KMED	k-medoids
mp	Melting point
MCL	Markush Combinatorial Library
MS	Mass Spectrometry
N	Total number of compounds
NSCLC	Non-Small Cell Lung Cancer
O/N	Overnight
PC	Population Coverage
PCs	Principal components
PCA	Principal Component Analysis
Phth.	Phthalimide
QSAR	Quantitative structure-activity relationship

Q-TOF	Quadrupole Time of Flight
Rf	Retention factor
RT	Room temperature
SAHN	Sequential agglomerative hierarchical non-overlapping
SAR	Structure-activity relationship
SC	Space Coverage
SMILES	Simplified Molecular Input Line Entry System
SPC	Supplementary Protection Certificate
TLC	Thin Layer Chromatography
t-SNE	t-Distributed Stochastic Neighbor Embedding
¹ H NMR	Proton nuclear magnetic resonance
¹³ C NMR	Carbon 13 nuclear magnetic resonance
WHO	World Health Organization
4AQ	4-aminoquinolines
8AQ	8-aminoquinolines

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	III
SUMMARY.....	VI
SUMARIO.....	VII
SUMARI.....	VIII
LIST OF ABBREVIATIONS.....	IX
TABLE OF CONTENTS.....	XI
Chapter 1 : Introduction	1
1.1. The Markush dilemma	1
1.1.1. The early stages of Drug Discovery process.....	2
1.1.2. The Free-Wilson R&D strategy.....	8
1.1.3. The product patent and its Markush structure.....	10
1.2. The Chemical Space	13
1.2.1. Molecular similarity	14
1.2.1.1. Molecular descriptors.....	15
1.2.1.2. Similarity coefficients	16
1.2.1.3. Describing the chemical space by dimensionality reduction	17
1.2.2. Exploring the molecular diversity: library curation.....	19
1.2.3. Clustering and partitioning tools	19
1.2.4. Assessment parameters: space and population coverage.....	22
1.3. Previous studies in the research group.....	23
1.4. Hypothesis	24
1.5. Objectives	24
Chapter 2: Rational vs. Traditional R&D methodology.....	26
2.1. Review of the latest drugs approved by the FDA (2008-2020).....	26
2.2. A glance of the <i>drugspace</i> described by ChEMBL.....	28

2.3.	Assessment of the actual chemical space exploration performed in seven cases of study	31
2.3.1.	Diverse drug selection and combinatorial library assembly	32
2.3.2.	Choosing the optimal clustering/partitioning methodology	35
2.3.3.	Assessing the explored space until date	39
2.3.3.1.	The explored regions of ChEMBL's drugspace	39
2.3.3.2.	Bibliographical representativeness in each chemical space.....	40
2.3.3.3.	Comparing BD and BCL coverage among the chemical space described by MCL43	
2.4.	Rational selection; an alternative towards a more efficient hit-to-lead strategy.....	50
2.5.	Discussion and conclusions.....	51

Chapter 3: Tafenoquine. Study of 8-aminoquinolines as potential antimalarial drug candidates..... 53

3.1.	Introduction to malaria and Tafenoquine.....	53
3.1.1.	Malaria	53
3.1.2.	The <i>Plasmodium</i> life cycle.....	54
3.1.3.	<i>P. vivax</i> treatment. The 8-aminoquinoline antimalarial drugs	58
3.2.	Chemical space assessment of Tafenoquine's analogs.....	60
3.2.1.	Tafenoquine library of analogs and its chemical space	60
3.2.2.	Choice of a suitable clustering methodology.....	62
3.2.3.	The explored space known until the date.....	65
3.3.	Rational Selection analysis.....	70
3.3.1.	Database curation; shaping the accessible chemical space.....	72
3.3.2.	Rational selection of Tafenoquine's analogs	75
3.4.	Synthesis of 8-aminoquinolines	76
3.4.1.	General synthetic Route	76
3.4.2.	Preparation of 5,6-dimethoxy-4-methyl-8-nitroquinoline (8)	78
3.4.3.	Preparation of 5-phenoxy-6-methoxy-4-methyl-8-nitroquinoline derivatives (11)	82

3.4.4. Preparation of <i>N</i> -(6-methoxy-4-methyl-5-(phenoxy)quinolin-8-yl)alkyldiamine derivatives (15)	86
3.5. Biological activity	92
3.6. Results discussion	98
3.7. Conclusions	102
Chapter 4: Dacomitinib. Study of substituted quinazoline derivatives as irreversible inhibitors of EGFR	103
4.1. Introduction to Dacomitinib	103
4.1.1. Non-small-cell lung cancer (NSCLC)	103
4.1.2. Tyrosine kinase inhibitors; a key to treat epidermal growth factor receptor (EGFR) and its mutations in NSCL	106
4.1.3. Dacomitinib and substituted quinazoline derivatives as irreversible TKI for NSCLC treatment.....	109
4.2. Chemical space assessment of Dacomitinib's analogs	112
4.2.1. Dacomitinib library of analogs and its chemical space	112
4.2.2. Choice of a suitable clustering methodology	115
4.2.3. The explored space known until the date.....	118
4.3. Rational Selection analysis.....	120
4.3.1. Database curation; shaping the accessible chemical space.....	120
4.3.2. Rational selection of Dacomitinib's analogs	123
4.4. Synthesis of quinazolines.....	126
4.4.1. General synthetic Route	126
4.4.2. Preparation of 7-fluoro-6-nitroquinazolin-4(3 <i>H</i>)-one (33)	127
4.4.3. Preparation of 7-fluoro- <i>N</i> -phenyl-6-nitroquinazolin-4-amines (28)	130
4.4.4. Preparation of 7-substituted- <i>N</i> -phenyl-6-nitroquinazolin-4-amines (29)	134
4.4.5. Preparation of dacomitinib analogs (30)	137
4.4.6. Global synthesis overview	144
4.5. Biological activity	146

4.6.	Results discussion	150
4.7.	Conclusions	157
Chapter 5: Experimental part		158
5.1.	Computational methods; introduction to PyLINS	158
5.1.1.	Dataset assembly and curation.....	160
5.1.1.1.	Combinatorial library enumeration.....	160
5.1.1.2.	Data curation and filtering	160
5.1.2.	Principal Component Analysis reduction	160
5.1.3.	Clustering methodologies	162
5.1.4.	Binning and Optimum Variance binning	166
5.1.5.	Library sizes and selections	169
5.1.5.1.	Bibliographic data search	169
5.1.5.2.	Library sizes	169
5.1.5.3.	Rational selection	170
5.1.6.	Space and Population coverage calculation	170
5.1.7.	t-SNE plots.....	171
5.2.	Synthesis and compound characterization	173
5.2.1.	Instrumentation	173
5.2.2.	Synthesis of 8-aminoquinoline derivatives	174
5.2.2.1.	Synthesis of 1,2-dimethoxy-4,5-dinitrobenzene	174
5.2.2.2.	Synthesis of 4,5-dimethoxy-2-nitroaniline	175
5.2.2.3.	Synthesis of 5,6-dimethoxy-4-methyl-8-nitroquinoline	176
5.2.2.4.	Synthesis of 6-methoxy-4-methyl-8-nitroquinolin-5-ol.....	177
5.2.2.5.	Synthesis of 5-chloro-6-methoxy-4-methyl-8-nitroquinoline	178
5.2.2.6.	Synthesis of 5-phenoxy-6-methoxy-4-methyl-8-nitroquinoline derivatives	179
5.2.2.7.	Synthesis of 5-phenoxy-6-methoxy-4-methylquinoline-8-amine derivatives ...	188
5.2.2.8.	Synthesis of 2-(oxoalkyl)isoindoline-1,3-dione derivatives	196

5.2.2.9. Synthesis of 2-(5-((4-(phenoxy)-3-methoxy-5-methylnaphthalen-1-yl)amino)alkyl)isoindoline-1,3-dione derivatives.....	198
5.2.2.10. Synthesis of <i>N</i> -(6-methoxy-4-methyl-5-(phenoxy)quinolin-8-yl)alkyldiamine derivatives	206
5.2.3. Synthesis of quinazoline derivatives	214
5.2.3.1. Synthesis of 7-fluoroquinazolin-4(3 <i>H</i>)-one	214
5.2.3.2. Synthesis of 7-fluoro-6-nitroquinazolin-4(3 <i>H</i>)-one	215
5.2.3.3. Synthesis of 7-fluoro- <i>N</i> -phenyl-6-nitroquinazolin-4-amines.....	216
5.2.3.4. Synthesis of <i>N</i> -(2-chloro-4-fluorophenyl)-7-(methylthio)-6-nitroquinazolin-4-amine (29{1})	222
5.2.3.5. Synthesis of 7-alkoxy- <i>N</i> -phenyl-6-nitroquinazolin-4-amines (method A)	223
5.2.3.6. Synthesis of 7-alkoxy- <i>N</i> -phenyl-6-nitroquinazolin-4-amines (method B)	227
5.2.3.7. Synthesis of <i>N</i> ⁴ -phenyl-6-nitroquinazolin-4,6-diamines.....	230
5.2.3.8. Synthesis of 7-substituted- <i>N</i> ⁴ -phenyl-6-nitroquinazolin-4,6-diamines	232
5.2.3.9. Synthesis of (<i>E</i>)- <i>N</i> -(4-(phenylamino)-7substituted-quinazolin-6-yl)-4-(Het)but-2-enamide (method A).....	241
5.2.3.10. Synthesis of (<i>E</i>)-4-bromo- <i>N</i> -(4-(phenylamino)-7-substituted-quinazolin-6-yl)but-2-enamide (method B)	247
5.2.3.11. Synthesis of (<i>E</i>)- <i>N</i> -(4-(phenyl)amino)-7substituted-quinazolin-6-yl)-4-(Het)but-2-enamide (method B).....	252
5.3. Determination of the biological activity	256
5.3.1. <i>P. falciparum in vitro</i> growth inhibition assay.....	256
5.3.2. EGFR enzymatic <i>in vitro</i> assay : measure of the residual activity	256
Chapter 6: Conclusions	260
Chapter 7: Bibliography	262
Annexes.....	281

Chapter 1 : Introduction

1.1. The Markush dilemma

In 1923, Dr. Eugene A. Markush (1888-1968), an American manufacturer of dyes and pharmaceuticals, filed a patent application with the United States Patent and Trademark Office (USPTO) claiming a set of organic chemical compounds considering that each of them may not have individually been supported by patent. After a set of rejections where he was objected for terminology vagueness or the use of broad terms, he finally won, in 1925, the Commissioner of Patent's confidence by coining the language that has come to characterize the claim format named for him¹. His US Patent No. 1506316 entitled "Pyrazolone Dye & Process of Making the Same"² was finally granted by a last modification which used the generic description through the expression: "*material selected from the group consisting of*". Since then, the claiming construction in metallurgy, biology and chemistry has been driven by the overuse of this term along with new alternative structures, changing the world of scientific intellectual property.

Hence, Markush *formulae* was originated in the USA with a decision of the Commissioner of Patents: *Ex parte Markush*, 1925 C.D. 126, 340 O.G. 839 (Comm'r Pat. 1924). However, for the surprise of the reader, no generic chemical structure was present in the above-mentioned patent. In the origins of the "Markush claim", only his language won the legal trial becoming the first inventor to use this style of claiming successfully in a US patent.

Ever since, the ritual of creating claims in a more generic manner has become the norm, and the so-called Markush structures now permit applicants to claim more than one chemical entity within one claim. These structures can be described for molecules with substituents at several positions, and frequently many thousands of potential compounds are defined in this way.

However, although the historical origins of these claiming structure advocated that Markush-type claims must be precise and unambiguous, this trend has led to the usage of vast generic aggrupration that defines libraries of thousands and even millions of chemical analogs.

More specifically, in the case of pharmaceutical product patents, broad molecular libraries are claimed but, when looking deeper, in all cases only few compounds have been described.



“It’s not a multiverse. It’s a copyright violation.”

Figure 1.1 Humorous metaphor to illustrate the vastness of a Markush patent claim.
[Web Source: <https://airmail.news/cartoons>]

At this point, two crucial questions may arise: is the patentability the main aim of scientific research, claiming broad unexplored sets of molecules? In fact, where would ideally the scientific community put their effort on: in protecting large unknown and undescribed molecules, or in building rigorous knowledge?

1.1.1. The early stages of Drug Discovery process

To get introduced in the methodology that it is being carried out at present in scientific investigation in the area of research and development (R&D) of new molecular entities (NME) with drug-like properties, a focus in the Drug Discovery process is needed. This comprises how new pharmaceutical compounds are discovered and, by a long-time process with several stages, are finally brought to the market (Figure 1.2).

However, this is not a trivial procedure, as launching a drug into the market actually takes an average of 15 years³, being the rate of success dramatically low – a third part only gets to the market as estimated by a research from McKinsey⁴ – as there are many drug candidates that are discarded in the pre-clinical or clinical trials. In fact, after all these steps, a last assessment is performed by regulatory agencies which gives the last consent to the drug by its approval, enabling them to be sold and consumed in their countries or regions. Moreover, apart from the time-consuming process, a corresponding high economical investment is needed to develop a new drug which has dramatically increased in the last decades. In 2000 the average cost – including capital costs and expenditures on drugs that fail to reach the market – consisted in US\$ 802 million⁵ and, at present, the expected cost to develop a new drug has been estimated to range from less than US\$1 billion to more than US\$2 billion⁶. In brief, when considering developing a new drug it has to be taken into account that the

profit from each successful drug must be large enough to cover not only the costs of the successful drug but also the expenses of other drugs that were cut due to failure during the development.

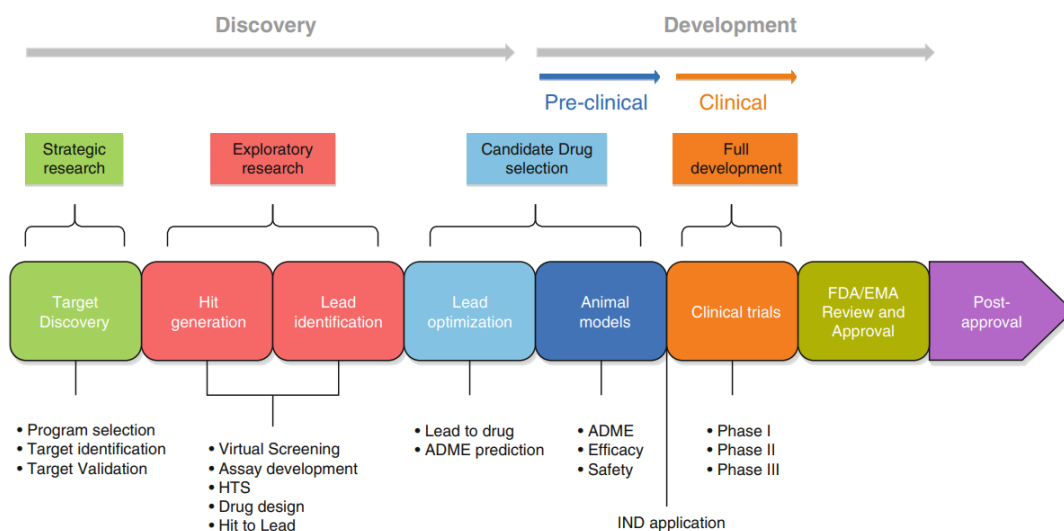


Figure 1.2. Schematic representation of the drug discovery process³.

The early Drug Discovery process consists in all the initial steps preceding the pre-clinical and clinical testing. Three main steps are normally followed: the strategic research, the exploratory research, and the candidate drug selection.

- **Strategic research:**

The first crucial point of the Drug Discovery process departs from choosing the desired area of study. This means which pathology is aimed to be treated – mainly considering the economic benefit and the previous knowledge and expertise by the pharmaceutical enterprise – in order to proceed with the identification and validation of the target that may be the key for disease modification. Due to cost-effective reasons, new drugs are normally developed for diseases and conditions that affect a large number of people or those new categories with potential market which no current treatments exist (such as Alzheimer's disease, obesity, Parkinson's disease, type 1 and 2 diabetes, and asthma) avoiding then, in most of the cases, rare or orphan diseases .

The target is the molecular system where the action of the drug molecule, also known as ligand, takes place controlling a certain biological process related to the sick state of a patient. Medicinal chemistry is the discipline that attempts to discover new drugs that can safely and effectively affect a target. Therefore, knowing the target and its structure is extremely valuable to the discovery team and can lead to a more successful drug search. The classical approach for target discovery is to study the inner mechanisms and signal pathways involved in a disease. Normally,

proteins that play key roles in the processes are candidate targets. In general, those that are promising for drug intervention are often described as *druggable*. Finding a proper target along a pathway requires defining a point of intervention in it that affects the diseased state without significantly impacting other biological processes. In general, in terms of side effects limitation, interacting with target toward the end of a biological pathway (*downstream*) causes fewer side effects than intervening further *upstream*. In fact, for diseases that are fairly well known, considerable information likely exists in the literature which may be supplementary to the previous in-house research performed by the research group area of interest which may provide wealth of proprietary information to facilitate the discovery of new targets.

Traditional drug discovery methods hold firmly to the concept of “*one target, one drug*” defending that the hit of other targets may present *off-target* undesirable side effects. However, the introduction of genomics has challenged this philosophy by revealing the complexity and interdependence of many biological pathways favoring the introduction of the multitargeted approach. This argues that by hitting several related targets weakly, a shorter quantity of drug can affect multiple pathways and give rise to a significant aggregated large biological effect.

- **Exploratory research:**

After choosing the receptor of interest, a screening process takes place. This step can take place *in silico* through computer-aided drug discovery techniques (CADD). In this step, a computational structure-based drug design (SBDD) or ligand-based drug design (LBDD) can be performed depending on the information available around the receptor structure crystallized with an original active ligand or the result of previous biologically tested molecules with potential active candidates or hit molecules (chemical entities with promising affinity for the target). Specialized software use force fields to estimate energies and forces associated with the ligand-receptor complex to identify molecules from large libraries that are likely to bind the target. This identification is carried out simulating ligand-receptor interactions (docking) of every compound in database with all the possible target conformations. The compounds are afterwards ranked from the ones with more affinity to less or even none depending on the given score⁷ (Figure 1.3) .

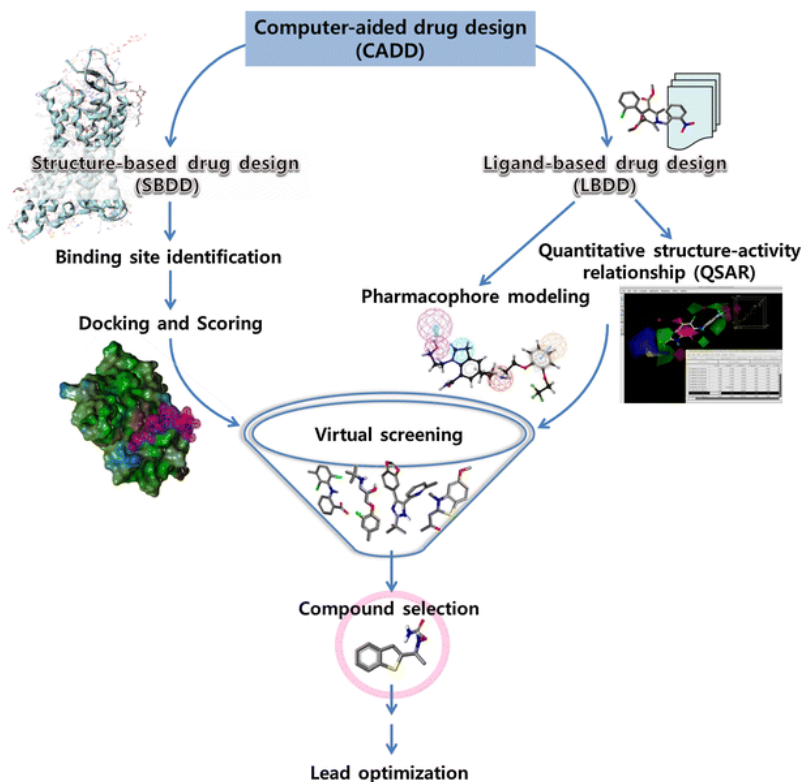


Figure 1.3 Representative workflow for computer-aided drug design from Macalino *et al.*⁸

In the pharmaceutical industry, the High-Throughput Screening (HTS) or Ultra-High-Throughput Screening (uHTS) techniques are also common to take place in parallel during this phase, being able to perform a screening from 500,000 up to 1 million compounds in a single assay⁹. This is an automated process which performs through a robotic equipment a large number of molecules quick biochemical testing screen that provides receptor binding or enzyme inhibition information aiming to find new candidates or a scaffold of interest to be optimized. The set of molecules normally tested correspond to a random screening of the molecular library in the pharmaceutical company's repository or a combinatorial library of interest which chemical similarity fits with the reported literature around the target of study. However, as this first attempt is being tested in isolated receptors, the success of this first approach cannot be taken for granted. A posterior cellular assay is needed to determine how these compounds behave in a more complex cellular environment. This approach has been modified by rational procedures involving a testing of a much smaller number of compounds believed to have potential activity on the basis of the knowledge of the receptor site or the mode of action by previous computer modeling studies involving structure activity relationships among others¹⁰. However, there are still occasions, for instance when the receptor or the mechanism of action is unknown where a largescale empirical testing is appropriate upon initiation for new in

vitro biological screening. In these cases, the chemical library of the big pharma industries is tested in order to proceed with further studies with the active scaffolds or direct drug repositioning (or repurposing). In other words, the HTS screening process serves as a filter on the route to locating a potential drug as the limited number of positive hits are used to direct the additional assays. This procedure implies a large cost of time and sources. As an approximation, if the total screening cost was as low as US 0.4\$ per compound, including the cost of the chemical synthesis, HTS disposables, capital costs and human resources, screening just one target with one million compounds would cost 400,000\$.

Once one or more molecules are identified as possible hits which show a degree of activity beyond a predetermined threshold, a process of drugs design is carried out and more dissimilar candidates may be synthesized to explore new scaffolds.

Finally, a hit to lead process takes place after assessing more factors such as the molecule's ability to cross membranes, cytotoxicity, and metabolism profile. Hence, a lead compound corresponds to the selection of a molecule that, after more accurate investigations, shows a specific and selective binding to the target and is able to modify its normal mechanism of action.

- **Candidate Drug selection:**

The stages of the discovery phase end with the final lead optimization. This step consists in the rational study involving the structural modifications of the selected lead in order to increase its activity and improve *in vivo* properties. This step along with the previous lead discovery phase, is a matter of study for medicinal chemists¹¹. Hence a subset of similar molecules to the original lead is listed, synthesized, and tested. These lead analogs are selected mainly taking in account their synthetical feasibility and the previous *in silico* studies performed. Structure-activity relationship (SAR) studies are normally performed to understand the best method to increase the lead's activity. Medicinal chemists consider in this phase the determination of the pharmacophore of the lead compound or, in other words, the minimal portion of the molecule structure required for significant biological activity⁷. Then, new structural motifs are added and assessed in order to improve the overall activity of the lead and also monitor its pharmacokinetic properties.

In parallel, virtual approaches also take part through an iterative strategy among all this exploratory research process; as more information is collected regarding the biological results obtained from new sets of hit molecules, better trained models can be build-up contributing to a more reality-like interaction study, find compounds with promising biological effect, great ADME (absorption, distribution, metabolism, and excretion) properties, low toxicity (Tox) values, lack of undesired secondary effects and hence improve the strategy in the search of new leads.

Pharmacokinetics studies the body interaction with the administered drug during its exposure which is mainly tracked by the quantitative determination of ADME parameters¹². Both control of ADME and Tox properties takes a key role in this optimization process as these are crucial for the final clinic success of a drug candidate. Indeed, a large number of compounds usually fail the initial development process due to tolerance, lack of bioavailability, toxicity and lack of efficacy of the drug in humans¹³ (Figure 1.5). Nearly 50% of the drugs are estimated to fail due to poor bioavailability and efficacy as a result of an unsuccessful intestinal absorption. Even more, the 40% of the estimated drug candidates normally fail due to toxicity and safety issues¹⁴.

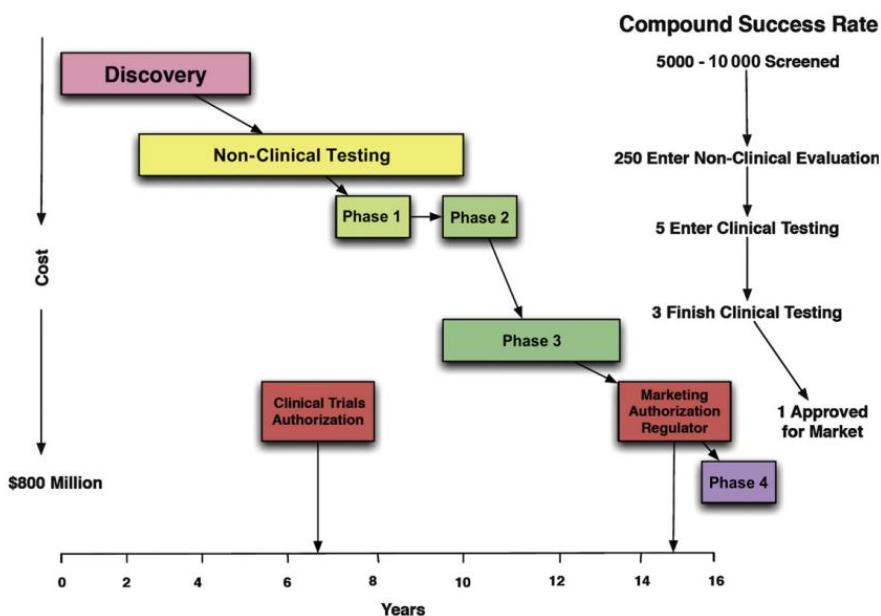


Figure 1.4 Cost and compound success rate during the different steps of drug discovery, development, and marketing authorization. Source: Haschek *et al.*¹³

Finally, once the lead has been effectively optimized and shows desirable properties, it becomes a promising drug candidate to develop and patent (which will be later discussed in section 1.1.3). After the lead found, it starts the final development process involving the animal (pre-clinical) and clinical trials which will track the compound pharmacodynamics, studying the mechanism and duration of action of the drug in the body, and monitoring any undesired side effects or drug-drug interactions. After this process, the drug will hopefully be approved by the regulatory and will be launch to the market. Still, it will be monitored during the years of its distribution and safety thanks to pharmacovigilance.

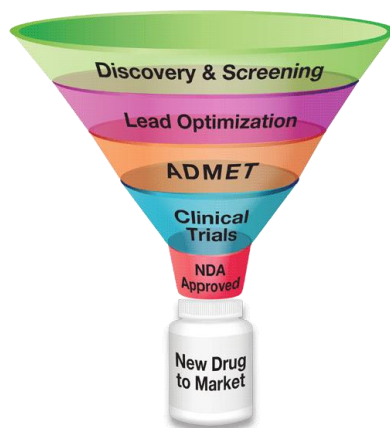


Figure 1.5 Graphical representation of all the succeeded steps until the final drug launch into the market. Source: Enzo Sciences¹⁵

1.1.2. The Free-Wilson R&D strategy

The concurrent design and synthesis of new compounds is a time-consuming and complex task in which the medicinal chemist seeks to balance off-target interactions, biological potency, toxicity, and pharmacokinetic properties. A critical step in the progression of the project is to choose the right compounds to synthesize from a large number of virtual compounds.

The fine-tuning of the lead compounds in the hit-to-lead step previously described departs from establishing the positions of functionalization in order to improve the molecular properties and activity. This phase is crucial to optimize the original hit in terms of activity increase, reduce its side-effects and provide easy and efficient administration to the patient. In other words, remarkable pharmaceutical activities are aimed through the lead search. Structural information on the receptor binding site can be hugely helpful in guiding the medicinal chemistry team regarding which substituents might be most beneficial. In fact, the consideration of the binding roles of some generic groups may be useful to consider if they are involved or not in the ligand-receptor interaction. Some strategies can be considered such as variation of substituents, extension of the structure, chain or ring extension, variation or contraction, isosteres, simplification or rigidification of the structure¹⁶. The most common SAR approach consists in assessing the functional group replacement for boosting the drug activity. These analogs of interest are limited to the ease of molecule preparation (synthetical feasibility) and the reagents availability. The SAR models are actually attracting increasing levels of interest in the pharmaceutical industry as a productive and cost-effective technology in the search for novel lead compounds.

Some authors such as Hansch^{17,18}, Free-Wilson¹⁹, Bocek-Kopecky^{20,21} or Fujita²² are considered the pioneers of SAR and quantitative SAR (QSAR) models by their mathematical contributions in the early 1960's. These QSAR approaches approximate the role between molecular

properties and the resultant biological points of interest in terms of using multivariate statistics. Between the different methods, Hansch, Free-Wilson, and modified Free-Wilson approaches stand out being widely practiced for modelling the biological response^{23–25}.

Specifically, the Free- Wilson (FW) analysis was one of the first mathematical techniques developed for series of chemical analogs. The basic idea in the FW *de novo* approach, also known as additivity model, is that the biological activity of a molecule can be described as the sum of the activity contributions of its specific substructures on a common molecular skeleton. It assumes that the presence or absent of each substituent gives an either additive or suppressive and constant effect to the biological activity regardless of the other subparts attached to the parent core. Thus, the different substituent effects are independent of each other. This can be described by a linear equation [Eq. 1]²⁶ where the regression coefficients show the contribution of every substituent (being S the number of substitutions sites and N_s the substituents per site) to the biological response (γ).

$$\hat{\gamma}_i = b_0 + \sum_{s=1}^S \sum_{k=1}^{N_s} b_{ks} \cdot I_{i,ks} \quad [\text{Eq. 1}]$$

Being b_0 the intercept of the model corresponding to the theoretical biological activity of the plain scaffold or pharmacophore, without any substituent, b_{ks} the regression coefficients and I the indicator variables normally called FW description which states the presence ($I_{i,ks} = 1$) or absence ($I_{i,ks} = 0$) of the substituent of interest in a certain location. Hence, these coefficients b_{ks} give the importance of each k substituent in each s site in terms of increasement or decrease of the response with respect to the unsubstituted structure.

As an explanatory example, Free and Wilson analysed¹⁹ the inhibitory activity against *S. aureus* by the compounds derived from the scaffold showed on Figure 1.6.

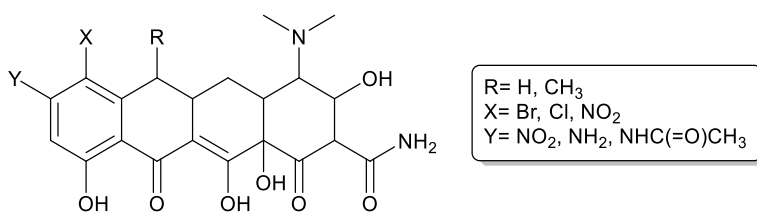


Figure 1.6. Example used by Free and Wilson to illustrate the application of their mathematical model

The enumeration of all the possible combinations of the R, X and Y substituents driven by the rule of product would lead to the creation of a combinatorial library of analogs. In this case, the study contained only eighteen compounds of possible interest ($2 \times 3 \times 3$) from which ten were synthesized, tested and used to train the model for further activity prediction of the resting compounds. In this case, the prediction could be explained by [Eq. 2] which account for 90% of the total variation.

$$\gamma = 75R_H - 112R_{CH_3} + 84X_{Cl} - 16X_{Br} - 26X_{NO_2} + 123Y_{NH_2} + 18Y_{NHC(=O)CH_3} - 218Y_{NO_2} \quad [\text{Eq. 2}]$$

Thus, the analog containing $R = H$, $C = Cl$, and $Y = NH_2$ would be the most promising molecule with a predicted biological response of 444. The actual biological response measured by Free and Wilson (γ) was 525. The authors sustained that the difference was originated in the biological variation of the *in vitro* procedure. However, the aim of this methodology is not to estimate exactly the biological activity but to rank the derivatives in terms of their potential activity.

The use of the FW model have showed fruitful results for many visible models²⁷⁻²⁹ and is highly attractive to chemists as there is a straightforward interpretation and clear structural explanation of the underlying relationships. However, it is limited to a controllable number of variables, compounds and excluding mutually correlated variables is a must³⁰. Still, the main general drawback of QSAR analysis (including FW) is that the prediction scope of these models is normally hampered by the design of the training sets and not explorative prediction can be provided for novel substituents that are not present in the original training set^{23,31,32}.

1.1.3. The product patent and its Markush structure

One of the main interests of pharmaceutical companies is to be profitable in order to stay in business. Therefore, protecting its work through highly restrictive patents is an important part of the process. When a new chemical compound is protected by a patent, the patent ownership can legally exclude others from using, making, selling, offering to sell or importing that compound for the lifetime of the patent which, in most jurisdictions, takes 20 years prom the date of filing. Consequently, compound patents are filed early in the lead optimization phase, a stage where enough knowledge of action of the optimized drug candidate has been constructed in order to invest in its legal protection.

Considering the years involving the drug development process (with its consequent clinical trials and regulatory assessment), the pharmaceutical company has only around twelve years to enjoy exclusive marketing privileges, recover its costs of development, and convert into profit. All things considered; a 3-to-5-year protection can be added through the request of a Supplementary Protection Certificate (SPC) which objective is to compensate some of the effective patent term lost during the development of the drug (Figure 1.7). Finally, once the patent expires, any competitor company may develop the formerly protected active principal ingredient (API) under a generic label.

HOW DOES IT WORK? TWO CASES TO ILLUSTRATE

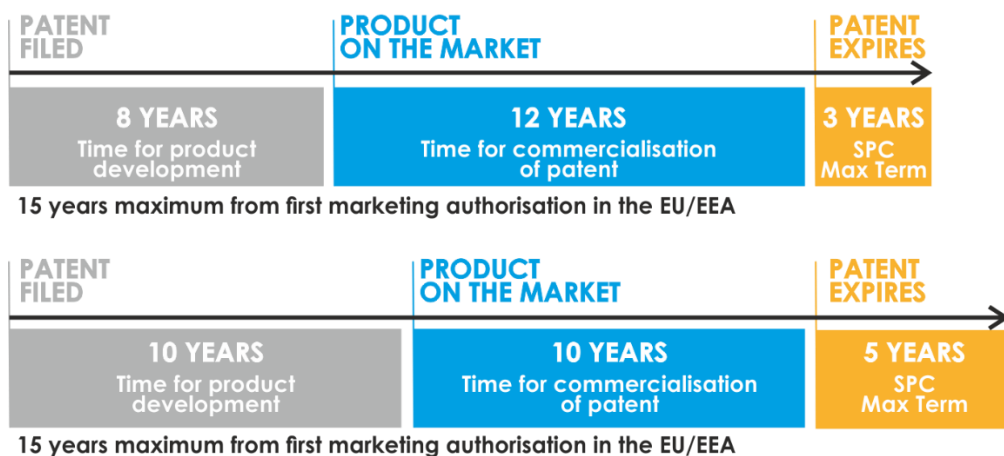


Figure 1.7 Example of patent protection depending on drug development process. Adapted from EFPIA³³

In the case of innovative drugs or therapies, a patent does not constitute a direct authorization to sell a product made with the invention; that is determined by the regulatory agencies through drug approval processes. These are designed to allow safe and effective drugs, combined formulations, or treatments to be marketed. Drug regulatory agencies, such as U.S. Food and Drug Administration (FDA), the European Medicines Evaluation Agency (EMA in Europe) or Pharmaceuticals and Medical Devices Evaluation Center (PMDEA in Japan), attempt to rely on premarketing scientific studies of the effects of drugs both in animals and humans to control if the new candidates have a favorable risk-to-benefit ratio.

At present, there are several types of patent or patent claims that are particularly relevant to pharmaceuticals. Among them, the most relevant are product patents, process patents and formulation patents. In all cases, the requirements for patentability are to consist undoubtedly in a novelty, to imply an inventive action (not an obvious or variant combination of known elements), to present an industrial application and to provide sufficient information in the description.

The common structure of a patent departs from the description of the issue; the title, the approach of the problem and state of the prior art, the general and detailed description of the novelty with examples and drawings. However, the most important part relies in the claims section, which determines the overall extension of protection. There are many sorts of claims depending on the patent type of interest.

In particular, the protection of a new drug candidate (product patent) normally presents the strongest claims such as crystalline forms, new isomers, alloys with different proportions etc. Among

them, it stands out the presence of large Markush structures. These, enable the inventor to protect an entire family of compounds rather than just the single drug candidate he plans to sell.

As introduced on the beginning of this section, this sort of general descriptions by one *formulae* was first used by Dr. Eugene A. Markush, who won a legal case concerning the validity of this type of claims and became the coiner of this sort of structures.

The generic or Markush structure is the primary tool used to condense the structural representation of sets of compounds. This representation used is described, along with the means in which it can be built from a reaction and precursor-based description of a combinatorial library. They are represented as the common core structure of a library of analogs with one or more functional groups attached (often represented by capital letters or R, referring to the term *residue* or *radical*) indicating the existence of variable substituents at that position. In fact, following the illustrative case example shown in Figure 1.8, the variability of these structure can be shown in one or more out of four different ways³⁴:

- Substituent variation: enumeration of a fixed list of alternatives (exemplified as R₁).
- Homology variation: Use of generic expressions which correspond to a potentially unlimited class of substituents, characterized by common structural features (exemplified as R₂).
- Position variation: change of attachment position in the core structure (exemplified as R₃).
- Frequency variation: alternation in multiplicity of occurrence of a group (exemplified as *m*)

These are afterwards described by large lists of specific chemical substructures that may be interchanged in all possible combinations depicting on a single page, libraries that would fill a book if fully enumerated³⁵.

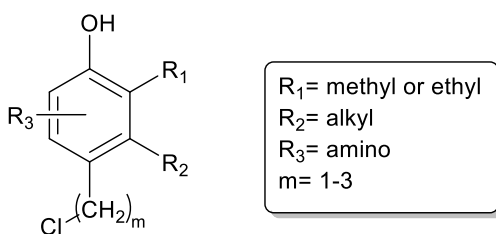


Figure 1.8. Example of Markush structure with all variability examples.

One of the key troubles with Markush structures, beside their inherent complexity and variability, is the representation of generic groups and the matching of these against specific examples of them^{36,37}. Extremely broad Markush claims involving generic categories, multiple ring forming attachments or event correlated sets of substituents tend to cover millions of small molecules with a single basic structure. As an example, the generic group “alkyl”, “heterocycle” or “cycloalkyl” contain

a broader set of analogs than the case of single atom-bon connection tables which would enumerate a few cases like “*n*-propyl” or “cyclopentyl” . In most of the cases, these structures may extend over many pages and contain as large quantity of variable moieties that it becomes hard and even practically impossible to visualize and understand what the applicant is seeking to protect. In fact, this is one of the main objectives while patenting, covering a combinatorial set as much large as possible in order to difficult the search strategy of the competitors. However, in the classical patent structure, as required, only a subset of combinations is present (with the description of the molecules and the procedure to obtain them) being a simplistic example of the scope and knowledge of the claimed library.

1.2. The Chemical Space

The accessibility of collections of thousands and millions of compounds for drug discovery has put forward the concept of chemical space in order to describe the aggrupation of all the molecules^{38,39}. In fact, every chemical database consisting in a set of molecules such as a family of analogs can describe a multi-dimensional chemical space when considering their molecular features. For instance, a combinatorial library derived from a Markush structure or even a diverse set of dissimilar molecules would also shape different chemical spaces (Figure 1.9). To do so, a previous assessment of the suitability of the chosen molecular features is needed in order to sufficiently represent the compounds of analysis. The tendency observed at present by experts is to build a reference chemical space based in large categories of compounds for which molecular properties are known by using methods such as molecular fingerprints or principal component analysis. Later, a projection of the measured features space is projected into a 2D plane or 3D volume by one or several known techniques⁴⁰. Finally, this chemical space is studied to analyze and find similarities among groupings of molecules showing different biological activity values in order to find which regions in the described chemical space of study require to be more deeply explored in order to synthesize new candidates.

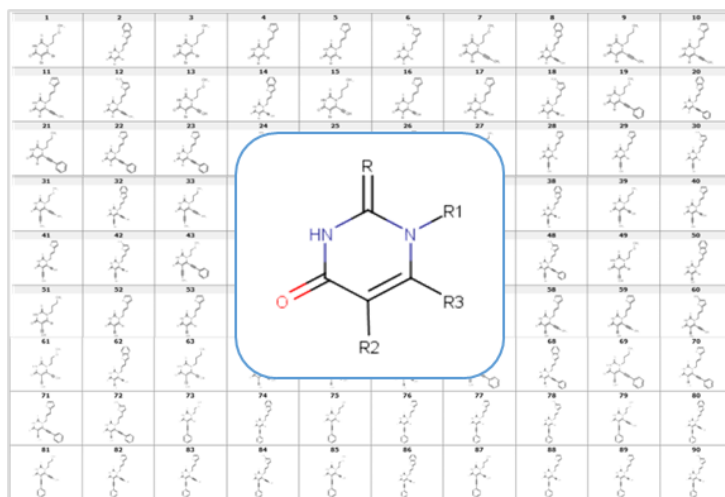


Figure 1.9 Example of a sample of the combinatorial library of HEPT analogs derived from a Markush structure⁴¹

When applied to medicinal chemistry, these new introduced terms – which are englobed in the cheminformatics field – open a door to methodologies containing plenty of opportunities to find new potential drug candidates through chemical space exploration. Computational chemists have endeavored to capture the essence of what makes a compound more drug-like rather than just any conventional organic molecule through vast drug-like libraries, and some successful results have been obtained. Still, it is crucial to understand that these computational tools can never fully capture the complete essence of this complex issue.⁴²

1.2.1. Molecular similarity

Johnson and Maggiora⁴³ enunciated in their Similar Property Principle that compounds that are structurally similar are prone to have analogous properties. Each molecule of a chemical library can be characterized by one or more descriptors (see section 1.2.1.1) and compared with the corresponding sets of features for each of the molecules in the database. Indeed, given an appropriate measure of intermolecular similarity, a list of possible drug candidates can be ranked after finding an active hit structure attending to its nearest neighbors. Consequently, molecular similarity is one of the most exploited concepts in cheminformatics and its related areas of drug discovery and medicinal chemistry^{44–47}. It is key for property prediction, cluster analysis, virtual screening, and molecular diversity evaluation.

Though, as molecular similarity is an abstract idea, not a physical observable, the measurement of molecular similarity is intrinsically subjective and no commonly applicable similarity criteria or rules exist, it depends on the context of study⁴⁷. In fact, the similarity system will be dependent on the features chosen to create the chemical space of study and the used metrics to calculate the distance matrix.

As a result, many similarity coefficients have been developed to provide a quantitative measure of the degree of structural connection between a pair of structural representations. The comparative can be based in real numbers (such as in molecular descriptors) or may be confined to dichotomous values that indicate the absence or presence of certain features (fingerprints).

1.2.1.1. Molecular descriptors

Molecular descriptors are used to represent numerically the structures and properties of the molecules of study. There are many molecular descriptors (Figure 1.10) that capture different aspects of molecules, but they are broadly classified according to their dimensionality⁴⁸. They are actually the basic tool in several chemoinformatics applications such as QSAR modeling, similarity/diversity analysis, and library searching²⁵.

These can be categorized in:

- 1D descriptors: include bulk and physicochemical property descriptors such as atom counts, molecular weight, or hydrogen bond donors.
- 2D-descriptors: topological descriptors which consider the internal atomic arrangement of compounds by fragments or connectivity indices.
- 3D-descriptors: derived from the conformation or 3D molecular structure of the molecule (e.g., 3D coordinates of the atoms in the molecule, molecular shape).

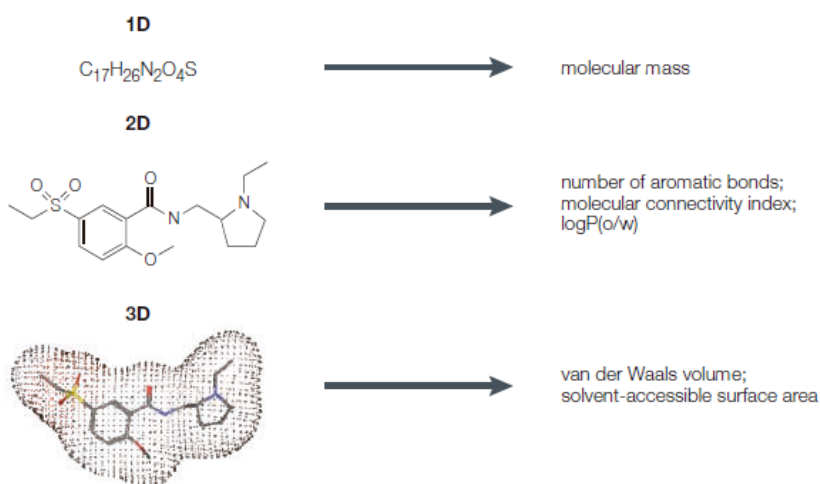


Figure 1.10. Representative molecular descriptors and their classification extracted from Bajorath *et al.*⁴⁹

Alternatively, fingerprints can also be considered for molecular description. In fact, some authors consider them as 2D fragment-based descriptors. These are also commonly known as structural keys that describe the molecular structure through bit strings where each bit indicates the existence or absence through binary values of a predefined substructure (Figure 1.11). Among the

most commonly used fingerprints stands out the MACCS keys⁵⁰ (also known as MDL keys) which consist in the fragment definition by 166 keys which are publicly available and accessible by open-source cheminformatics software packages .

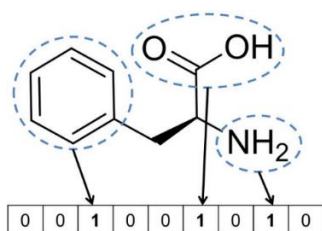


Figure 1.11 Representation of a molecular substructure key-based fingerprints, where bits are set according to the presence (1) or absence (0) of the substructures into the molecule. Source: Bajusz *et al.*⁵¹

Diversity selections are constantly performed for large databases consisting in hundreds of thousands or, in the case of virtual combinatorial libraries, millions of molecules. To be possible, an amazingly fast calculation of the molecular descriptors is needed, and so this alone may prevent the use of those sort of demanding types of descriptor calculations (such as those 3D structure-related features)⁵². Finding the optimal set of descriptors for each target is a time-consuming procedure, which explains the attraction of those descriptors that claim to be universally relevant (1D & 2D molecular descriptors).

1.2.1.2. Similarity coefficients

Dissimilarities between objects are normally based on their distance in multidimensional geometric space which can be calculated by a metric. Hamming (also known as Manhattan or city-block distance) or Euclidean distances are two examples of a more general class of distance metrics named Minkowski distance which follows the general formula shown in [Eq. 3].

$$D_{A,B} = \left[\sum_{j=1}^{j=n} (|x_{jA} - x_{jB}|)^t \right]^{\frac{1}{t}} \quad [\text{Eq. 3}]$$

Being $D_{A,B}$ the distance between the two objects, j the attributes, n the total number of attributes of an object and x_{jA} the value of the j th attribute of object A (same applies to B). The index t determines the different metrics, where $t=1$ for Hamming distance and $t=2$ for Euclidean. The main difference between the two described and other metrics such as Tanimoto, Dice and Cosine coefficients is that the derived from Minkowski's formula consider a common absence or low values (in the case of continuous variables) of attributes as evidence of similarity whereas the latter do not⁵³.

The Euclidean distance is considered the optimal metric as it remains as the most popular measure for molecular comparison from the basis of many different clustering methods. The

implementation of Euclidean-based classification algorithms have shown to be particularly effective for continuous databases on empirical studies found in literature^{54,55}. Euclidean chemical spaces are normally preferred as they can be readily visualized. Even more, through algebraic vectoring measures like the center versus the edges of the overall space or regions (centroids), density of molecules, coverage, direction, and orthogonality can be measured. However, when considering databases consisting only in binary variables, Tanimoto coefficient is the measure of choice for fragment-based (such as 2D bit-strings or fingerprints) chemical similarity work.

1.2.1.3. Describing the chemical space by dimensionality reduction

The chemical representation of a particular database is influenced by the descriptor vector type and length. Thus, the choice of the feature set is the first step to reproduce the expected depth description of the dataset. There is a large variety of descriptors, so their selection is necessary to select the most appropriate to a given application.

Thousands of descriptors have been reported in the literature, allowing limitless possibilities to define chemical spaces. To visualize the diversity of these compounds on the basis of these properties, two approaches can be carried out: (a) a previous feature selection in terms for dimensionality simplification, (b) the use different dimensionality reduction methodologies. Plus, the latter can be considered aiming to preserve as much of the significant structure of the high-dimensional data as possible in the low-dimensional map.⁵⁶

One of the most performed statistical approaches is known as principal component analysis⁵⁷⁻⁶⁰ (PCA) and it is normally used to map descriptors' spaces. The key idea of PCA is to find a new system of coordinates in which the input data with original dimension can be explained by many less variables in a reduced matrix. The plot normally is represented by the plane of the first two principal components (PCs, Figure 1.12) or the space of the first three PCs (that typically involves above the 60-80% of the overall variance).

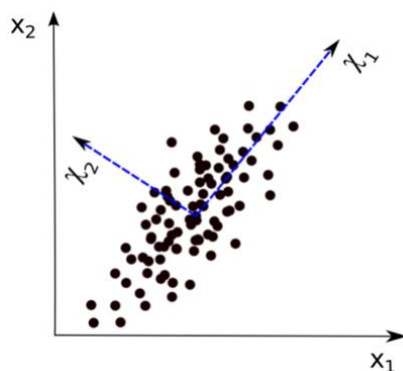


Figure 1.12. Graphical representation of a PCA transformation in 2PCs. Figure extracted from Sorzano et al.⁴⁰

In such property space maps described by PCA, compounds with related structural, physicochemical and sometimes biological activities are generally grouped together.³⁸ PCA-based mapping is a fast and deterministic dimensionality reduction method but, as it omits non-linear feature interactions, overloaded regions of data can be normally perceived.

Another approach for high-dimensional data processing is the t-SNE (t-Distributed Stochastic Neighbor Embedding), developed by Van der Maaten⁵⁶ which has gained large popularity for data visualization. Briefly, data points are embedded in a low-dimensional space (2D or 3D) attending to joint probabilities (local densities) and divergence is minimized. Overall, although it is a good tool for visualization maps for large datasets, it is not recommended or still unclear the use for further reduction purposes as it loses high resolution and a subsequent predictive power.

An illustrative example with Fisher's Iris dataset (Fisher, R. A. Ann. Eugen. 1936, 7 (2), 179–188) is used in Figure 1.13 to evidence the differences between the two visualization methodologies only by its visual inspection. Although the PCA system contains more rigorous information (with a 95% of explained variance) for further clustering issues or rational selection, the plot derived by the t-SNE embedding represents the data plotted in an easier and more intuitive way. This example may seem too simplistic as it departs from a matrix containing 150 samples and 4 features. However, when using a larger dataset, the choice of the proper visualization plot will make a difference to better comprehend the different cluster distributions.

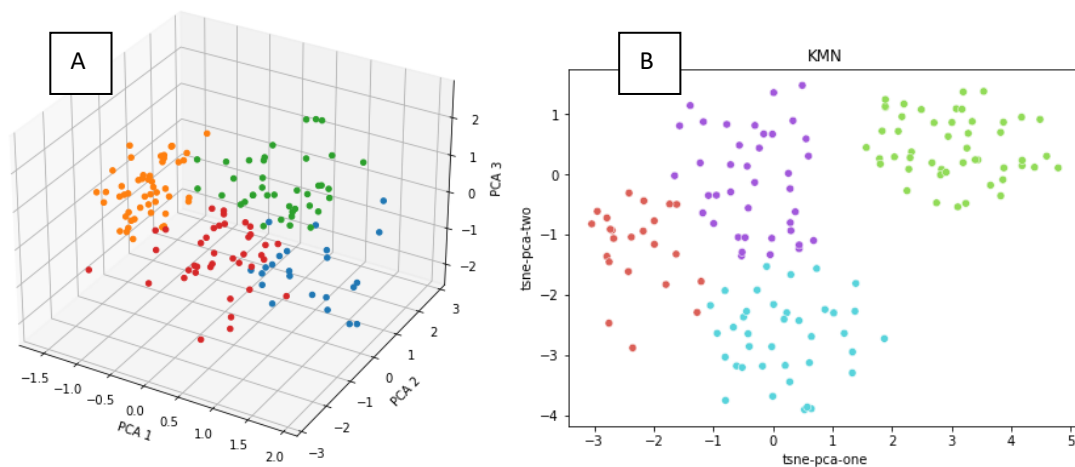


Figure 1.13 Example of a Fisher's Iris Dataset to exemplify the visual difference between (A) a new 3D-system reduced by PCA (B) a new 2D-system defined by t-SNE.

Other alternative methods have been designed and used in literature but have not been considered in this study such as Self-Organizing Maps (SOM)⁶¹ which treats non-linearities in a better way than PCA, or the probabilistic alternative called Generative Topographic Mapping (GTM)⁶².

1.2.2. Exploring the molecular diversity: library curation

All the possible organic compounds that could comprise a universal chemical space is so vast and almost infinite that only a fraction of it has been explored. These can be represented in a unique chemical space attending to their physicochemical properties. Researchers can screen millions of compounds in the search of some that are biologically active. In fact, pharmaceutical companies currently own compounds libraries containing around 10^6 chemical entities. *Chemography*^{63,64} is the named term to the exploratory analysis of this maps where diverse set of molecules are placed. Through this searching methodology, new regions with biologically active analogs can be found.

However, not all the theoretical compounds imagined in researchers mind can be synthesized or show the characteristics to be used as drugs. Thus, the challenge for the scientist is to find spots in this vast chemical space where clusters of therapeutically useful compounds resides⁶⁵. Hence, a chemical space described by a dataset of interest should be filtered attending to their synthetical feasibility and their drug-likeness in order to be empirically studied.

The first point depends on our current extensive knowledge of organic chemistry. The second term was coined by Lipinski in 1994, in its introduction to the 'rule of five', where he stated that orally administrated drugs are far more likely to reside in areas of chemical space defined by a limited range of molecular properties. His seminal analysis of the Derwent World Drug Index stated that, historically, the 90% of the orally absorbed drugs presented the four following characteristics – which are all multiple of five, as an origin of the rule's name – : the compound must present less than 5 H-bond donors, 10 H-bond acceptors, its molecular weight must be lesser than 500 Daltons (Da) and the calculated Log P lesser than 5. These five characteristics enhance a better solubility and permeability to facilitate the drugs absorption. In fact, the overall chemical space is believed to contain at least 10^{60} organic molecules below 500 Da of possible interest for drug discovery^{38,66,67}.

Finally, after pondering the restrictions or filters concerning to the synthetical feasibility or the *druglikeness* profile of the candidates, the last aim in lead generation experiments is to select a diverse selection of samples in order recognize all type of biological activity among a compound collection through clustering, partitioning or QSAR analysis. Once a number of active compounds have been identified for the desired screening, a test of all compounds corresponding to that cluster – or a newly inner diverse subselection – with those hits should usually result in a several number of possible hits being identified, allowing the widest possible choice of lead compounds for optimization⁶⁸.

1.2.3. Clustering and partitioning tools

The assessment of a chemical space can be achieved through many points of view. As an example, plenty of studies describe similarity-based approaches^{53,69,70}, cluster-based selections⁷¹,

dissimilarity-based selections⁷², cell-based selections^{73,74} or optimization-based selections^{75,76}. The recent introduction of non-supervised machine learning techniques⁷⁷ used for both classification (via clustering or partitioning methodologies) and regression is highly effective to assess small datasets and select subsets which might describe significantly the chemical diversity of the overall chemical space^{78–80}.

Several methods have been developed to assess and classify scatterings or distributions of compound sets in a chemical space of study. These can be mainly categorized as cluster-based methods and cell-based methods.⁸¹ Compound clustering and partitioning techniques have a long-standing tradition in computer-aided drug discovery since Willet⁸² work and other review articles developed by Bratchell⁸³, Barnard and Downs⁸⁴, and Downs and Willet⁸⁵. The present main uses of chemical datasets classification are to find representative subsets for HTS purposes, combinatorial chemistry, and to increase the diversity of internal datasets for other reprofiling uses⁸⁶. One of the principal applications of these methodologies is to classify databases of compounds in smaller groups of similar compounds in order to choose representative compounds for biological testing when there is availability (Figure 1.14).

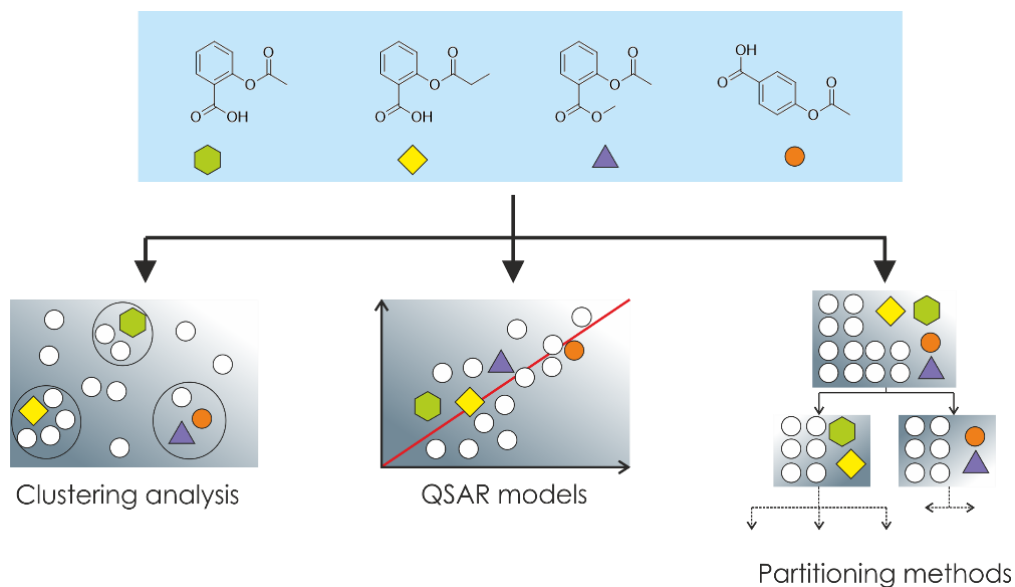


Figure 1.14 Different methods applicable for virtual screening. Adapted figure from Bajorath *et al.*⁴⁹

Clustering is the process of subdividing a group of objects – in this study, molecules – into groups (clusters) according to their proximity and, hence, similarity in the inner cluster and an inter-cluster dissimilarity.

As introduced, since decades ago, many authors^{84,86,87} have attempted to describe these methodologies for molecular mapping. Altogether, the clustering process for chemical structures normally follows the scheme:

- A set of attributes or descriptors on which to base the comparison of structures are selected. Normally, as a rule of thumb, structural and physicochemical properties are considered.
- The database is built-up by calculating these descriptors and the structures are characterized.
- Coefficients of similarity, dissimilarity, or distance measurements between all the possible pair of samples in the dataset are calculated.
- The desired clustering methodology is applied to group those structures presenting similar structures by the calculated matrix of distances or the similarity coefficients (which can be calculated between the existing objects or new abstract objects such as the centroid of a sample set or mean value).
- The resultant clusters are analyzed in order to select a diverse sample of compounds.

Compound clustering depends essentially on the calculation of intermolecular distances in the chemical space of reference. The use of large datasets may difficult the computational and time cost of these algorithms due to the constant pair-wise distance comparison. Thus, an alternative classification methodology can be performed through partitioning algorithms. These later methods are based in establishing a consistent reference frame that allows the assignments of coordinates to each database independently of others. This results in a more computational efficient method as, when particularly considering cell-based partitioning methods, the chemical space is subdivided in tiny boxes or bins through a binning algorithm (Figure 1.15). This is based in low dimension chemical spaces obtained through data reduction (e.g. PCA, SOM).

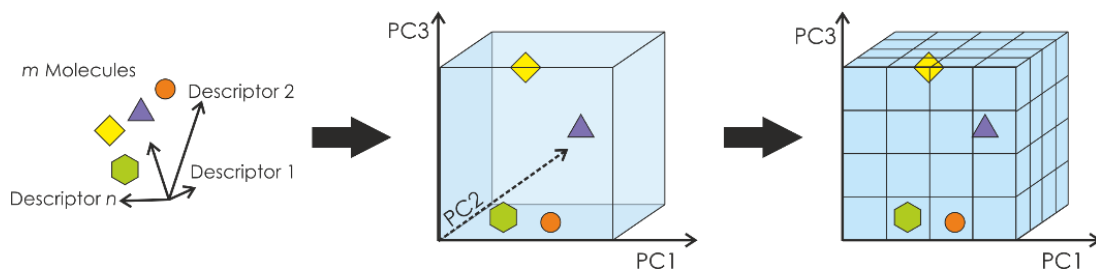


Figure 1.15 Schematic representation of chemical space reduction to PCA Chemical Space and further partition through binning. Adapted figure from Bajorath *et al.*⁴⁹

1.2.4. Assessment parameters: space and population coverage

In contrast with the above-mentioned FW analysis which tends to lead to highly focused studies of chemical libraries, the selection of a compound inside of each cluster might be more representative to describe all the cluster population. This representative sample can be randomly chosen (*cherry-picking*) or can be the closest to the center of the cluster (*centroid*).

Additionally, in order to contrast different selections and chose the most favorable option, the molecular representativeness and space description of the selected diverse subsets can be assessed using two objective functions: the space coverage and the population coverage^{78,88}.

On the one hand, Space Coverage (SC) represents the percentage of selected (occupied) clusters or bins (k_{oc}) by a given number of selected compounds over the total number of partitions, k [Eq. 4].

$$SC = \frac{k_{oc}}{K} \cdot 100 \quad \text{[Eq. 4]}$$

On the other hand, given a database of N compounds, Population Coverage (PC) measures the SC weighted by the occupancy in each cluster or bin, by dividing the population of the occupied clusters (n_{oc}) among N [Eq. 5].

$$PC = \frac{n_{oc}}{N} \cdot 100 \quad \text{[Eq. 5]}$$

The procedure to calculate SC and PC is exemplified in Figure 1.16. Considering the selection of 9 compounds among a database consisting in 50 molecules ($N=50$, represented as dots). After applying a clustering or partition-based method, the chemical space is divided into 16 partitions (clusters or occupied bins, $k=16$) as shown in scheme A. If the 9 selected compounds (depicted as stars in B) are distributed in 7 clusters ($k_{oc}=7$, in orange), this corresponds to 43.8% SC. Moreover, in this colored clusters are involved 28 samples of the database as shown in C. Hence, the 7 selected molecules are representative of 28 out of the 50 compounds included in the database, which represents 56.0% PC.

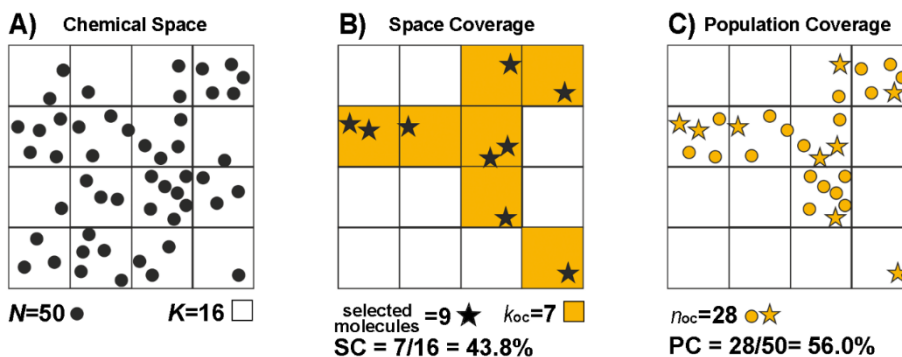


Figure 1.16. Schematic representation to illustrate the space and population coverage of a dataset of 50 molecules in its chemical space.

1.3. Previous studies in the research group

Willet *et al.*⁸⁹ implemented by first time in the mid-1980s a methodology based in rational selection of subsets through a previous dataset clustering to improve the Structural Representative File of Pfizer Central Research (UK) which was being performed via an unproductive time-consuming intellectual operation. In their study, they proved that an effective clustering procedure such as Jarvis-Patrick Clustering Methodology, did not overlook certain classes of compounds being the new selection more representative, consistent, and free of bias. They hypothesized then that the use of clustering and similarity techniques would be proven to be highly effective in providing computational support for pharmaceutical research. This served as first evidence of an efficient optimization of the first screening steps in drug discovery throughs diverse selections using cheminformatic tools.

Later, in 2003, R. Pascual *et al.*^{41,90} studied the case of a family of similar compounds related to an approved drug to discuss the chemical space explored in R&D in order to suggest alternatives for an optimal dataset description through, once more, rational or diverse selection. The goodness of 11 available clustering methodologies and their subsequent standard diversity selection methods was assessed with the combinatorial library of HEPT – an inhibitor of HIV-1 transcriptase inverse – containing 125,396 analogs. It was analyzed the distribution of the 180 compounds described in literature in the combinatorial accessible space of analogs and results showed that indeed the conventional and intuitive procedure of selection had restrained the coverage by a 9% in its space giving smaller values than a random selection (38%).

A selection of only 25 compounds was proven to represent the 90% of the population. Afterwards, R. Puig de la Bellacasa *et al.*⁹¹, designed a synthetic route for a selected combinatorial sub-library analogs. After the synthesis of the derivatives and biological testing, some molecules exhibited an outstanding HIV-1 inhibitory activity showing a broad activity range (with EC_{50} between

0.05 μM and $>90 \mu\text{M}$), even better than the reference compound (HEPT, $\text{EC}_{50} = 3.3 \mu\text{M}$). Among these, one was 67-fold and 40-fold more active than HEPT, the original hit.

1.4. Hypothesis

The Drug Discovery strategy includes a hit-to-lead process which involves large research on the synthesis of derivatives suggested from an original hit that had previously shown biological activity against a specific target. Therefore, this process implies the synthesis of many analogs leading to the description of a chemical sub-library which leads a highly focused study (traditionally based on a Free-Wilson approach) on the chemical space nearby the hit compound. However, when this drug is finally patented, a wider chemical space derived from a Markush structure is described, theorizing that some analogs within may present biological activity. Nevertheless, this claim presenting a Markush structure does not imply the proven synthesis of all the molecular library but just a focused small population of it.

Thus, we hypothesize that there is a great part of the chemical space of these libraries that is unexplored and can hide potential lead candidates which may surpass the activity of the original hit or even minimize its undesired side effects. Moreover, we suggest an alternative methodology which would optimize the chemical space exploration through clustering and rational selections algorithms through cheminformatics.

We aim to proof that a first clustering classification could chief a pathway to next discovery steps by dictating which compounds could be further synthesized and tested. Therefore, our methodology suggests the use of diversity selection through clustering and rational selection-based practices in the drug discovery process in the selection of trial compounds for activity testing in early biological screens for hit exploration or even to describe families or analogs for further reprofiling purposes. We hypothesize that if any of the selected compounds proves to be active when tested in the bioassay of interest, then it is worth to assay other compounds in its cluster in a looped methodology since some of them may also show potential activity.

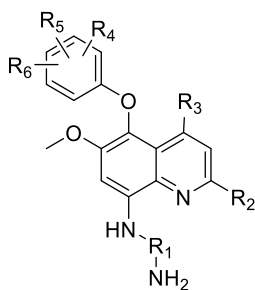
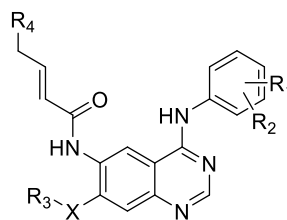
Finally, through our suggested procedure not only optimized leads can be obtained but selections of representative compounds that can be used to maximize the efficiency of random screening in lead-discovery programs providing more systematic methods of selection than in the used in many current empirical screening systems.

1.5. Objectives

The main goal of this thesis is to proof that there is still space to be explored in the resulting combinatorial libraries derived from the claimed Markush structures in the drug patents. Additionally, we defend that rational selection can perform a better exploration of these chemical spaces

discovering new hidden reservoirs with potential biological activity or, at least, describing the dataset for further reprofiling purposes. With this purpose, the following specific objectives are defined:

- Proof the advantages of rational selection against the actual R&D methodology through the chemical space assessment of several real case studies of recently FDA approved drugs and their derived combinatorial libraries.
- Perform a proof of concept to give insights into the advantages of the suggested exploratory strategy with the study of Tafenoquine and Dacomitinib combinatorial libraries (**10**, **30** respectively) derived from their in-patent Markush structures. Improve their chemical space description through rational selection, synthesis, and biological assessment to contrast the results with the reported analogs until date.

**10****30**

Chapter 2: Rational vs. Traditional R&D methodology

2.1. Review of the latest drugs approved by the FDA (2008-2020)

The description and exploration of the *druggable* chemical space has always been of great concern given the overwhelming number of molecules that can be obtained by fragment combination (which was barely estimated to 10^{60} small molecules)⁶⁶. In light of this scenario, finding new molecules with drug-like properties become as hard as finding a needle in a haystack, at least without the support of computer-aided techniques. As the total exploration of this vast space is computationally unmanageable, many cheminformatics tools have been developed for a better mapping of the chemical space by means of molecular descriptors, enabling the rational navigation through it, contributing to new hit discovery^{67,92,93}.

Although the introduction of innovative technologies in the Drug Discovery pipeline in the beginning of the century (bio and cheminformatics, combinatorial chemistry and high-throughput screening, to name but a few), the pharmaceutical industry has always been challenged by increasing costs in research and development (R&D) due to complex variables^{94,95}. Hence, there are still new tools to be exploited such as Artificial Intelligence, Deep Learning, Machine Learning or drug repurposing⁹⁶⁻¹⁰⁰ to improve the R&D efficiency process, lower its costs and offer a more diverse spectrum of drug-like small molecules to treat as many diseases as possible.

In the present study, a first attempt to shape the already known *drugspace* was performed with the data collected by the FDA's latest approved drugs in the past years. A total of 483 new drugs have been approved by the U.S. FDA regulatory system in the period of 2008-2020 (see Annex I). This increasing tendency of approvals has contradicted the initial predictions probably due to the new modalities developed that have enabled advances in therapeutic areas and disease pathways¹⁰¹. Even with the growth of approved number of therapeutic Biologic Applications (BLAs) observed in the last years (Figure 2.1), 366 of the total approved drugs (76%) consist in New Molecular Entities (NMEs) and, more specifically, 348 are small molecules. Hence, the development of small molecules is still booming in the pharmaceutical research sector. Over this period, the median number of NME and BLA approvals per year was 39, with a low of 21 in 2010 (with 13 NMEs and 8 BLAs) and a high of 59 in 2018 (with 42 NMEs and 17 BLAs).

These compounds can be classified depending on its target treatment in different categories attending to the Index of Classification of Diseases (ICD-11¹⁰²). According to the ICD-11, the main aim of the actual research is focused on neoplasms (26.7%), infectious or parasitic diseases (12.4%), endocrine and nutritional diseases (10.8%), diseases of the nervous system (10.4%) and diseases of the circulatory system (3.9%) as shown in Figure 2.1.

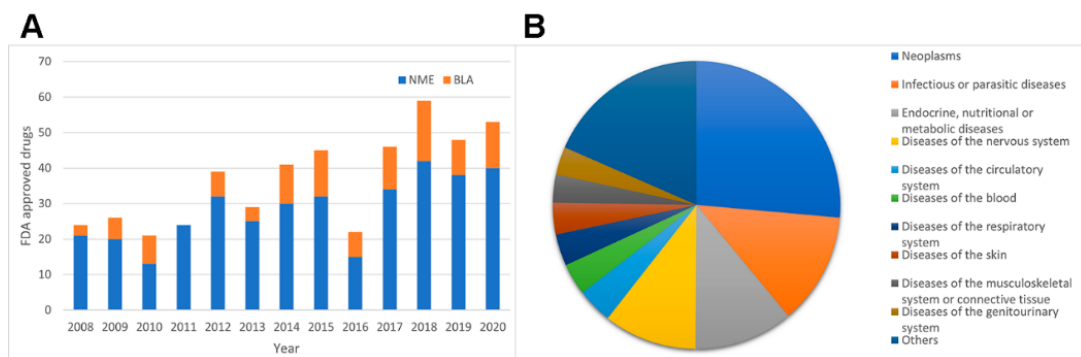


Figure 2.1 (A) Bar plot of the approved NMEs and BLAs drugs by the U.S. FDA in the period of 2008-2020. (B) Pie chart representing the number of drugs approved by the U.S. FDA in the period of 2008-2020 classified by the ICD-11.

After curating the data of the latest 348 small molecules approved, a resulting database of 336 compounds can be obtained and the features' space consisting in 206 1D and 2D molecular descriptors (see Annex II) reduced to 41 PCs to explain the 95% of its variance. Even approximate, the graphical representation of the first 3 PCs provides a useful depiction about the distribution of FDA compounds into its chemical space. Since no structural motif is exclusively related to a particular disease, molecules are spread out in space without any correlation with their ICD-11 categorization as expected (Figure 2.2).

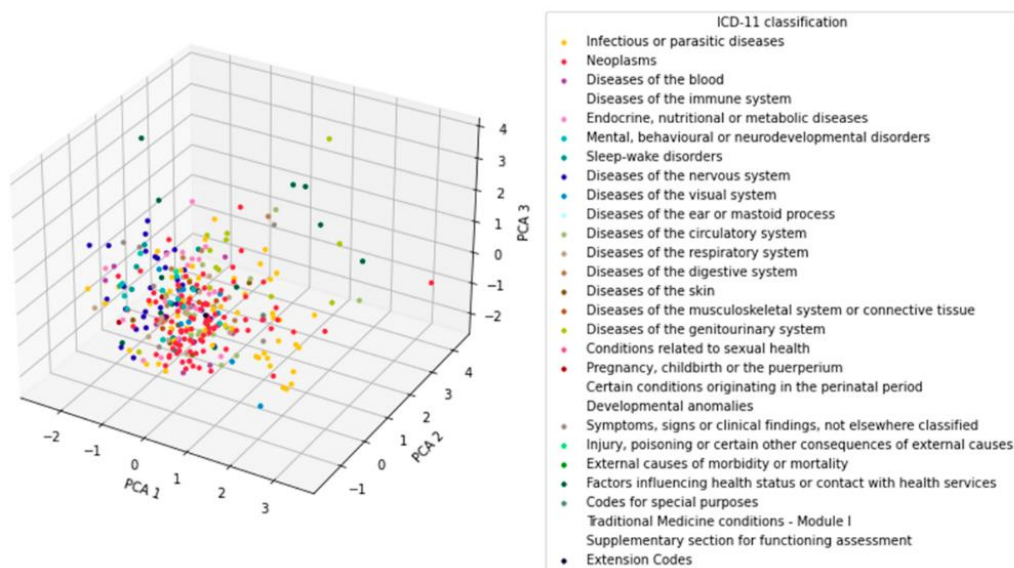


Figure 2.2 Chemical space described by the firsts 3 PCs (53% of variance explained) of the dataset of small molecules approved by the FDA in 2008-2020 classified according to ICD-11. Not represented categories are not colored in the legend.

2.2. A glance of the *drugspace* described by ChEMBL

The drug-like chemical space or *drugspace*, is nowadays challenged by its dataset description which can be explored using automation and algorithms to improve the probability of new pharmaceutical hit discoveries. Moreover, many chemical spaces derived from different databases can be overlapped and hence, contrasted, to enrich the information of the possible scaffolds that may present biological activity.

Many authors have attempted to describe a global chemical space through different approximations. A global *drugspace* map was firstly introduced by Oprea *et al.*⁶³ through their standard tool for compound prediction within the same PCA model called chemical global positioning system (ChemGPS) which by using one training set 423 satellite and core structures can preserve cluster characteristics of local models. This approach was later enlarged with the ChemGPS-NP for natural product related research which avoided extrapolation defending a more robust prediction engine. Other approaches based in ligand-receptor interactions for guiding medicinal chemists was also introduced by Rabal *et al.*¹⁰³ through their biologically relevant Chemical Space Navigator (RCS) or the Hierarchical Generative Topographic Mapping introduced by Zabolotna *et al.*⁶⁴ can be considered for a global *drugspace* depiction. Here, far from describing and suggesting a standard methodology for a global chemical space exploration, a huge PCA-dimensional chemical space is used in discussion for the assessment of the largest possible druggable chemical space known until date.

A new approach to shape an optimal *drugspace* has been carried out to gain insight into the significance of the FDA database within the chemical space. To localize the drug-like environment approved in the latest years in a more extensive *drugspace*, it has been interpolated into the reduced space derived from ChEMBL database. This contains curated bioactive molecules with drug-like properties and is a database of reference in the medicinal scientific community (its access is available through a web-based interface). This largest dataset was previously lowered from 1,914,538 to 1,896,082 compounds (*N*) after data filtering (discarding those inorganic and macromolecular compounds) and the dataset consisting in the calculated 206 1D and 2D molecular descriptors (see Annex II) was reduced to 52 PC containing 95% of variance.

At first glance, Figure 2.3 reveals that FDA (in blue) shows a sparse distribution in ChEMBL chemical space supporting that there is correlation between both datasets.

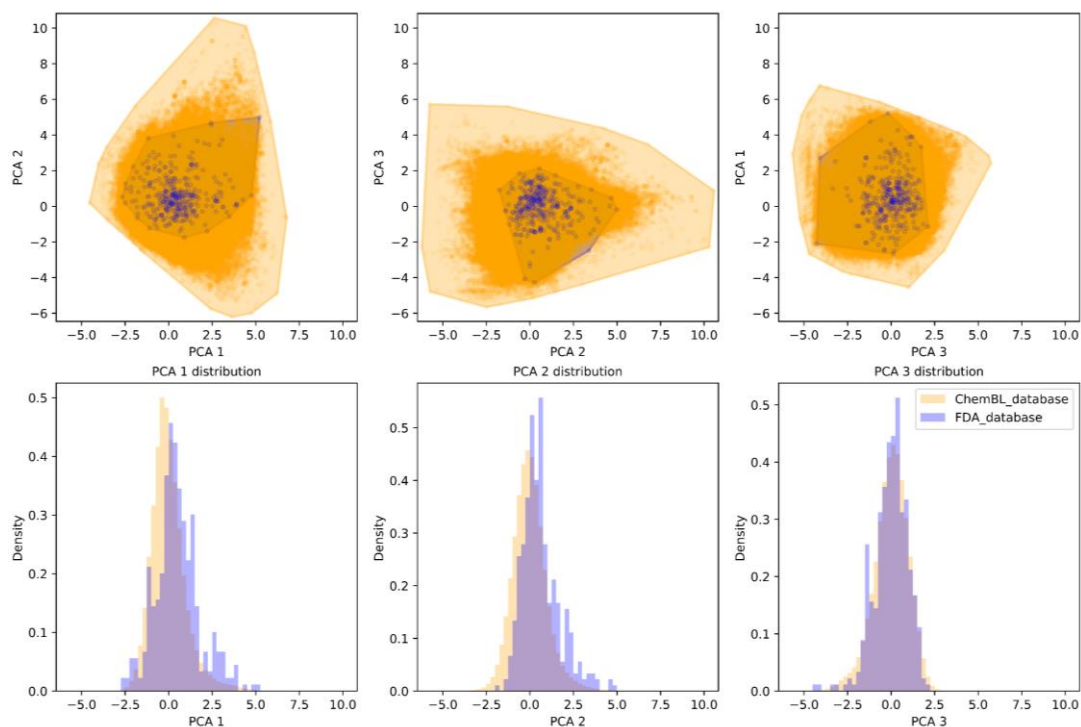


Figure 2.3 Three planes of 3 the first PCA plots and their density plots to represent FDA dataset (in blue) in ChEMBL chemical space (in orange)

In the present thesis, to delve into the unknown space and increase the chances to find new and more active hits a clustering-based selection methodology has been chosen as it has been proved to be efficient and effective for similarity matching and classification in chemical information systems⁸⁹. Unsupervised Machine Learning tools such as data clustering have been chosen to assess mathematically the space overlap or coverage between datasets and selection of compounds. The goodness of eight clustering (HRC single, HRC complete, HRC median, HRC average, HRC centroid, HRC, Ward, KMN and KMED) and two partitioning methodologies (binning and optimum variance binning) have been tested to explore and evaluate the chemical spaces in terms of space overlap or coverage between dataset and selected compounds (later applied in section 2.3.2). The later partitioning binning implemented (OV binning) has been developed as an adaptation of the optimum binning procedure reported by Pascual *et al.*⁴¹ (see Chapter 5) as previous work in the group evidenced the need of more concise partitioning methodologies. In order to assess the overlapping between datasets the measurement of the already introduced coverage parameters have been used in discussion; space and population coverage (SC and PC).

However, in this specific case, due to the size of ChEMBL dataset, only partitioning methods have been computationally easy-to-handle for a coverage discussion. Looking for an optimal standard

selection, a study was performed for the standard sizes⁷³ of $\sqrt{N}/2$, \sqrt{N} and $10 \cdot \sqrt{N}$ clustering formation to assess the explored space by the FDA latest approved drugs.

To illustrate an example, considering ChEMBL as a case of *drugspace* to be explored, when studying the selection of $\sqrt{N} \approx 1,377$ compounds, the binning partitioning presents a total amount of 1,216 occupied bins and 1,276 bins are occupied using OV binning. The latest approved 336 small molecules (FDA database) represents a 4.1% SC and 92.5% PC of the chemical space of ChEMBL database in the case of binning partitioning. For OV binning partitioning methodology, a resultant 7.3% of SC and 86.0% of PC are obtained. It is worth noting that SC and PC percentages obtained with FDA compounds are below the random selection, whether considering a cherry-picking choice of \sqrt{N} (1,377), $10 \cdot \sqrt{N}$ (13,770), $\sqrt{N}/2$ (689) or same size as FDA database compounds (336) (Table 2.1). These results evidence that, although the approved molecules seem to be spreadly located in different regions of the *drugspace* of study – as expected, not overlap is seen due to patentability purposes –, they are, in terms of molecular representativeness, highly located in a small area representing less than 10% SC dismissing regions were potential new scaffolds could be found.

Even more, considering, in this case, that a rational selection of 336 molecules from different bins could optimally cover the 27.6% SC (336 out of the 1,216 occupied bins) or 99.8% PC if rationally picking samples of the most populated bins via Binning partitioning and the 26.3% SC (336 out of the 1,276 occupied bins) and 99.7% PC via OV Binning, it is confirmed that there is still space with drug-like properties still to be explored. Instead, in terms of PC, it can be observed that there is a good representation of the overall *drugspace* population. This fact makes evident that there are a few hyper-populated bins originated by a hit-focused analysis and a large number of less populated bins which hide druggable space still to be profoundly investigated. At the light of the results, it was firstly verified in this case that a random selection of the same number of compounds would better describe the space.

Table 2.1 Space and Population Coverage results obtained for FDA database and random selections in the ChEMBL's chemical space in $\sqrt{N} \approx 1,377$ bins.

	Binning k=1,216		OV Binning k=1,276	
	SC %	PC %	SC %	PC%
FDA samples	4.1	92.5	7.3	86.0
336 Random Samples	5.5	95.0	5.8	87.5
336 Rationally Selected Samples	27.6	99.8	26.3	99.7
1,377 Random Samples	7.3	97.0	12.6	93.9

Table 2.2 Space and Population Coverage results obtained for FDA database and random selections in the ChEMBL's chemical space in $\sqrt{N}/2 \approx 689$ bins.

	Binning k=598		OV Binning k=515	
	SC %	PC %	SC %	PC%
FDA samples	6.6	94.2	11.3	92.4
336 Random Samples	5.8	95.2	10.1	91.9
336 Rationally Selected Samples	56.2	100.0	65.2	99.9
689 Random Samples	6.9	96.1	12.3	93.3

Table 2.3 Space and Population Coverage results obtained for FDA database and random selections in the ChEMBL's chemical space in $10 \cdot \sqrt{N} \approx 13,770$ bins.

	Binning k=11,925		OV Binning k=12,981	
	SC %	PC %	SC %	PC%
FDA samples	1.2	61.6	1.6	51.1
336 Random Samples	0.9	78.1	1.4	61.9
336 Rationally Selected Samples	2.8	92.4	2.6	82.9
13,770 Random Samples	5.1	86.8	7.5	77.5

In all cases of size selections, ($\sqrt{N}/2$ and $10 \cdot \sqrt{N}$ sizes are shown Table 2.2 and Table 2.3) the results agree that a rational selection of 336 candidates do better represent in terms of SC and PC the coverage of the *drugspace* described by ChEMBL proving to be a promising exploratory methodology.

2.3. Assessment of the actual chemical space exploration performed in seven cases of study

Frequently, the common hit-to-lead optimization methodology departs from a slight structure modification of the original hit and, next, the biological activity is compared to see if there it has been any improvement to continue with that chemical variation. The actual following procedure is to perform a Free-Wilson approach which results in a number of sequential modifications that explores the surrounding chemical space of the initial active compound.

However, neither the real space explored, the bibliographical data (BD), nor the fragment combination of the studied molecules (BCL or bibliographic combinatorial library) significantly represent the space defined by the combinatorial libraries derived from the Markush structure (Markush combinatorial library, MCL) when comparing it with the coverage a statistically random sampling of \sqrt{N} molecules (being N the size of the dataset, see Chapter 5 for the other introduced library sizes). Thus, we sustain that there is a large part of the chemical space claimed on patents that remains unexplored and it can hide potential leads that may surpass the activity of the original hit or reduce undesired side effects. In our opinion, rational selection algorithms could assist traditional methodology to optimize the selection of representative compounds of a given chemical space.

To demonstrate that rational selection can improve the efficiency to describe a chemical space, we have studied seven drug patents approved by the FDA (six of them in the last 12 years) referring to different therapeutic targets, with different sizes (in terms of the number of analogs included in the Markush structure) and including on-patent (ONP) and off-patent (OFP) drugs. OFP drugs have been included in the study to discuss if the exploration of the chemical space is limited due to the situation of their property rights. Accordingly, Leflunomide (the oldest approved drug considered) has been added to the study to verify that this exploratory deficiency is still present in an elder patent.

2.3.1. Diverse drug selection and combinatorial library assembly

We have tested the goodness of rational selection over the traditional approach (described by BD) using seven different datasets obtained from the Markush structure stated in the patents of Dacomitinib, Abemaciclib, Tafenoquine, Ertugliflozin, Rufinamide, Azilsartan Medoxomil and Leflunomide drugs^{104–111}, retrieved from the Orange Book¹¹². To demonstrate our hypothesis, it has been chosen seven different drugs attending to their ICD-11 category, therapeutic target, number of analogs – with a computationally handleably number of compounds –, their year of approval, applicant, and scaffold (Table 2.4). These diverse features have been explicitly considered to minimize a biased effects in the result due to common structural features or others. Consequently, we can better relate the results of these real cases to an extended traditional R&D Drug Discovery methodology.

Table 2.4 Information related to drugs under study.

Case Study Drug	Year of approval	Applicant	Disease Treatment	Status	Markush Patent Pub. Year
Dacomitinib	2018	Pfizer Inc.	Metastatic Non-Small-Cell Lung Cancer	ONP	2005
Abemaciclib	2017	Eli Lilly & Co	Breast Cancer	ONP	2010
Tafenoquine	2018	GSK	Malaria	OFP	1986, 2002
Ertugliflozin	2017	Merck	Diabetes	ONP	2010
Rufinamide	2008	ESAI	Lennox-Gastaut Syndrome	OFP	1991
Azilsartan Medoxomil	2011	Takeda	Hypertension	ONP	2006
Leflunomide	1998	Sanofi	Rheumatoid Arthritis	OFP	1978

Corresponding combinatorial libraries (Markush Combinatorial Library, MCL) were developed from the Markush structure stated in the patent of each drug (Figure 2.4).

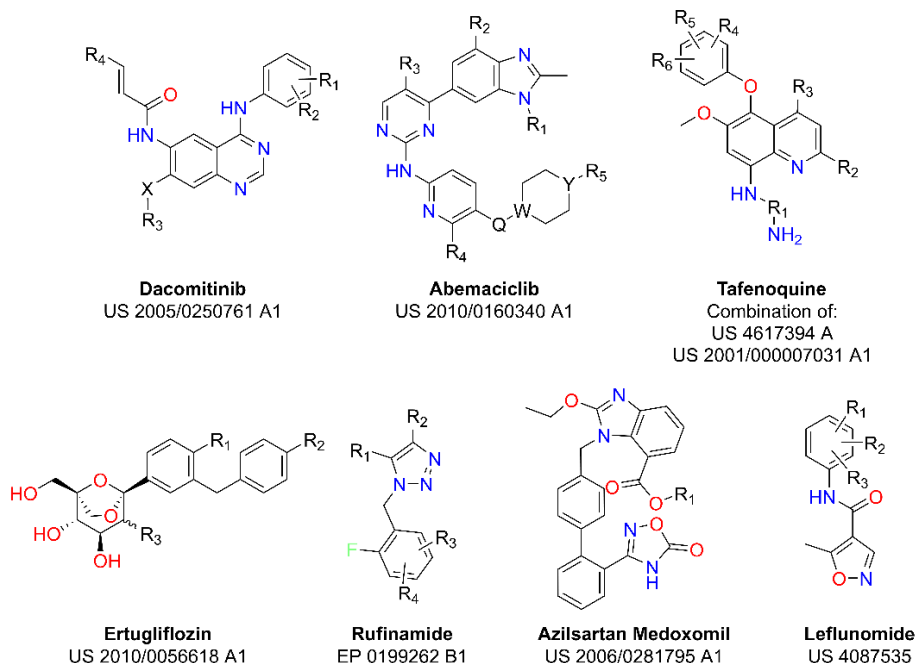


Figure 2.4 Markush structure reported for the seven drugs under study.

Markush combinatorial libraries were fully enumerated following the specifications as described in the patent's claims. As introduced, Markush structures may present vague generic

substructure definitions such as “ R^{16} is hydrogen, alkyl, alkenyl, aryl, aryl-substituted alkyl, cycloalkyl, cycloalkenyl, cycloalkyl-substituted alkyl or cycloalkenyl-substituted alkyl”¹¹³. These unprecise terms where neither number of carbons nor homology are specified may lead the generation of large sublibraries only for their substituent enumeration leading to a huge Markush-derived combinatorial library which could not be fully enumerated. In consequence, we have specifically selected patents with clear radical descriptions to avoid inaccuracy. In fact, among the patents selected only Dacomitinib presented ambiguous definitions like; $-(CH_2)_m$ -Het where m can be an integer 1 to 4 and Het a heterocyclic moiety among a given list but where no connectivity to the ring is specified. Hence, as an example, for the case of $-(CH_2)_2$ -morpholino, three possible substituents could be considered as the methylene could be linked to the 2,3 or 4 position of the morpholine structure. Consequently, this vague definition results in a dataset enumeration of more than 10 billion original compounds.

Therefore, only for the unique case of Dacomitinib the original size of MCL was reduced to a computationally handleable size by only combining the fragments present in the 51 compounds described in claim 5. Including all the positional substitution in the aryl ring described by R1 and R2, the unmanageable dataset of more than 10 billion original compounds was reduced to 16,530 ($19 \times 15 \times 58 = 16,530$). Thus, the final MCL of this specific case describes a expectedly explored subspace of the originally described by the its patent claim. The final list of substituents considered in the study concerning to the Dacomitinib Markush structure shown in Figure 2.4 is explained in Table 2.5. The complete list of substituents for each dataset can be found in the Annex III.

Table 2.5 Fragments in SMILES used to build the 16,530 compounds library of Dacomitinib analogs.

R1	R2	X-R3		R4	
-H	-H	-O-C	-O-CCF	-CN1CCCC1	-Cn1cncc1
-F	-F	-S-C	-S-CCF	-CN1CCC(F)CC1	-CN1CCCC1
-Br	-Br	-N(H)-C	-O-CC(F)(F)F	-CN1CCCC(F)C1	-CCN1CCCC1
-Cl	-Cl	-O-CC	-O-C(F)CF	-CN1C(F)CCCC1	-CCN1CCC(F)CC1
		-O-CCC	-O-CCCN	-CN1CCOCC1	-CCN1CCCC(F)C1
		-O-C(C)C	-O-CCCN1CCOCC1	-CN1CCCCC1	-CCN1C(F)CCCC1
		-O-C(F)(F)F	-O-CCN1CCCC1	-CN1C=C1	-CCN1CCOCC1
		-O-CF		-CN1CC=CCC1	-CCN1CCCCC1
				-CN1CCNCC1	-CCCN1CCCC1
				-CN1CCN(C)CC1	

On the contrary, in the antimalarial Tafenoquine case, the library of analogs was extended to the combination of the structures found in two expired patents which included this drug in order to enlarge the combinatorial library to the maximum patently protected chemical space in the last decades.

Both Tafenoquine and Dacomitinib cases have been studied in detail and will be further explained in the following Chapter 3 and Chapter 4 respectively.

2.3.2. Choosing the optimal clustering/partitioning methodology

To broach the chemical space distribution discussion and the representativeness of certain areas described by selected compounds, suitable clustering methodologies are aimed. It is well known that depending on the dataset distribution some methodologies will be more convenient than others^{114,115}. However, as the object of this study is to prove the lack of chemical space exploration through many points of view and diverse patent cases, several methodologies have been considered. The goodness of eight clustering and two partitioning methodologies was initially studied to assess the different chemical spaces described in each patent of study.

In order to proceed with data clustering, its previous features assembling and dimensionality reduction through PCA was performed. Dacomitinib's database was reduced to 22 PC, Abemaciclib's to 19 PC, Tafenoquine's to 13 PC, Ertugliflozin's to 28 PC, Rufinamide's to 12 PC, Azilsartan medoxomil's to 11 PC and Leflunomide's to 14 PC. The original databases of study containing the analogs SMILES and their calculated descriptors can be found as files in Annex VII.

HRC with a bottom-up approach with six linkage methods (single, complete, median, average, centroid, and Ward), KMN and KMED, binning, and OV binning were firstly used to divide the chemical spaces into k -optimal clusters in order to see the grade of homogeneity in the population distribution per cluster. The k -optimal value was firstly calculated through average silhouette criterion^{116–118} using KMN as the standard clustering methodology for all cases. The silhouette score is a method of interpretation and validation of the consistency within the clusters. This describes the degree of cohesion (in terms of similarity) of a sample in its proper cluster compared to others. It ranges from -1 to 1; the highest the score, the best match of the sample to its cluster has. When considering the mean silhouette coefficient of all the objects, the optimal cluster configuration can be achieved once the highest value is obtained. This score can be assessed by the elbow method, which determines the model that fits best to the dataset distribution. When considering a line chart, the *elbow* can be easily recognized as a strong inflection point when considering a certain k number of clusters.

In this study, the maximum number of clusters to study was settled to a range from 1 to N (being N the number of samples for each combinatorial library). Only in the cases of the largest datasets (Abemaciclib and Ertugliflozin), the total calculation could not be achieved. Figure 2.5 shows the plots obtained calculating the mean silhouette score (Y axis) for a given range of calculated clusters (X axis). The range steps were settled to 100 although a finest approach was considered for tafenoquine database to discard if the lack of elbow perception was due to a poor range step definition. Still, as no inflection points can be perceived, no optimal cluster were obtained for all cases through this methodology as the average silhouette score measured shows a smooth concave curve trend being the optimal value of k unclear. This may be caused by the high correlation of the data due

to their combinatorial nature which results in a not very clustered databases and the large size of the sets which commonly suffer from the *curse of dimensionality*^{119,120} and lack of efficiency.

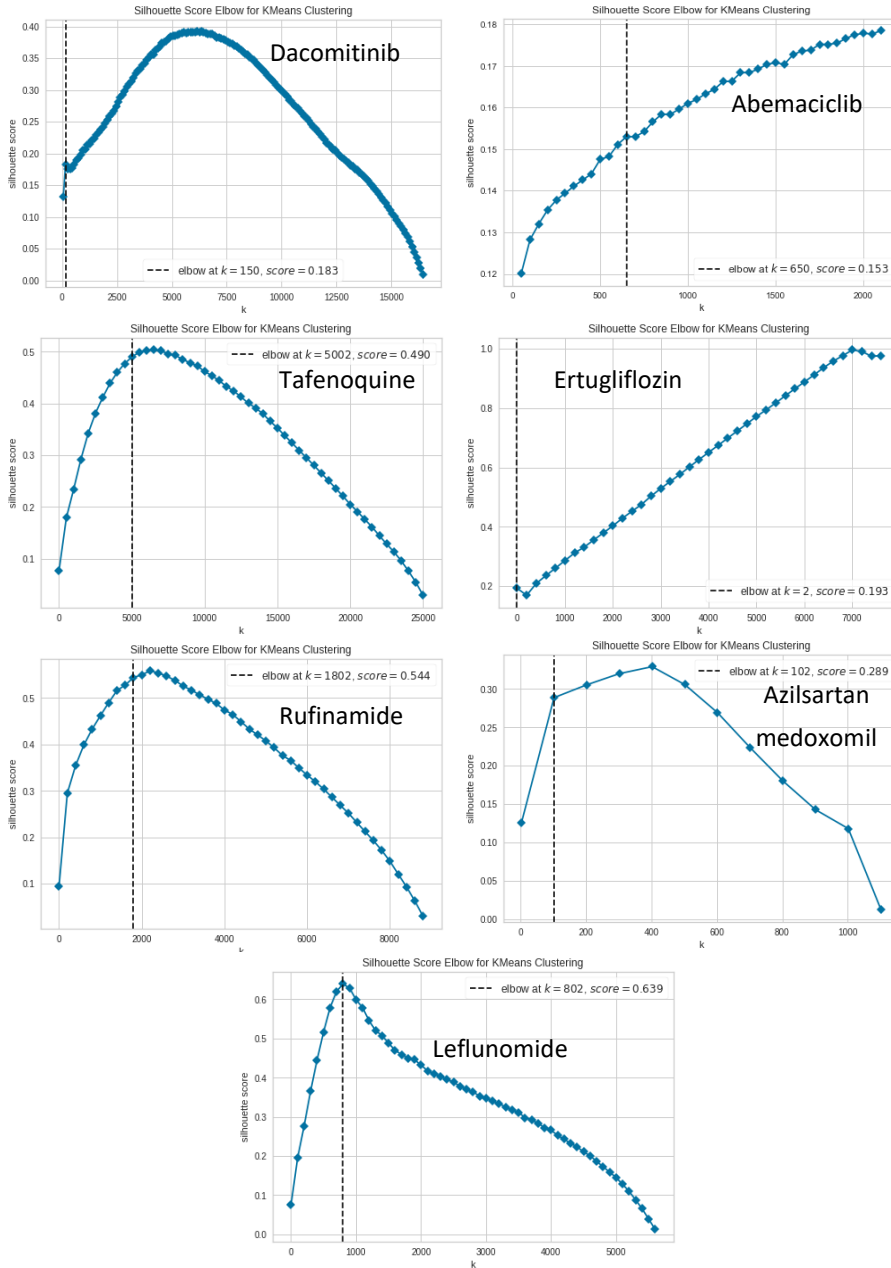


Figure 2.5 Silhouette Score plots describing the poor convergence to an optimal number of clusters for all the cases of study

Given the first inconclusive attempts, new standard values commonly used as a rule of thumb in clustering analysis were considered ($\frac{\sqrt{N}}{2}$, \sqrt{N} and $10 \cdot \sqrt{N}$) to find the most representative partitioning size to be applied in further comparative discussions.

The case of Tafenoquine's chemical space (Table 2.6) serves as an explanatory example to discuss the population distribution using each different size. The aim of this approach was to find the most balanced distribution of data avoiding as many singletons and hyper populated clusters (corresponding to high standard deviation values) as possible. The obtained results evidenced that the \sqrt{N} clusters seemed to be a good compromise. This size first showed more balanced population frequencies per cluster than $\frac{\sqrt{N}}{2}$, which results showed large values in their standard deviations. Moreover, when compared to $10 \cdot \sqrt{N}$ clusters (a very large number of clusters for these datasets), \sqrt{N} clustering presented a much lesser number of singletons.

Table 2.6 Population distribution per cluster (k) analysis in Tafenoquine's analogs chemical space varying the clustering standard sizes. The mean value (\bar{x}) of the population frequencies per cluster and its subsequent standard deviation (σ) along with the ratio of singletons to number of clusters (%s) are considered in discussion.

	$k = \frac{\sqrt{N}}{2} = 80$		$k = \sqrt{N} = 160$		$k = 10 \cdot \sqrt{N} = 1600$	
	$\bar{x} \pm \sigma$	%s	$\bar{x} \pm \sigma$	%s	$\bar{x} \pm \sigma$	%s
HRC single	318±2582	28.7	159±1816	26.2	16±63	18.1
HRC complete	318±184	0.0	159±93	0.0	16±11	0.0
HRC median	318±847	0.0	159±333	0.0	16±23	5.7
HRC average	322±348	0.0	159±151	0.0	16±14	0.7
HRC centroid	322±2294	7.6	159±1436	6.9	16±25	7.4
HRC Ward	318±113	0.0	159±54	0.0	16±6	0.0
KMN	318±72	0.0	159±38	0.0	16±5	0.0
KMED	318±125	0.0	159±64	0.0	16±7	0.2
Binning	411±269	0.0	209±165	0.0	20±25	10.8
OV binning	509±555	0.0	173±201	0.7	20±25	9.8

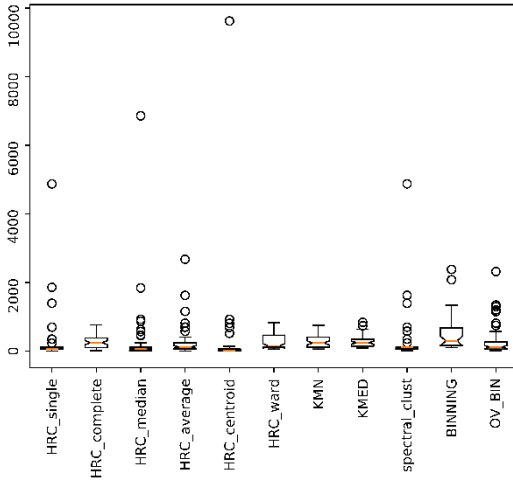
Hence, after choosing \sqrt{N} as the best clustering size for our study, the choice of the different clustering methodologies was needed so the population distribution was assessed for all the cases. The methods of choice were fixed to those algorithms giving an acceptable uniform compound distribution. Therefore, at the light of the results observed in Figure 2.6, HRC single, median, and centroid clustering methods were discarded as they tended to represent a unique or few hyper-populated clusters (represented as outliers of the corresponding boxplots) and a high number of singletons losing homogeneous representativeness of the chemical space. This distribution is the common result of many hierarchical agglomerative clustering with a bottom-up approach and was

observed for all the cases of study. Spectral clustering was also discarded as it could not be applied to all the databases due to not convergence drawbacks of the algorithm for some of the distributions.

Therefore, HRC average, HRC complete, and Ward were considered hierarchical agglomerative clustering methodologies, KMN and KMED as non-hierarchical relocation clustering, and OV binning as a representation of a cell-based partitioning method.

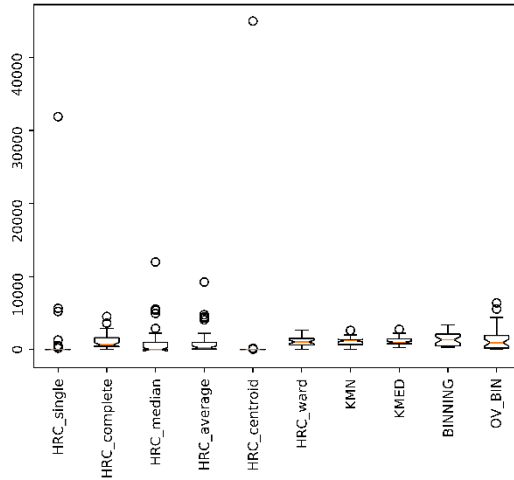
DACOMITINIB

Population Distribution



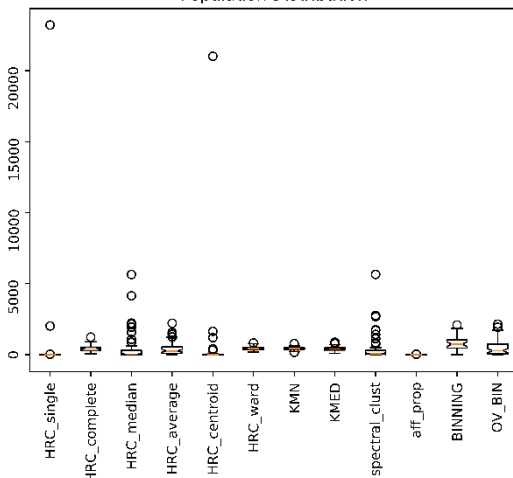
ABEMACICLIB

Population Distribution



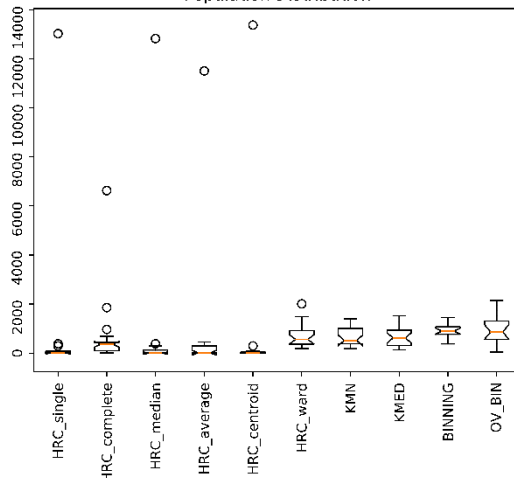
TAFENOQUINE

Population Distribution



ERTUGLIFLOZIN

Population Distribution



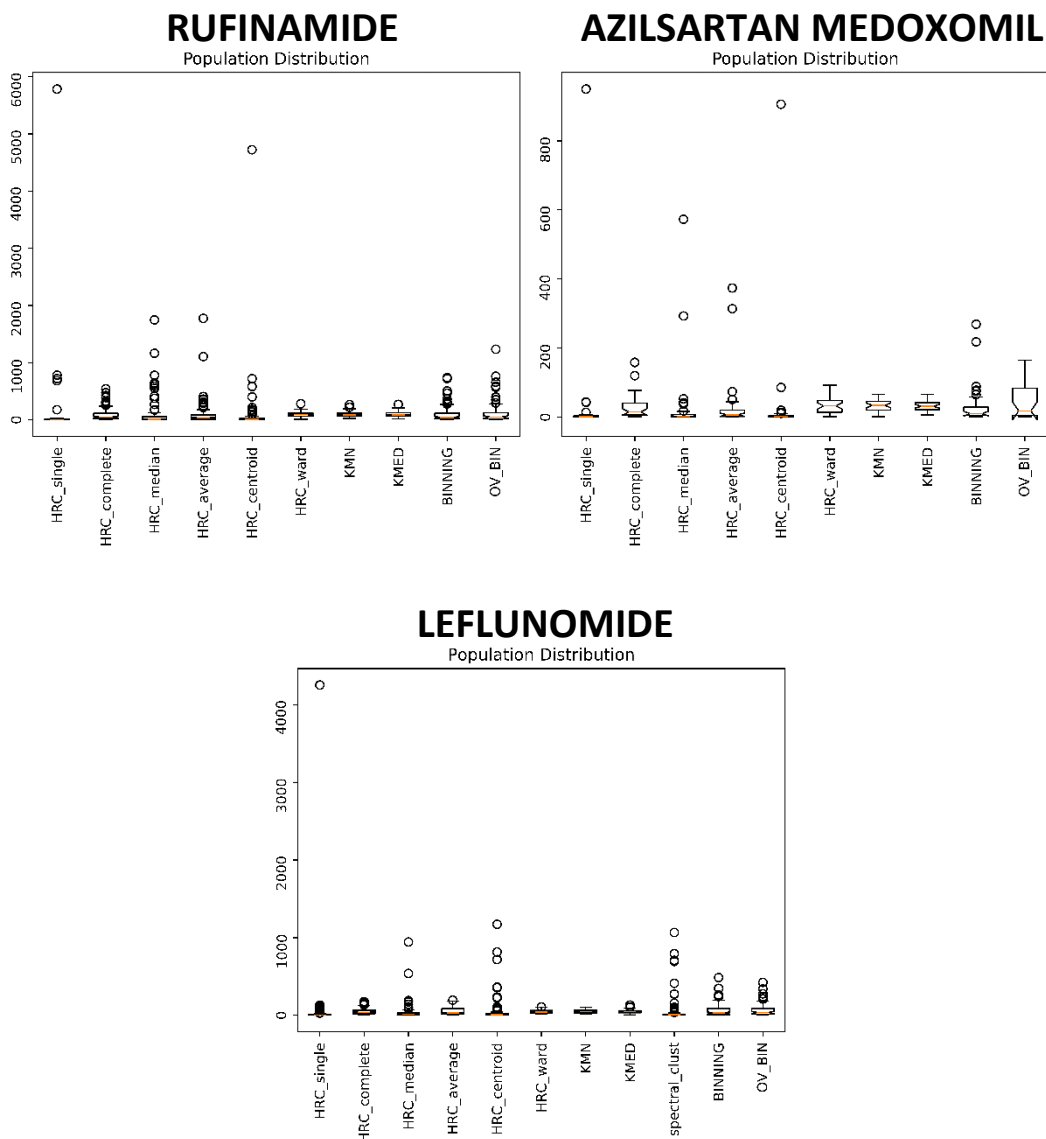


Figure 2.6 Boxplots of population distribution for the 6 libraries of analogs and \sqrt{N} clusters.

2.3.3. Assessing the explored space until date

2.3.3.1. The explored regions of ChEMBL's drugspace

With the purpose of visualizing the chemical subspaces' location in the predefined ChEMBL's *drugspace*, the combinatorial libraries were interpolated into the ChEMBL's PCs leading to the colored regions shown in Figure 2.7. Contrary to the scattering observed in the FDA's chemical space due to a categorization attending to their disease class, the sample distribution of libraries of analogs with a

common scaffold, are located highly focalized in the large drugspace described by the ChEMBL dataset. Moreover, as an overlap of these subspaces is noticeable, this would give us a hint that there are compounds with common drug-like properties between the combinatorial libraries of different drug derivatives that may leave the door open for further repurposing or multitarget studies.

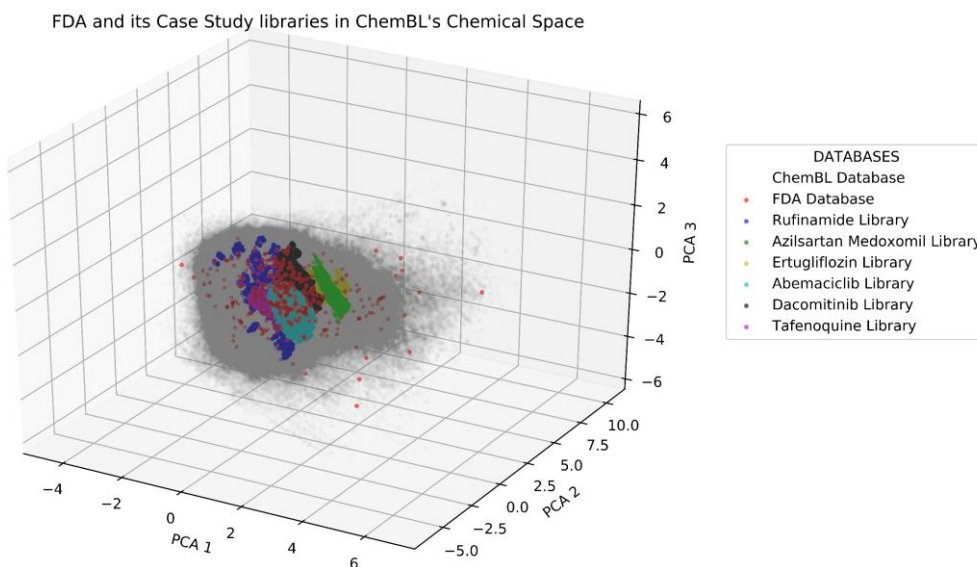


Figure 2.7. Drugspace described by the firsts 3 PCs (52% of variance explained) of the filtered database of ChEMBL with FDA data and the combinatorial libraries of drugs of study.

2.3.3.2. Bibliographical representativeness in each chemical space

In order to assess the real space explored until date in terms of molecular novelty, the search of the analogs described both in the patents and literature was enforced. Bibliographical databases (BD) refer to all the compounds derived from the Markush structure that have been reported as synthesized and, in some cases, biologically tested. To describe all the explored space known until the date for each patent, we include not only the individual molecules explicitly declared on the patent's claims but also all ulterior derivatives found in PubChem¹²¹ (even if their declared current application is different from the described in the original patent).

It is worth mentioning that in this study we have only used publicly available compounds. As an example, the set of 58 Tafenoquine analogs found in PubChem includes the 7 molecules claimed in patents^{105,109} and 51 molecules described in the literature (belonging to the same Markush structure). Among these novel molecules; 27 are included in other patents or articles concerning malaria¹²²⁻¹²⁷, 5 are related to parasitic diseases¹²⁸, 14 are present in a study of the inhibition of

human monoamine oxidase A¹²⁹, and 5 are structurally described but no biological activity was reported.

Regarding the rest of the analyzed patents, apart from the molecules stated in each patent, 10 Dacomitinib analogs, 27 Abemaciclib analogs, 51 Tafenoquine analogs, 2 Ertugliflozin analogs, 9 Rufinamide analogs and 102 Leflunomide analogs have been reported until date in literature. With the exception of Leflunomide, the number of molecules described are lower than the expected to significantly represent their Markush *drugspace* as it is shown in Table 2.7. Leflunomide's greater number of *BD* may be related to the year of the drug approval, as it matches with the eldest studied structure with an expired patent from 1978. It stands out in what manner the number of studied compounds (*BD* sets) differs excessively from the size of the library derived from the Markush structure (*MCL*). On the one hand, the low number of derivatives would evince the presence of synthetical limitations that hamper the obtention of more compounds.

On the other hand, it would confirm the application of a highly focused exploratory methodology. For this reason, we propose the creation of a second combinatorial database (named Bibliographic Combinatorial Library, *BCL*) for each patent by combining only the substituents present in *BD* aiming to represent the real combinatorial space synthetically accessible until the date.

Table 2.7 Information related to the drugs under study. The number of analogs included in *MCL* (*N*), in Bibliographical Data (*N_{BD}*) and in Bibliographical Combinatorial Library (*N_{BCL}*) and the root square of *N*.

Drug Name	<i>N</i>	<i>N_{BD}</i>	<i>N_{BCL}</i>	\sqrt{N}
Dacomitinib	16,530*	60	798	129
Abemaciclib	45,696	41	736	214
Tafenoquine	25,472	58	600	160
Ertugliflozin	14,194	21	56	120
Rufinamide	8,959	22	144	95
Azilsartan Medoxomil	1,110	4	9	34
Leflunomide	5,641	114	2,844	76

To assess the degree of representativeness of *BD* in the chemical space claimed in a patent (*MCL*), *SC* and *PC* values have been calculated when the *MCL* space is divided into *N_{BD}* partitions. Needless to say, an optimal space exploration such as the derived via rational selection would result in a 100% of both *SC* and *PC*. However, results show that a selection of an equal-sized set using random sampling (calculated as the mean value of the coverages obtained with 5,000 repetitions) has proved to better represent the chemical space than *BD* compounds (Table 2.8 and Table 2.9).

Table 2.8 Comparison of SC and PC values obtained by random selection and BD compounds when dividing the MCL of Dacomitinib, Abemaciclib and Tafenoquine in N_{BD} clusters.

		<i>Dacomitinib</i>		<i>Abemaciclib</i>		<i>Tafenoquine</i>	
		<i>BD</i>	<i>Random</i>	<i>BD</i>	<i>Random</i>	<i>BD</i>	<i>Random</i>
HRC average	SC	26.7	41.1	22.0	35.1	17.2	46.8
	PC	74.2	80.1	66.2	85.9	12.8	78.4
HRC Complete	SC	50.0	54.0	34.1	52.0	25.9	59.3
	PC	69.0	72.8	44.9	74.0	19.3	67.6
HRC Ward	SC	41.7	53.5	34.1	58.6	17.2	61.8
	PC	71.5	73.2	43.9	68.8	18.8	65.3
KMN	SC	48.3	57.6	34.1	60.3	22.4	62.5
	PC	67.3	69.4	39.1	67.4	21.0	64.7
KMED	SC	55.0	57.6	34.1	59.1	24.1	61.4
	PC	68.0	68.8	44.4	68.3	21.8	65.9
OV binning	SC	25.0	42.9	28.1	51.8	24.0	49.2
	PC	59.1	78.3	46.2	84.0	35.4	82.9

Table 2.9 Comparison of SC and PC values obtained by random selection and BD compounds when dividing the MCL of Ertugliflozin, Rufinamide, Azilsartan Medoxomil and Leflunomide in N_{BD} clusters.

		<i>Ertugliflozin</i>		<i>Rufinamide</i>		<i>Azilsartan Medoxomil</i>		<i>Leflunomide</i>	
		<i>BD</i>	<i>Random</i>	<i>BD</i>	<i>Random</i>	<i>BD</i>	<i>Random</i>	<i>BD</i>	<i>Random</i>
HRC average	SC	9.5	19.8	36.4	30.6	50.0	26.4	27.2	50.1
	PC	83.8	88.4	6.4	86.2	1.3	98.7	11.6	76.4
HRC Complete	SC	23.8	38.7	22.7	43.6	50.0	26.3	26.3	55.3
	PC	59.4	78.8	4.0	79.7	1.3	98.7	17.9	71.5
HRC Ward	SC	19.0	57.7	13.6	70.8	50.0	41.0	15.8	60.6
	PC	26.9	69.7	6.2	67.1	82.7	87.0	15.7	66.8
KMN	SC	23.8	58.7	18.2	58.9	50.0	52.1	15.8	60.8
	PC	23.5	69.3	7.9	69.4	59.9	84.8	13.1	66.7
KMED	SC	19.0	57.7	27.3	61.6	50.0	68.3	18.4	59.8
	PC	29.8	70.8	27.4	66.6	46.8	68.8	20.8	67.6

		<i>Ertugliflozin</i>		<i>Rufinamide</i>		<i>Azilsartan Medoxomil</i>		<i>Leflunomide</i>	
		<i>BD</i>	<i>Random</i>	<i>BD</i>	<i>Random</i>	<i>BD</i>	<i>Random</i>	<i>BD</i>	<i>Random</i>
OV binning	SC	13.3	65.6	28.6	50.3	25.0	56.6	26.4	53.0
	PC	18.2	82.2	9.6	90.1	1.7	80.2	32.5	83.0

For example, after dividing the Tafenoquine MCL space into 58 clusters, using the HRC average method, the 58 molecules included in BD showed 17.2% SC and 12.8% PC while a random selection was able to achieve 46.8% SC and 78.4% PC. This trend is observed regardless of the clustering or partition-based method used and in all the cases. Only the application of HRC average, HRC complete and HRC Ward cluster algorithms on Azilsartan Medoxomil led to better SC results than averaged random selections; although in this case results are clearly determined by the low number of compounds to select ($N_{BD} = 4$). Overall, these results evince that the chemical space defined in a drug patent is poorly described.

In contrast, the alternative suggested use of rational selection would always cover 100% of the current partitioned chemical space by selecting at least one molecule of each cluster. Then, following the above-mentioned example, a rational selection of 58 molecules (one per each cluster) would represent the 100% SC and PC of Tafenoquine's chemical space. However, this does not imply the synthetical feasibility of the chosen compounds, so a visual inspection by a medicinal chemist would be necessary for further steps.

2.3.3.3. Comparing BD and BCL coverage among the chemical space described by MCL

Quantifying the reduction of the MCL space that implies the use of BCL might be of key importance since its compounds are the ones genuinely expected to be synthetically feasible. This confirmation relies on the fact that BCL compounds include only the fragments coming from real studied candidates, constituting a more legitimate representation of the drug patent. In other terms, BCL dataset would describe the real Markush structure known until date so it would be the suitable structure to be presented in a fictitious patent claim at present.

For this purpose, the coverage of BD and BCL in MCL libraries were compared by assessing their distribution along with each principal component and their density plots when projected on the space of the first three principal components (Figure 2.8). As expected, BD and BCL progressively improve the description of MCL although neither of them, except for Dacomitinib and Leflunomide's analogs (later discussed), significantly cover the chemical space derived from the Markush structure.

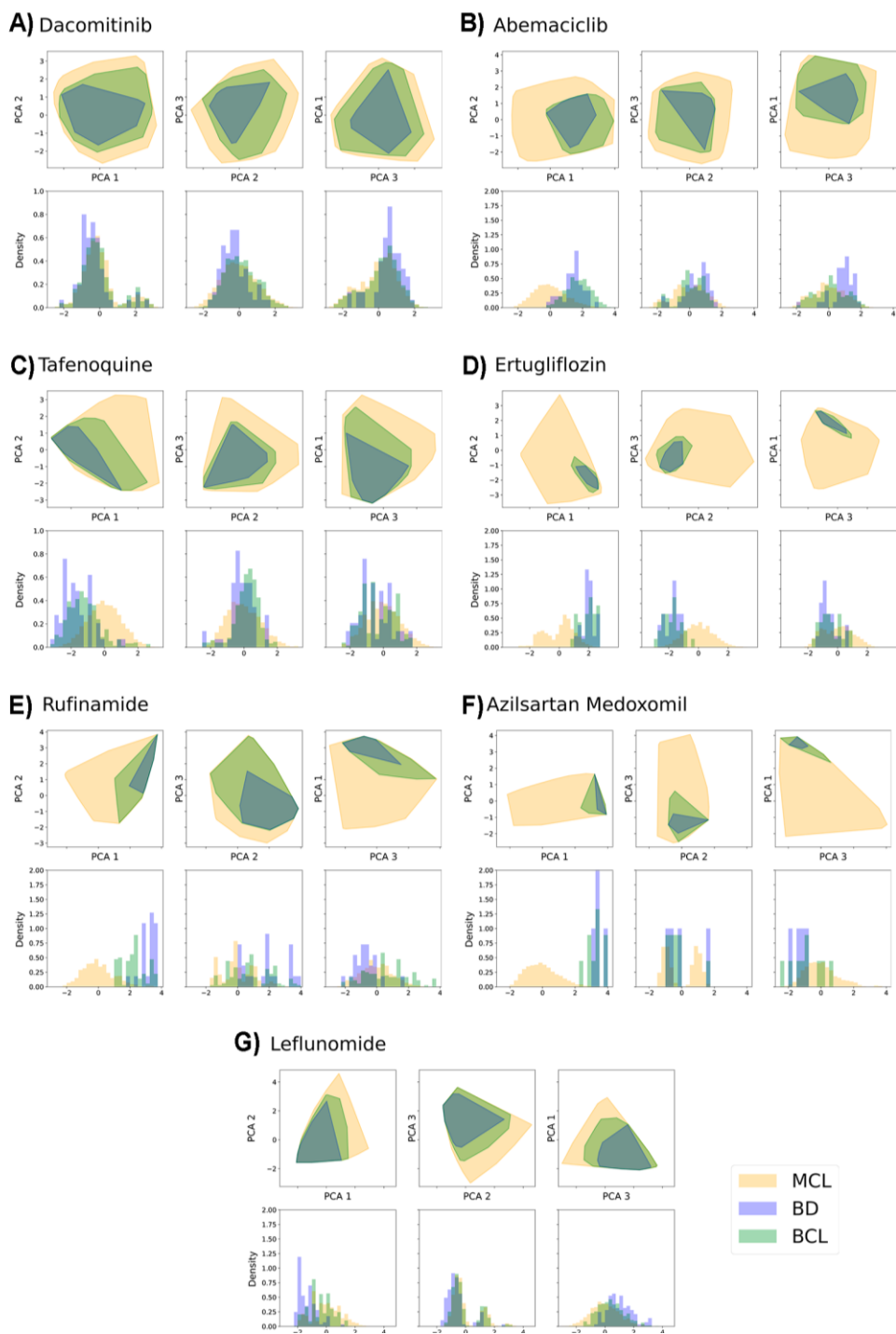


Figure 2.8 For each subfigure: in the first row there is the graphical representation of the scatter plot contour of MCL (pale yellow), BD (blue) and BCL (green) libraries when projected on PC1-PC2 (left), PC2-PC3 (middle) and PC3-PC1 (right) planes. In the second row, the density

histogram for the considered libraries along PC1 (left), PC2 (middle) and PC3 (right). This scheme is repeated for all patents: A) Dacomitinib, B) Abemaciclib, C) Tafenoquine, D) Ertugliflozin, E) Rufinamide, F) Azilsartan Medoxomil and G) Leflunomide.

To quantify this observation, cluster analysis was performed setting the number of clusters to a standard size of \sqrt{N} . Firstly, according to the results, aligned with the previous study, random selection once again showed to better represent the overall space rather than compounds derived from the current R&D methodology (represented as BD), even with a higher number of clusters. The results of all the libraries can be found from Table 2.10 to Table 2.16.

Firstly, in contrast to other libraries, Dacomitinib's results concerning the BD representativeness showed high SC and PC values (Table 2.10) and a very centered distribution (as seen in the previous Figure 2.8). This is explained by the reduction of the substituents considered in library enumeration (only those in claim 5, see section 2.3.1). This affected not only the MCL size (getting 16,530 compounds) but also the BCL (798 compounds). Consequently, BCL comprises a significant number of analogs of the used dataset of 16,530 compounds. So, much lower SC and PC values would be expected when considering the original database.

Table 2.10 SC and PC results for a number of $k = \sqrt{N}$ clusters for Dacomitinib analogs.

<i>Dacomitinib</i>	<i>#Analog</i> s	<i>HRC average</i>		<i>HRC complete</i>		<i>OV BIN</i>	
		<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>
Bibliographic data (BD)	60	20.9	50.3	27.1	42.1	16.8	52.0
Random (BD)	60	31.2	54.8	34.4	46.4	30.2	70.6
Bibliogr. Comb. data (BCL)	798	79.1	92.1	86.8	90.9	64.3	92.3
Random (BCL)	798	86.9	98.4	95.9	98.8	81.9	98.1
Random (\sqrt{N})	129	49.4	74.7	56.2	69.9	45.1	83.5

<i>Dacomitinib</i>	<i>#Analog</i> s	<i>HRC Ward</i>		<i>KMN</i>		<i>KMED</i>	
		<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>
Bibliographic data (BD)	60	29.5	38.4	27.9	35.8	31.0	43.0
Random (BD)	60	35.9	46.4	30.1	41.6	35.2	44.2
Bibliogr. Comb. data (BCL)	798	92.3	92.3	90.7	92.4	86.8	90.8
Random (BCL)	798	98.5	99.2	98.6	99.3	97.4	99.0
Random (\sqrt{N})	129	59.6	67.1	60.1	66.7	58.0	68.5

The results showed for Abemaciclib database in Table 2.11 are totally in concordance with the general observed trend. As BD analogs are smaller than \sqrt{N} , both its coverage and its combinatorial database (BCL) coverage of the chemical space described by MCL leaves much to be desired in comparison to a random choice of BD and BCL sampling which are more representative. In

fact, even a random choice of \sqrt{N} samples (which is significantly smaller than BCL) is more representative in terms of SC and PC.

Table 2.11 SC and PC results for a number of $k=\sqrt{N}$ clusters for Abemaciclib analogs.

<i>Abemaciclib</i>	#Analog	<i>HRC average</i>		<i>HRC complete</i>		<i>OV BIN</i>	
		<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>
Bibliographic data (BD)	41	4.7	63.8	9.8	9.9	15.9	23.7
Random (BD)	41	6.9	81.4	16.7	24.6	23.2	50.8
Bibliogr. Comb. data (BCL)	736	26.6	31.8	30.4	27.5	46.8	51.0
Random (BCL)	736	64.9	95.3	85.2	95.2	74.0	97.9
Random (\sqrt{N})	214	42.9	81.2	55.1	71.2	54.3	89.0

<i>Abemaciclib</i>	#Analog	<i>HRC Ward</i>		<i>KMN</i>		<i>KMED</i>	
		<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>
Bibliographic data (BD)	41	10.3	10.1	10.3	11.0	10.7	14.7
Random (BD)	41	17.2	19.7	17.3	19.0	17.1	20.9
Bibliogr. Comb. data (BCL)	736	24.3	24.2	25.7	24.4	27.1	33.0
Random (BCL)	736	93.3	96.0	94.3	96.3	92.1	95.6
Random (\sqrt{N})	214	60.5	66.2	61.4	65.4	59.3	67.3

The results showed in Table 2.12 for Tafenoquine's database are in agreement with the observations detailed above for Abemaciclib case of study.

Table 2.12 SC and PC results for a number of $k=\sqrt{N}$ clusters for Tafenoquine analogs.

<i>Tafenoquine</i>	#Analog	<i>HRC average</i>		<i>HRC complete</i>		<i>OV BIN</i>	
		<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>
Bibliographic data (BD)	58	8.8	6.4	11.9	9.7	13.6	17.0
Random (BD)	58	27.0	45.4	29.1	37.1	27.1	53.6
Bibliogr. Comb. data (BCL)	600	38.8	33.2	33.8	28.3	36.0	49.3
Random (BCL)	600	80.4	95.6	90.6	96.1	74.7	96.1
Random (\sqrt{N})	160	50.9	74.8	57.7	69.1	48.2	79.8

<i>Tafenoquine</i>	#Analog	<i>HRC Ward</i>		<i>KMN</i>		<i>KMED</i>	
		<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>
Bibliographic data (BD)	58	7.5	7.0	9.4	8.0	10.0	8.6
Random (BD)	58	30.0	32.9	30.3	31.7	20.3	45.4

<i>Tafenoquine</i>	#Analog	<i>HRC Ward</i>		<i>KMN</i>		<i>KMED</i>	
		<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>
Bibliogr. Comb. data (BCL)	600	26.9	27.7	37.5	36.0	44.4	41.2
Random (BCL)	600	95.9	97.0	96.8	97.4	94.7	96.8
Random (\sqrt{N})	160	61.4	65.4	62.4	64.4	60.6	66.2

Ertugliflozin's database was the only case where KMED clustering could not be calculated due to its particular chemical space distribution which did not present a convergence for the algorithm. Overall, the results showed in Table 2.13 for Ertugliflozin's database are in agreement with the observations detailed above.

Table 2.13 SC and PC results for a number of $k = \sqrt{N}$ clusters for Ertugliflozin analogs.

<i>Ertugliflozin</i>	#Analog	<i>HRC average</i>		<i>HRC complete</i>		<i>OV BIN</i>	
		<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>
Bibliographic data (BD)	21	8.3	12.2	6.7	9.2	6.8	7.8
Random (BD)	21	13.4	41.5	14.7	25.4	20.2	51.0
Bibliogr. Comb. data (BCL)	56	12.5	17.7	10.8	12.3	8.1	8.0
Random (BCL)	56	26.5	65.3	30.8	56.5	36.3	76.5
Random (\sqrt{N})	120	50.9	74.8	57.7	69.1	48.2	79.8

<i>Ertugliflozin</i>	#Analog	<i>HRC Ward</i>		<i>KMN</i>		<i>KMED</i>	
		<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>
Bibliographic data (BD)	21	7.5	8.0	8.3	8.5	-	-
Random (BD)	21	15.9	18.6	15.9	17.9	-	-
Bibliogr. Comb. data (BCL)	56	10.0	10.5	11.7	11.9	-	-
Random (BCL)	56	36.2	41.5	36.5	40.4	-	-
Random (\sqrt{N})	120	60.2	66.6	61.1	65.9	-	-

Rufinamide's small BD and BCL coverages can be explained by the focalized distribution presented in the above introduced Figure 2.8. Moreover, the results are in agreement with the population distribution in their clusters (see Figure 2.6) which show a few outlying overpopulated clusters for some clustering methods such as HRC average and HRC complete. However, as for the other introduced cases, the results showed in Table 2.14 for Rufinamide's database agree with the general observed trend.

Table 2.14 SC and PC results for a number of $k = \sqrt{N}$ clusters for Rufinamide analogs.

<i>Rufinamide</i>	#Analog	<i>HRC average</i>		<i>HRC complete</i>		<i>OV BIN</i>	
		<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>
Bibliographic data (BD)	22	11.6	2.0	7.4	1.1	8.4	2.2
Random (BD)	22	15.0	54.1	18.3	39.1	20.5	56.5
Bibliogr. Comb. data (BCL)	144	28.4	8.3	23.2	8.3	31.0	9.7
Random (BCL)	144	46.0	86.6	56.1	84.8	54.0	90.4
Random (\sqrt{N})	95	37.6	80.7	46.5	77.3	45.8	85.5

<i>Rufinamide</i>	#Analog	<i>HRC Ward</i>		<i>KMN</i>		<i>KMED</i>	
		<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>
Bibliographic data (BD)	22	5.3	3.2	6.3	2.7	4.2	2.9
Random (BD)	22	20.3	24.5	20.3	24.7	20.3	25.0
Bibliogr. Comb. data (BCL)	144	12.6	8.7	15.8	8.7	15.8	11.8
Random (BCL)	144	72.9	80.4	73.0	80.1	72.1	80.4
Random (\sqrt{N})	95	59.6	67.5	59.4	67.3	59.1	67.7

Again, bibliographical databases only show better results in the HRC average method applied to Azilsartan Medoxomil, due to its size (Table 2.15).

Table 2.15 SC and PC results for a number of $k = \sqrt{N}$ clusters for Azilsartan Medoxomil analogs.

<i>Azilsartan Medoxomil</i>	#Analog	<i>HRC average</i>		<i>HRC complete</i>		<i>OV BIN</i>	
		<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>
Bibliographic data (BD)	4	11.8	0.5	8.8	0.5	13.0	0.5
Random (BD)	4	8.8	52.7	10.5	25.4	15.0	33.6
Bibliogr. Comb. data (BCL)	9	26.5	1.9	20.6	5.6	26.1	3.9
Random (BCL)	9	14.5	69.3	20.3	45.4	27.1	57.6
Random (\sqrt{N})	34	30.5	84.8	45.6	79.2	51.0	88.8

<i>Azilsartan Medoxomil</i>	#Analog	<i>HRC Ward</i>		<i>KMN</i>		<i>KMED</i>	
		<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>
Bibliographic data (BD)	4	8.8	1.3	8.8	1.3	5.9	1.3
Random (BD)	4	11.0	15.7	11.1	14.5	11.1	13.5
Bibliogr. Comb. data (BCL)	9	20.6	10.6	20.6	11.2	8.8	3.3
Random (BCL)	9	22.6	31.6	22.8	29.5	23.0	27.6

<i>Azilsartan Medoxomil</i>	#Analogos	<i>HRC Ward</i>		<i>KMN</i>		<i>KMED</i>	
		<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>
Random (\sqrt{N})	34	56.1	72.6	57.8	71.4	60.3	68.4

Additionally, Leflunomide (Table 2.16) is the unique example in which the number of molecules found BD (N_{BD}) was greater than \sqrt{N} , so the size of its combinatorial sub-library (*BCL*) had better representativeness in its chemical space for almost all cases (with the exception HRC average PC value) of rather than a random choice of \sqrt{N} analogs.

Table 2.16 SC and PC results for a number of $k = \sqrt{N}$ clusters for Leflunomide analogs.

<i>Leflunomide</i>	#Analogos	<i>HRC average</i>		<i>HRC complete</i>		<i>OV BIN</i>	
		<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>
Bibliographic data (BD)	114	28.9	16.2	34.2	21.3	29.2	36.2
Random (BD)	114	62.8	84.7	65.5	82.7	59.9	87.7
Bibliogr. Comb. data (BCL)	2,844	76.3	67.5	84.2	77.3	70.8	86.2
Random (BCL)	2,844	99.0	100.0	99.6	100.0	98.0	99.9
Random (\sqrt{N})	76	52.8	74.7	54.3	72.4	50.8	80.8

<i>Leflunomide</i>	#Analogos	<i>HRC Ward</i>		<i>KMN</i>		<i>KMED</i>	
		<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>	<i>SC</i>	<i>PC</i>
Bibliographic data (BD)	114	22.4	16.4	22.4	17.9	22.4	23.3
Random (BD)	114	74.2	79.8	74.7	79.7	74.1	79.9
Bibliogr. Comb. data (BCL)	2,844	76.3	74.4	75.0	73.7	86.8	88.7
Random (BCL)	2,844	100.0	100.0	100.0	100.0	100.0	100.0
Random (\sqrt{N})	76	60.7	66.8	61.1	66.3	60.7	66.8

Overall, although having changed the number of partitions, results agree with the ones described in the previous section: both SC and PC percentages obtained using a random selection are better than those obtained with the bibliographic database, either BD or BCL. Generally, contrasting the obtained results with \sqrt{N} and N_{BD} partitioned space one may estimate an optimal and synthetically manageable number of samples that could be chosen to be synthesized.

Although the BCL library includes a higher number of compounds, in many cases, it is not able to represent the MCL space hence, this result would serve as evidence that the synthetically accessible chemical space for each dataset is still poorly known as the rational R&D methodology has followed a mainly focused trend around some original hits. In fact, the use of a rational selection of

BD compounds would increase the efficiency of both traditional and cherry-picking methodologies in terms of coverage. As an example, when applying a rational selection of 58 Tafenoquine analogs, an optimal value of 36.3% of SC could be achieved (58 out of 160 clusters).

2.4. Rational selection; an alternative towards a more efficient hit-to-lead strategy

In light of the results of our study, enough evidence has been exposed to prove the lack of chemical space exploration in the traditional R&D methodology. This is the result of a procedure that relies in the hit-to-lead optimization step which commonly aims to find an optimal compound around the original hit, commonly involving a Free-Wilson approach during the process (Figure 2.9). The synthesis and biological evaluation of these analogs is performed leading to a drug candidate and, accordingly, a Markush structure is settled defending that its involved analogs may present the same biological behavior as the original hit. In fact, even the fragment combination of the reported structures (BCL) does not represent properly the drug's chemical space derived from the Markush structure (MCL).

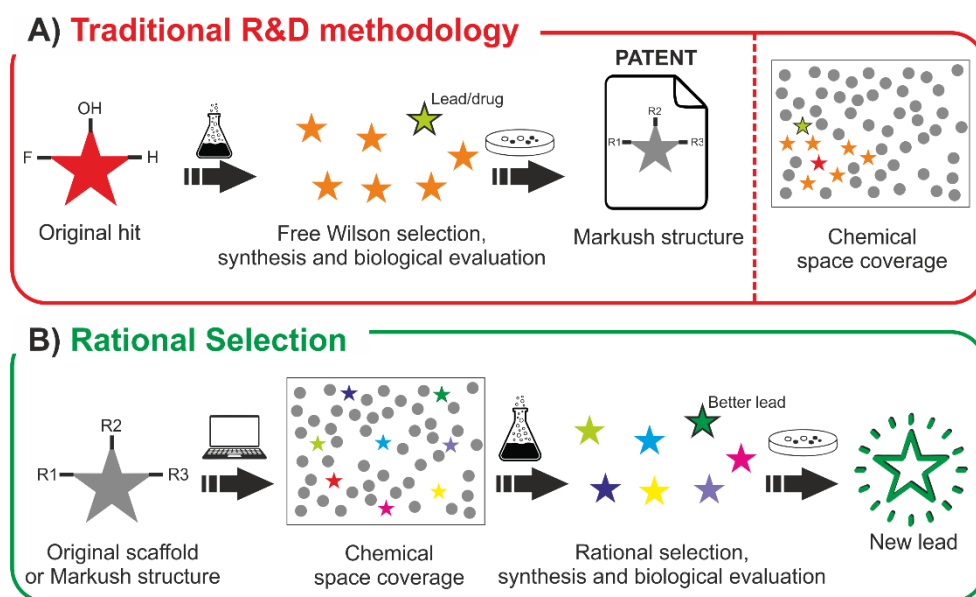


Figure 2.9 Hit-to-lead process comparative workflow between the traditional methodology (A) and the approach suggested by the authors after the reported study (B).

Hence, we defend an alternative or complementary approach which departs from the combinatorial library obtained from a theoretical Markush structure or from the fragment combination of an original scaffold explored in previous studies (which would ensure the synthetical feasibility of its analogs). Secondly, a computational study, involving the space clustering or partitioning, is suggested to rationally choose a handleable number of compounds to synthesize and

test that may unveil a better lead in unexplored regions. This strategy can be reproduced in an iterative process. Consequently, if the most active cluster matches with the original hit subspace, a later optimization process around its surrounding area would be also applied in the search for new leads. Thus, this new methodology would better consider the chemical diversity of the original Markush set, being the rational selected compounds a significant representation for the issue of study or further repurposing approaches.

2.5. Discussion and conclusions

The hit-to-lead process in drug discovery has been traditionally based on the application of the Free-Wilson approach according to which, after hit identification, the structure of the drug candidate is progressively modified attempting to improve its biological activity. Hence, the resulting procedure allows for exploring the surrounding chemical space of the initial hit compound but there could be regions that remain unexplored, compromising the R&D efficiency.

We have assessed how well the chemical space claimed in a patent is actually explored, using seven patents as examples. For all cases, the space explored in the literature (BD) for each combinatorial library is very small, about 20% on average when clustering the chemical space in \sqrt{N} clusters. Moreover, results show that even a random selection (by cherry-picking) may lead to a better coverage than the molecules reported in the literature. These results are in agreement with results previously reported by our research group regarding the study of the chemical space described by HEPT analogs^{41,91}.

It has also been evidenced that, in most cases (5 out of 7), even the synthetically accessible combinatorial library (BCL), resulting from the fragmental combination of the molecules described in the literature, are not entirely representative of the chemical space (with lower values than a random study of \sqrt{N} molecules).

Neither the real space explored (BD) nor the fragment combination of the studied molecules (BCL) significantly represent the space defined by the combinatorial libraries derived from the Markush structure, especially when they are compared with the coverage obtained by a statistically random sampling of \sqrt{N} molecules. Thus, there is a large part of the chemical space claimed on patents that remains unexplored and it can hide potential leads that may surpass the activity of the original hit or reduce undesired side effects. Rational selection algorithms could assist traditional methodology to optimize the selection of representative compounds of a given chemical space. This could be applied to explore many pharmacokinetic profiles such as toxicity, biological activity, or solubility among others.

Results reinforce the proposal to integrate the rational selection in the R&D process in early drug discovery or combine it in a mixed methodology involving a local optimization around the

original hit or new lead candidates. Nevertheless, it should be noted that many big libraries derived from a Markush structure may present candidates with problematic or even unfeasible synthesis and, hence, proper data curation is mandatory before proceeding to a definitive rational selection.

Chapter 3: Tafenoquine. Study of 8-aminoquinolines as potential antimalarial drug candidates

3.1. Introduction to malaria and Tafenoquine

3.1.1. Malaria

Malaria is an infectious disease caused by a protozoa parasite of the genus *Plasmodium* and it is transmitted by female *Anopheles* mosquitoes. In humans, malaria is caused by five species: *P. falciparum*, *P. vivax*, *P. malariae*, *P. ovale* and *P. knowlesi*. While the first is the deadliest malaria parasite causing the highest number of deaths worldwide and the most prevalent in the African continent, the *P. vivax* transmission occurs across the tropics and reaches into subtropical and temperate climate. Thus, *P. vivax* is the most predominant in most countries outside of sub-saharian Africa, such as India, Afghanistan, Pakistan, Brazil, or Guatemala.

This life-threatening disease caused an estimated 241 million cases and 627,000 related deaths in 2020 among 85 countries (see Figure 3.1), being Africa the most affected continent¹³⁰. In short, malaria is a constant dangerous threat for approximately half of the globe's population, while among the other half, there are many people who travel to areas where malaria is endemic for many reasons. Consequently, the World Health Organization (WHO) has prioritized the reduction of malaria case reduction with the objective of lowering morbidity and mortality associated with this disease by 90% by the year 2030. Such goal requires an approach to greatly reduce its transmission in endemic populations primarily by using drugs for both prophylaxis and/ or treatment along with vector control and early diagnosis procedures.

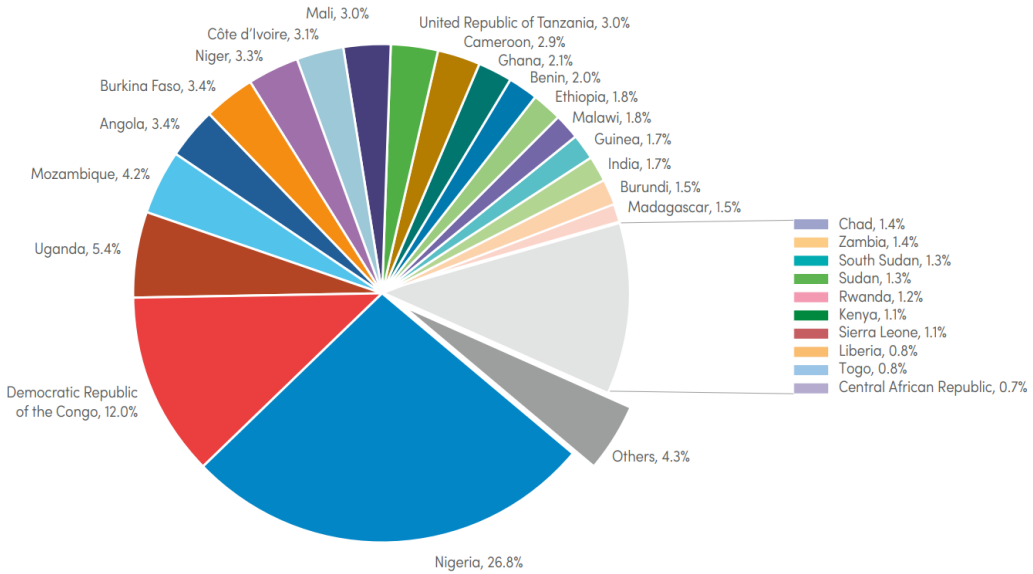


Figure 3.1. Countries that were malaria endemic in 2020. Source: World Malaria Report 2021 ¹³⁰

3.1.2. The *Plasmodium* life cycle

In order to comprehend the underlying mechanism of the drugs currently in use for malaria treatment, the understanding of the infection process in the host is needed.

The infection departs from a female *Anopheles* mosquito (the vector host) bite to a human. This simple and casual process injects the infectious sporozoite form of the parasite to the host from the vector’s salivary gland. Once the parasite infects the host, the sporozoites migrates shortly to the liver hepatocytes (Figure 3.2) entering into an incubation period consisting in growth, segmentation and finally sporulation (exoerythrocytic cycle) that might last from 7 to 30 days depending on the specie where the parasite resides while continuing a life cycle producing mitotically high numbers of schizonts. Finally, the mature form of the parasite leaves the liver as merozoites to invade the bloodstream.

After the infection of the red blood cells by the merozoites, the erythrocytic cycle takes place producing new forms as blood schizonts which ultimately burst releasing merozoites which will reenter erythrocytes whilst producing high fevers in the host. This cycle is continuously repeated in a course of single malaria episode until is suppressed by the host’s immunity system or by drugs. During this asexual phase (schizogony), an average of 8-32 new merozoites and gametocytes per infected red blood cells are produced. Finally, when the female *Anopheles* mosquitoes feed on blood of the new infected host, they ingest the *gametocytes* and these later begin the sporogonic phase (sexual phase) in the gut and migrate to the salivary glands perpetuating the infection^{131–135}.

During the first infective period (exoerythrocytic cycle), no symptoms may be manifested. However, after the incubation period, the release and circulation of malaria parasites into de bloodstream as merozoites may result in fever, sweats, nausea, headaches, nausea, vomiting or body aches are typically to take place. When the patient receives the right treatment at this stage, the prognosis is great, but when *P. falciparum* infections are left untreated, they might develop into severe malaria, which may be lethal in the majority of cases if no cure is supplied. Occasionally, different level of mental disorientation, coughs or muscle pains may occur. Hence, malaria treatment relies predominantly on drugs that target the disease-causing asexual blood stages.

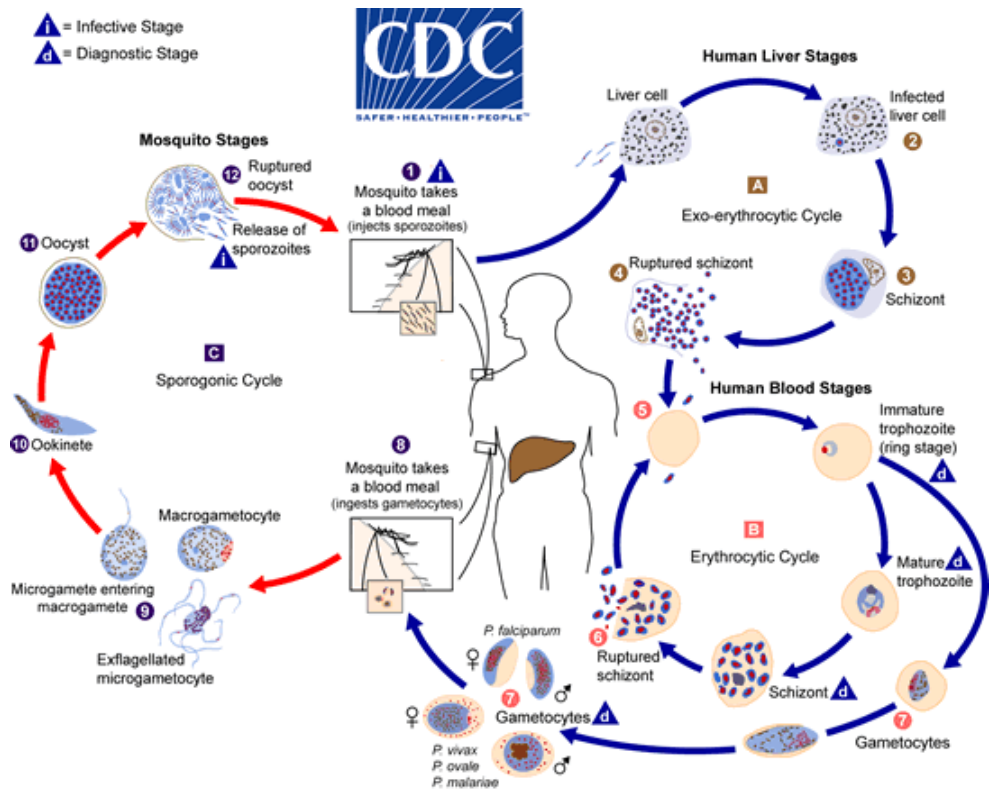


Figure 3.2 Life cycle of the Plasmodium parasite. Source: CDC, 2020¹³⁶

More specifically, unlike *P. falciparum*, *P. vivax* and *P. ovale* have a typically low blood-stage parasitemia with gametocytes emerging before illness manifests. Their hepatic schizonts hibernate before migrating to the bloodstream in dormant liver stages, called hypnozoites. These are capable to reactivate the process and cause the relapse after weeks, months or even years after the first episode. During this period, the parasite remains undetectable by the human immune system or any current diagnostic technique complicating the prophylaxis and treatment. Indeed, the global burden of infection and disease imposed by endemic *P. vivax* transmission is complicated by its biology as

both active and latent infection take place in difficult-to-access tissues. Consequently, all the described characteristics affect both *P. vivax* geographic distribution and transmission patterns as shown in Figure 3.3.

To boost malaria elimination, countries need to focus on vector control measures and aggressive treatments to prevent the transmission and relapses due to *P. vivax* as it is considered the most widespread specie around the globe.

However, the study of malaria seems not to be interesting for pharmaceutical industries as it does not provide financial incentive mainly due to their prevalence in developing countries rather than in the developed world. The development of a drug intended to treat this disease would not allow the recovery of the capital invested during its research. Thus, it is considered an orphan disease and the indication of its drugs are also considered as orphan¹³⁷. However, coordinated WHO global efforts have been adopted since the World Health Assembly in may 2015 to develop the WHO Global Malaria Programme to control and eliminate malaria in the period from 2016 to 2030¹³⁸.

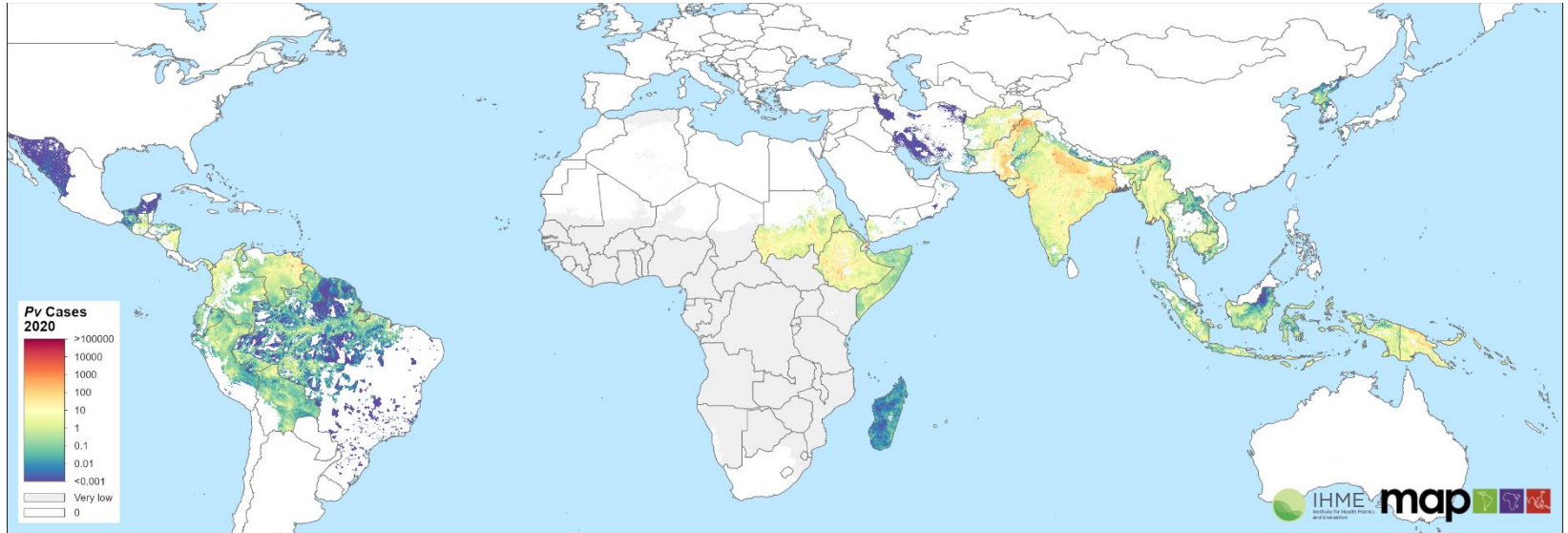


Figure 3.3 The global clinical cases (per 1,000 population) in all ages of patent *Plasmodium vivax*, 2020. The numbers of cases predicted to occur in each 5 x 5 km² pixel are shown on a spectrum of blue to red. Areas where *P. vivax* is known to be endemic, but there was not sufficient information to generate a prediction, are shown in light grey. Global national shapefile obtained from the Malaria Atlas Project (MAP: <https://malariaatlas.org/>)

3.1.3. *P. vivax* treatment. The 8-aminoquinoline antimalarial drugs

Malaria treatment depends on many factors such as the species of malaria parasite causing the infection, disease severity, and the geographical region where the host was infected. Even more, the latter two attributes help determine the probability of the developed resistance of the parasite to certain antimalarial drugs. Further factors such as weight, age or pregnancy condition might also limit the available options for malaria treatment.

A wide variety of preventing drugs for malaria disease have been discovered since quinine was first isolated from the chinchona tree and used for its treatment in the early 1800s. After the depiction of *Plasmodium* life cycle, several drugs or drugs combination serving as chemotherapy were developed to treat the different stages of the parasite development. There are four possible strategies for drug therapy¹⁰:

1. Kill the sporozoites injected by the mosquito or prevent its entrance into the liver.
2. Kill the schizonts residing in hepatocytes or prevent their development to merozoites. This mechanism is the most common use and its drugs are considered as blood schizontocidal.
3. Kill the merozoites in the red blood cells or prevent their conversion to gametocytes.
4. Kill the gametocytes before entering in other *Anopheles* female mosquito. The drugs with this mechanism of action are considered gametocidal.

Present antimalarials can be categorized into multiple classes such as: aryl-amino alcohols, artemisinins, antifolates, antibiotics, 4-aminoquinolines (4AQ) and 8-aminoquinolines (8AQ). Since 2019, six drugs have been approved by the FDA and are currently available by prescription for malaria prophylaxis. In fact, several are in widespread use today. They are, in terms of the year of approval: chloroquine, primaquine, mefloquine, doxycycline, the combination of atovaquone and proguanil, and tafenoquine¹³⁹. Among these, primaquine and tafenoquine are the currently 8AQ drugs available of choice against *P.vivax* and relapsing forms of malaria, and chloroquine the 4AQ to treat uncomplicated *vivax* malaria¹⁴⁰ (Figure 3.4).

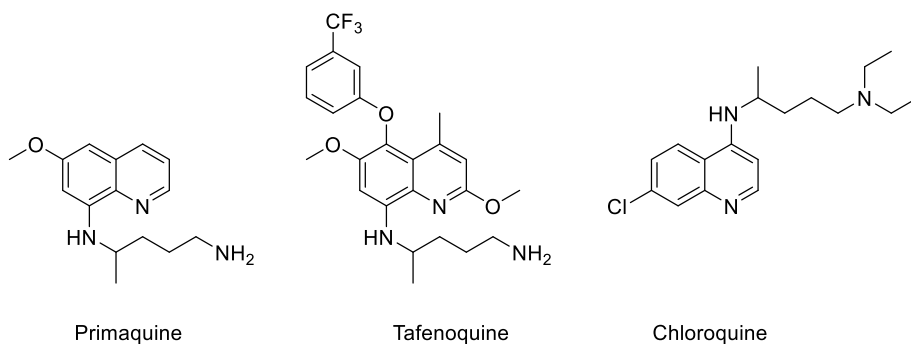


Figure 3.4 Quinolinic based drugs currently in use against *P.vivax*

Specifically, Tafenoquine is a single-dose 8-aminoquinoline derivative that was approved by the FDA in 2018 and was termed as the “radical cure for *P. vivax* malaria”. It has been proved to be the most potent and less toxic 8AQ analogue against this specie. Although, this compound shows great inhibitory activity for both blood and liver stages of *P. falciparum* and *P. vivax*, its main interest relies in the relapse prevention for the latter. Indeed, the challenges in controlling and eliminating *vivax* malaria are normally related to its ability to relapse from long-lasting formant liver stages (as hypnozoites) and its high transmission capability caused by the continuous production of gametocytes along with the shorter growth cycle in the vector host compared to other *Plasmodium* species¹⁴¹.

This drug belongs to 8-aminoquinoline (8AQ) antimalarial drugs, it was first synthesized and named as WR238605 by scientists at the Walter Reed Army Institute of Research in 1978 as a part of the US antimalarial drug program started in 1963^{142–144}. Its first research attempts explored through preclinical studies revealed its potent tissue schizonticide activity, its good oral bioavailability, its reduced toxicity and its longer half-life in comparison which its precursor, primaquine^{145–149}. GlaxoSmithKline’s legacy in the research and development of tafenoquine as a likely drug for malaria started 20 years ago. Later, in 2008, GSK entered into a collaboration with the not-for-profit drug research partnership Medicines for Malaria Venture (MMV) to develop this drug as an anti-relapse medicine for patients infected with *P. vivax*¹⁵⁰. In fact, Tafenoquine’s properties totally fit with MMV requirements. As an example, liver schizonticidal activity is a key factor of next generation antimalarials as proposed by MMV, other critical components include prophylactic and chemoprotective liver stage efficacy when single exposure¹⁵¹. Still, the precise mechanism of action of tafenoquine is yet unknown at present, as is so for primaquine¹⁵².

Nevertheless, as a primaquine analog, it has inherited one of the main drawbacks of its predecessor causing severe hemolytic anemia in people with glucose-6-phosphate dehydrogenase (G6PD) deficiency^{152–154} which may result in hemolytic anemia (breakdown of the red blood cells)

which may cause arrhythmias, cardiopathies or even heart failure. Consequently, as no robust test has been developed to quantitatively measure the G6PD levels in endemic areas, tafenoquine administration has not been widely adopted. Suitably, an alternative to this original hit would be necessary to lower this secondary effect and optimally find a more active drug candidate.

3.2. Chemical space assessment of Tafenoquine's analogs

3.2.1. Tafenoquine library of analogs and its chemical space

When looking into the patents containing Tafenoquine's structure such as its already expired original patent from 1986 (US4617394¹⁰⁹) or a latter from 2002 (US6376511¹⁵⁵), it can easily be found the common Markush structures in their concluding claims section.

Through the previous discussed chapter (see Chapter 2) we realized and first proved that although the patents tend to describe wide chemical spaces through Markush structures stated in claims, the optimization of the new principal active ingredient developed – in the case of study; Tafenoquine – is frequently driven by a simple Free-Wilson approach. This procedure leads to a highly focused study on the chemical space nearby a hit compound leaving many regions unexplored which may present highly biological active reservoirs.

In this case of study, the enumeration of the 8AQ analogs was performed as a combination of both of the Markush structures stated on the above-mentioned patents. Both patents' Markush structures share the same 8AQ scaffold and differs from the R₂ substituent and the number of substituents in the ring in C5 position. Plus, the additional valuable information added by the most recent patent is that it contains 7 specifically described molecules and hence, protected that are part of the BD dataset. The resulting Markush structure is described in Figure 3.5 and its corresponding list of substituents is showed in Table 3.1.

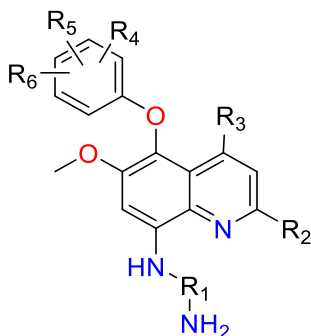


Figure 3.5 Markush structure of Tafenoquine analogs

The complete combinatorial library of 8AQ analogs, consists in 25,472 compounds obtained as a result of 8x2x2x796 possible combinations, being 796 all the possible phenolic structures (containing R4, R5 and R6 in the 5 available ring positions) for the C5 quinolinic site.

Table 3.1 Fragments in SMILES used to build the 25,472 compounds library of Tafenoquine analogs.

R1	R2	R3	R4	R5	R6
-C(C)CCC-	-H	-H	-H	-H	-H
-C(C)CCCC-	-OC	-C	-Cl	-Cl	-Cl
-C(CC)CCC-			-Br	-Br	-Br
-C(CC)CCCC-			-F	-F	-F
-CCCC(C)-			-C(F)(F)F	-C(F)(F)F	-C(F)(F)F
-CCCCC(C)-			-OC	-OC	-OC
-CCCC(CC)-					
-CCCCC(CC)-					

The database was desalted, protonated at pH 7, and their partial charges were calculated using MMFF94x forcefield. The total 206 1D and 2D molecular descriptors available in MOE were calculated (the complete list of molecular descriptors is available in the Annex II) and the resulting matrix of 25,472 x 206 was reduced through PCA. The final chemical space was lowered to 13 PCs which explained the 95.5% of the cumulative variance and ultimately explored through different clustering and partitioning techniques. The overall process can be resumed as shown in Figure 3.6.

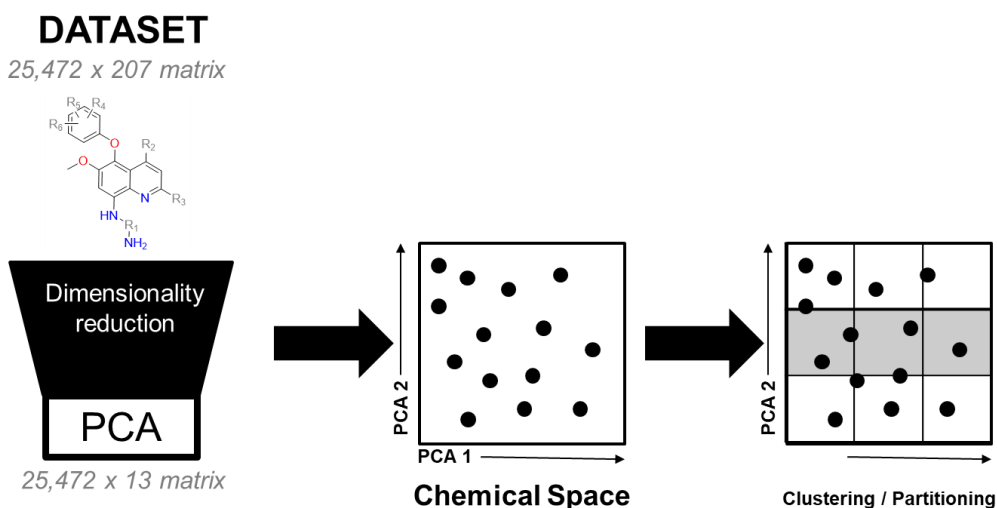


Figure 3.6. Overall process for Chemical Space description and assessment.

3.2.2. Choice of a suitable clustering methodology

In order to perform a further rational selection of a tractable number of representative molecules to be synthesized in the laboratory, a first assessment of the chemical space and choice of the suitable clustering methodology is needed. In our case, Tafenoquine's chemical space was firstly assessed attending to the classification of the data in $\sqrt{N} \approx 160$ clusters. It was studied the data distribution for nine clustering and two partitioning algorithms: hierarchical agglomerative clustering (HRC) with six linkage methods (single, complete, median, average, centroid and ward), *k*-means (KMN), *k*-medoid (KMED), spectral clustering (spectral clust) simple binning (BINNING) and the developed algorithm of optimum variance binning (OV BIN). The last two partitioning methodologies were able to group the data in 122 and 147 occupied bins respectively.

In order to contrast all the classificatory methodologies, the centroid of each cluster/bin was picked leading a selection of 160 candidates for each case – with the exception of the 122 and 147 compounds selection for partitioning methodologies– and the values of space and population coverages was calculated when overlapping these subselections into the other clusters resulting from the different algorithms. The obtained results are shown through two heatmap plots showed in Figure 3.7 and Figure 3.8.

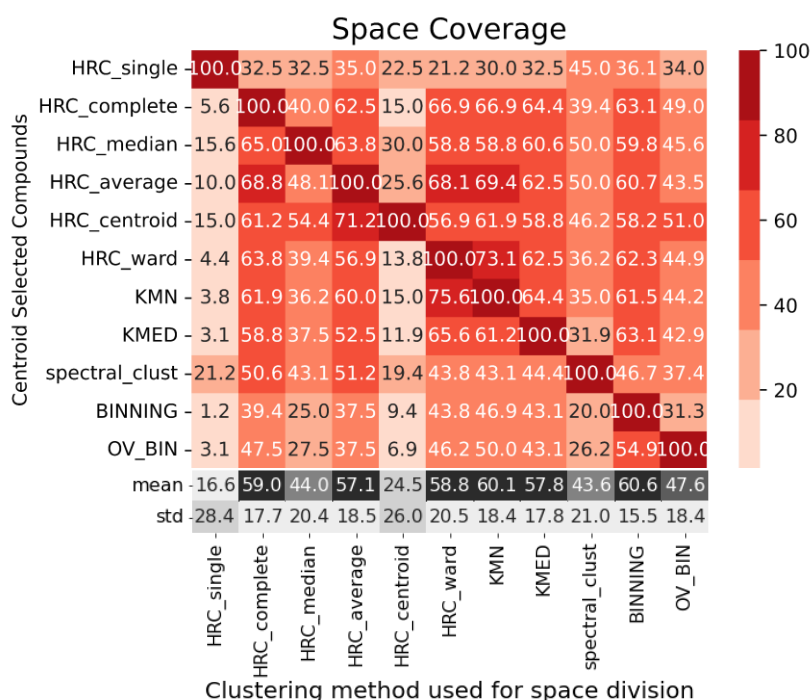


Figure 3.7 Space coverage heatmap for the clustering and partitioning methodologies assessing Tafenoquine's chemical space in 160 groups.

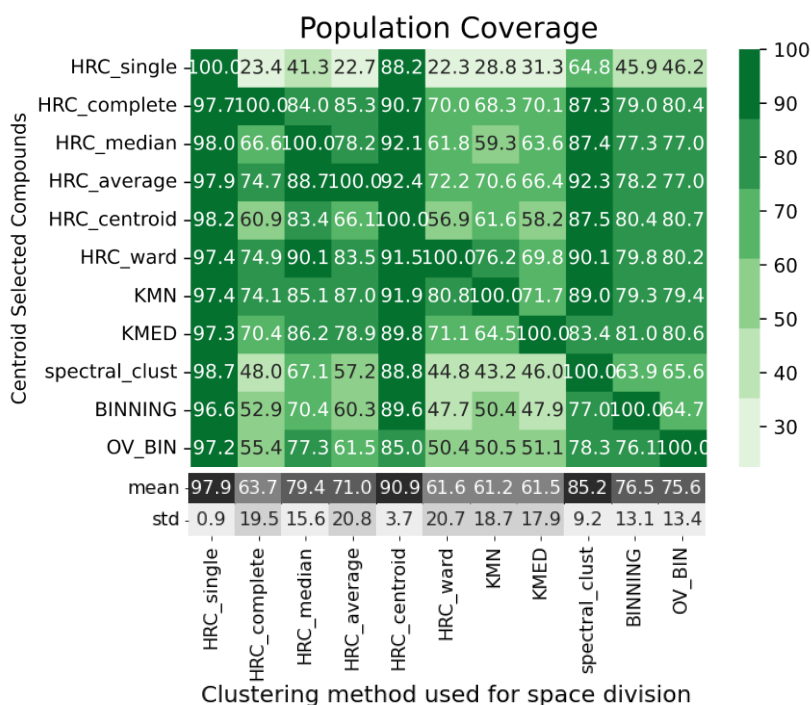


Figure 3.8. Population coverage heatmap for the clustering and partitioning methodologies assessing Tafenoquine's chemical space in 160 groups.

At a first glance, it stands out the poor SC values and large PC values obtained when studying the representativeness of the selections in HRC single and HRC centroid. This can be explained by the presence of an overpopulated cluster in both cases; the cluster 70 in HRC single contains 23,004 analogs and the cluster 85 in HRC centroid contains 18,162 compounds. These clustering methodologies commonly show many singletons – clusters with a unique compound – leading to a population distribution highly unbalanced. By this terms, HRC single and HRC centroid were firstly discarded to rationally represent the chemical space of the 8AQ analogs.

The deeper study of the population distribution in the different clustering methodologies performed and exposed via boxplot (shown in Figure 3.9 and Figure 3.10) finally gave us enough insights to decide which cluster could properly classify the database of study.

Firstly, Figure 3.9 confirmed the dissuasion to use of HRC single and HRC centroid clustering due to the overpopulated cluster above-mentioned. Furthermore, HRC median and spectral clustering were also rejected due to the same reasons although not as many singletons were observed as in the first two cases.

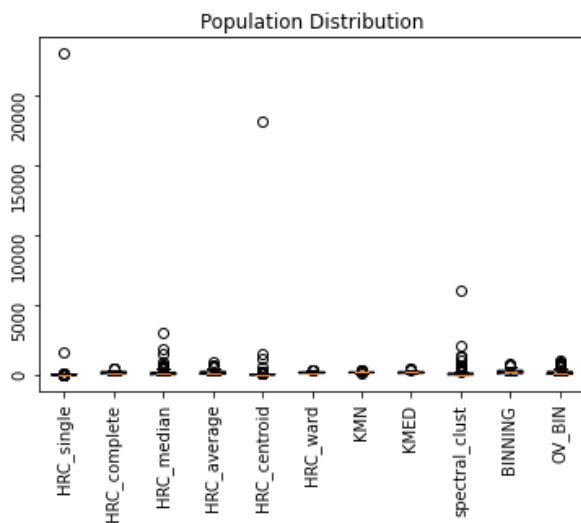


Figure 3.9 Population distribution of Tafenoquine database of analogs for k=160 (122 and 147 occupied bins for BINNING and OV BIN respectively) considering outliers

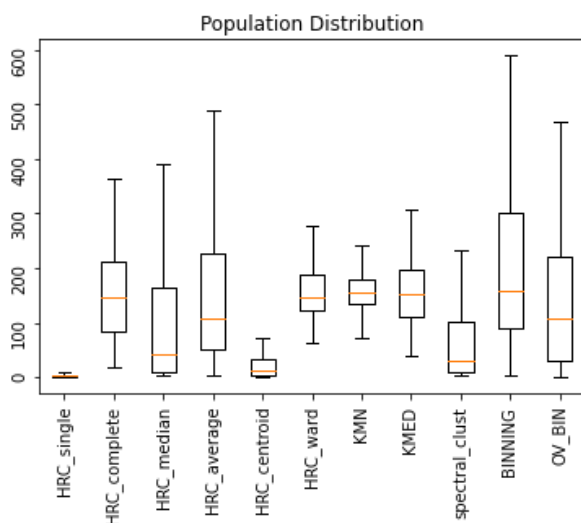


Figure 3.10 Population distribution of Tafenoquine database of analogs for k=160 (122 and 147 occupied bins for BINNING and OV BIN respectively) discarding outliers

Secondly, as the aim of the study is to find the most balanced distribution among clusters or cells it was reinforced the rejection of spectral clustering and HRC median at the light of the results shown in Figure 3.10.

After this first two criteria, whichever resting clustering methodology or partitioning could have been used. Hierarchical agglomerative methods were chosen as the favorite approaches for this

first study as it was meant to ensure the reproducibility of the study and these methodologies have a consistent distribution ruled by its fixed dendrogram – which is based mainly on the distance matrix and the linkage method performed – independently on the number of clusters considered. As the choice of linkage impacts the cluster formation, it is challenging to determine which clustering linkage works best. This can be overcome by the comparison of the cophenetic correlation coefficient detachedly on the number of clusters studied. This value is a measure to calculate of how faithfully the dendrogram preserves the pairwise distances in contrast with the original unmodeled data points. As a conventional correlation coefficient, the closest to 1 the highest quality solution is presented. Thus, the different HRC linkage behaviors were assessed considering its cophenetic value.

Table 3.2 Cophenetic coefficients measured for the six different HRC linkage methods applied to Tafenoquine’s combinatorial library of analogs consisting in 25,472 compounds.

Metric	Linkage	Cophenetic value
Euclidean	Single	0.408
Euclidean	Complete	0.264
Euclidean	Median	0.251
Euclidean	Average	0.520
Euclidean	Centroid	0.429
Euclidean	Ward	0.406

Finally, at the light of the results shown in Table 3.2, average linkage stood out as the linkage methodology of choice as it showed the largest cophenetic value.

3.2.3. The explored space known until the date

After choosing HRC average as the preferred clustering methodology to assess Tafenoquine’s chemical space, research of the compounds described in bibliography was conducted to discuss the explored space until date. This was accomplished programmatically checking the presence or absence of all the molecules of the chemical space of study in PubChem¹⁵⁶ database which contains publicly available compounds.

Accordingly, 58 molecules were found in bibliography which would represent the known chemical space at present and will be thereafter named as Bibliographical Data (BD). This low number of studied analogs may likely be related to the orphan condition of the studied drug and the disease itself. These not only include the individual molecules explicitly declares on the patent’s claims but

also derivatives which may be described in other patents or literature for the original application (as antimalarial drug) or for a different one. The resumed disease categorization for each analogue, its Pubchem CID and its original source as found in the public database is summarized on Table 3.3. It is worth noting that some of the molecules were found in the database but no record to source was found, these molecules have also been considered as part of the bibliographical data as they are chemically described.

Table 3.3. Information review of the data found in PubChem for the Tafenoquine database of analogs.

Library ID	CID PubChem	Disease	Source
24	115358	Malaria & Leishmania	Articles & Patents
904	133354	Inhibition of recombinant MAO	Articles & Patents
2496	182707	Malaria	Articles
2424	328150	Malaria & Hsp90 Inhibition	Articles & Patents
820	456317	Malaria & Pneumocystis carinii	Articles & Patents
838	473589	Inhibition of recombinant MAO	Articles & Patents
1616	11444465	Malaria	Patents
2388	11840454	Malaria	Articles
796	12827752	Malaria	Article
2412	12827754	Malaria	Articles
2568	12827758	Malaria	Articles
2425	12827760	Malaria	Articles
2430	12827762	Malaria	Articles
2766	12827764	Malaria	Articles
1174	12827766	Malaria	Articles
2520	12827768	Malaria	Articles
928	12827772	Malaria	Articles
833	13004209	Inhibition of recombinant MAO	Articles & Patents
976	13004212	Malaria	Patents
4004	13268300	Malaria	Articles
13556	13268302	Malaria	Patents
16740	13268304	Malaria	Patents
7188	13268306	Malaria	Patents
19924	20471245	Unspecified	-
13640	20471248	Unspecified	-
13712	20471249	Unspecified	-

Library ID	CID PubChem	Disease	Source
13574	20471252	Parasitic and opportunistic infections	Patents
13569	20471256	Parasitic infections	Patents
13773	22615922	Malaria	Patents
14003	22615925	Parasitic infections	Patents
1130	22615927	Malaria	Patents
20141	22615930	Parasitic infections	Patents
16957	22615932	Malaria	Patents
15232	53432170	Unspecified	-
1267	53897188	Malaria	Patents
7405	54372920	Malaria	Patents
1037	54526203	Inhibition of recombinant MAO	Articles & Patents
799	57335751	Inhibition of recombinant MAO	Articles
798	57392561	Inhibition of recombinant MAO	Articles
808	57392562	Inhibition of recombinant MAO	Articles
802	57396037	Inhibition of recombinant MAO	Articles
922	57399500	Inhibition of recombinant MAO	Articles
907	57399501	Inhibition of recombinant MAO	Articles
940	57399502	Inhibition of recombinant MAO	Articles
832	57399503	Inhibition of recombinant MAO	Articles
814	57401260	Inhibition of recombinant MAO	Articles
797	57401261	Inhibition of recombinant MAO	Articles
4	68485773	Unspecified	-
803	68824793	Malaria	Patents
12760	69401460	Malaria	Articles & Patents
15148	70697515	Malaria	Articles
2460	86020681	Malaria	Articles
8756	86133435	Malaria	Articles
8936	86133437	Malaria	Articles
8792	86133444	Malaria	Articles
8828	86133450	Malaria	Articles
13866	87983613	Parasitic infections	Patents
2394	145924641	Malaria	Articles

Shortly, when considering the disease of study for the set of 58 Tafenoquine analogs found in PubChem four main groups can be observed; there are 34 compounds that have been studied to treat malaria diseases¹²²⁻¹²⁷ (including in the subset of seven molecules explicitly described in the drug's patent claims^{105,109}), 5 have been tested in studies concerning parasitic infections¹²⁸, 14 have been tested for the inhibition of Monoamine Oxidase A (human) recombinant (MOA)¹²⁹ and 5 compounds with unspecified target of study.

When representing these four categories of bibliographical data found in the drug's chemical space in the first three principal components (Figure 3.11 which represents the 68.3% of the explained variance) it can be easily perceive that the data known until date is highly localized in a focused area of the chemical space. However, the data overlapping hinders the region inspection for each category.

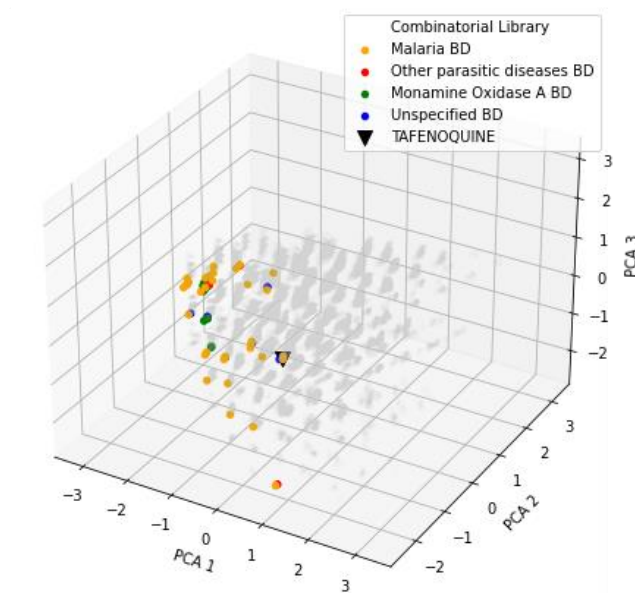


Figure 3.11 First three principal component plot displaying BD in Tafenoquine's chemical space

As an alternative, we represented the data in a t-SNE plot as showed in Figure 3.12. This representation developed by Van der Maaten *et al.*⁵⁶ (see Chapter 5) gives us a glance of the distribution of the samples in a 2D graph by merging the nearest analogs, hence, fewer spots than the stated may be represented as the nearest analogs (in terms of Euclidian metric) tend to be merged in this sort of representation.

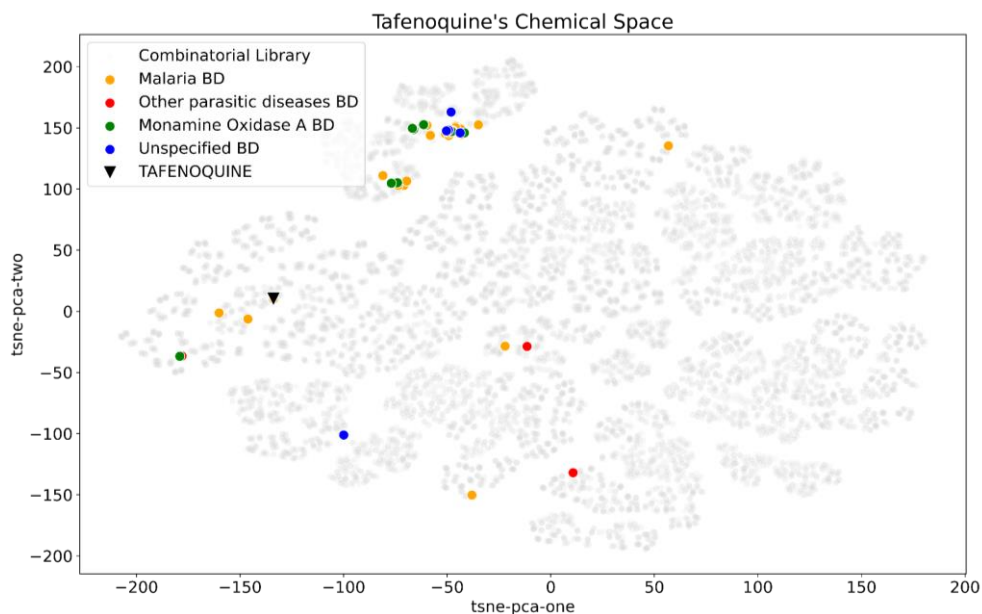


Figure 3.12 Bibliographical data displayed in Tafenoquine's chemical space categorized by disease

Independently on the representation, PCA and t-SNE figures confirm by an easy visual inspection that there is still a huge fraction of the space unexplored. Indeed, when representing only the derivatives described in the drug's patent claims (Figure 3.14) a clear data overlap is seen meaning that are three analogs out of the unique seven studied derivatives published by the developers (Figure 3.13) that are highly correlated in terms of chemical similarity as they are all isomers.

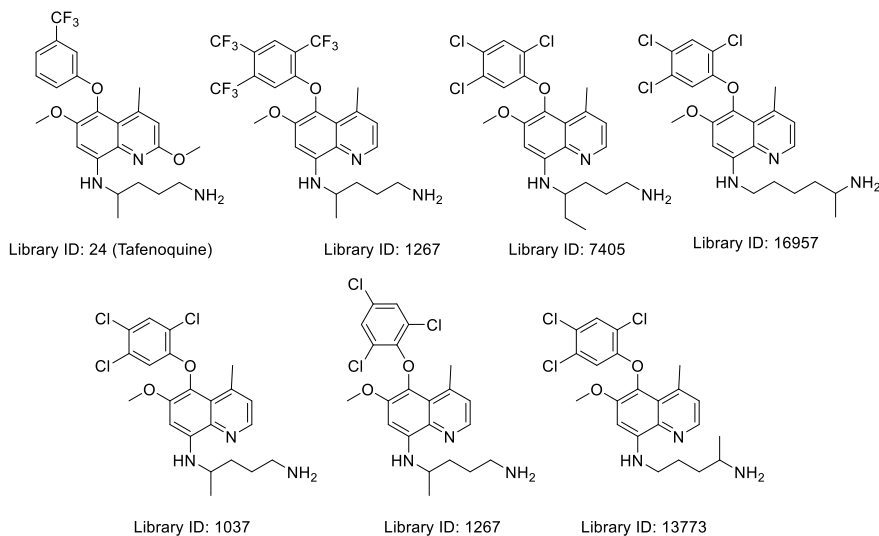


Figure 3.13 Molecules described in the patents.

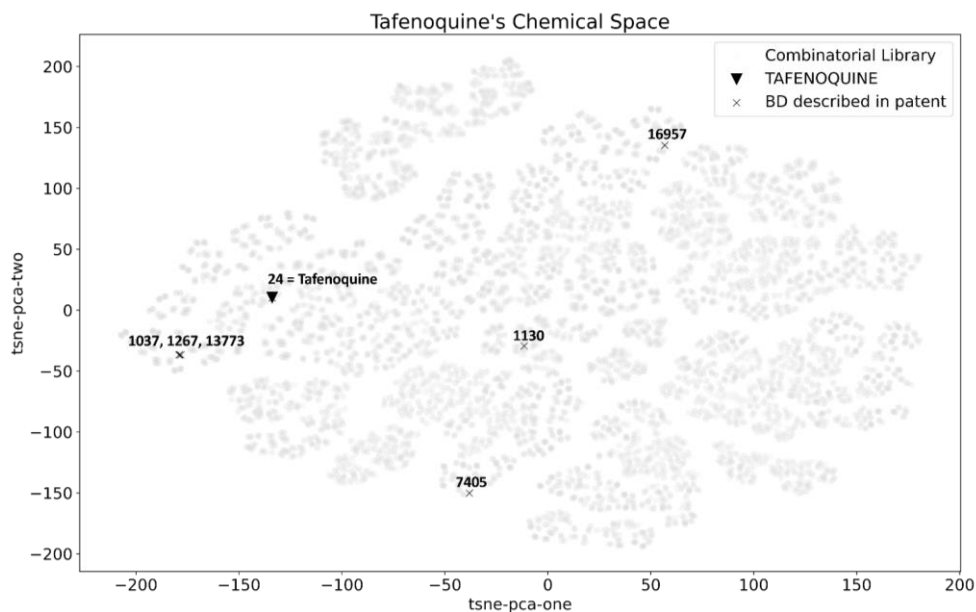


Figure 3.14 Display of the seven molecules described in the drug's patent in Tafennoquine's chemical space. The numbers shown in the plot correspond to the library index of the analogs represented.

In fact, as stated in Chapter 2 for the standard size of $\sqrt{N} \approx 160$ clusters calculated via HRC average, the overall subset of 58 analogs (N_{BD}) would represent the 8.8% SC and the 6.4% PC. In contrast, a random choice of N_{BD} analogs (calculated as the mean of 5,000 iterations) would better represent the chemical space with 27.0% SC and 45.4% PC. These short values might seem quite obvious when discussing the representativeness of only 58 compounds in a larger number of clusters. Nevertheless, the same trend occurs when clustering the space in N_{BD} (58) clusters. The BD data only covers the 17.2% SC and 12.8% PC while a cherry-picking approach selection of N_{BD} samples could achieve to cover a mean value of 46.8% SC and 78.4% PC. Once again, it should be highlighted that an optimal space exploration such as the derived via rational selection would result in a 100% of both SC and PC.

3.3. Rational Selection analysis

Through this study we aim to reinforce the use of an alternative R&D exploratory methodology based on rational selection in order to better represent the chemical database of Tafennoquine's analogs and possibly find regions with hidden potential biologically active derivatives. Moreover, the significant representation of the database can also describe the chemical space for further reprofiling purposes with the simple rational choice of a few compounds.

As a goal of this thesis, a proper assessment through rational analysis has been implemented in order to better explore the Tafenoquine's chemical space and possibly discover new regions with potential biological activity. The rational analysis departs from picking a tractable number of molecules to be synthesized in the laboratory. The choice of this size depends on the research team's human, economical and chemical resources along with the time expectations for the project taking in account a previous designed synthetic route. In our case, considering that only seven analogs have been described by the drug developers and our research plan in terms of time and resources, we considered synthesizing ten compounds to better describe this chemical space.

As the rational selection methodology selects a compound per each cluster intending to represent the other analogs within the same group –as they are expected to be highly similar –, the chemical database was divided in 10 clusters using HRC algorithm with average linkage. The clustering performance of the database is shown in Figure 3.15.

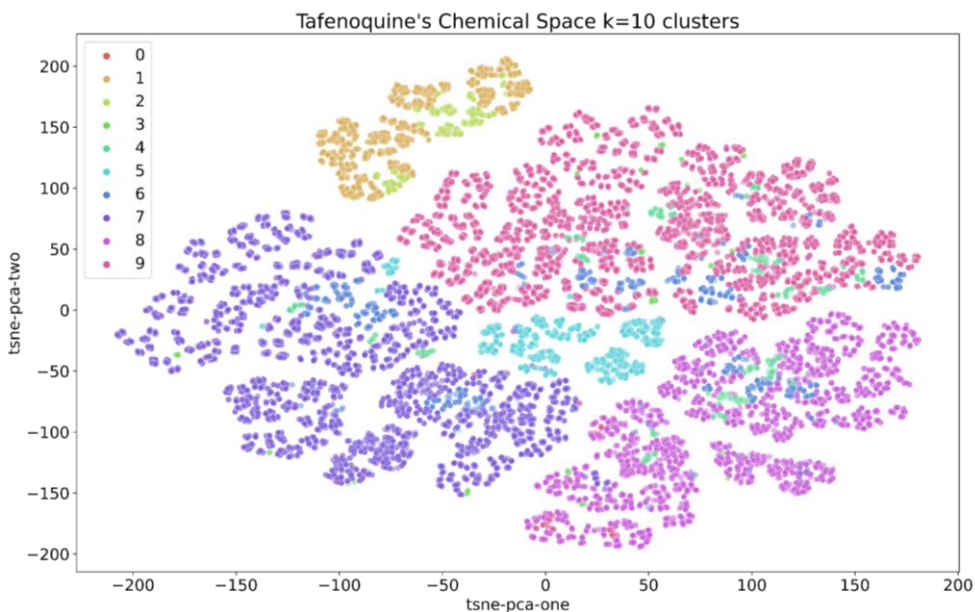


Figure 3.15 t-SNE representation of Tafenoquine's chemical space after HRC average clustering ($k=10$)

However, the computational manageability of these large datasets derived from chemical combinatorial libraries does not imply that all the analogs may be synthetically feasible or drug-like molecules. Hence, a further curation of the dataset is needed.

3.3.1. Database curation; shaping the accessible chemical space.

Three factors have been considered in order to shape the real accessible space considering our time and resources limitations and the drug-likeness of the derivatives: its synthetic feasibility, the commercial availability of the reagents and Lipinski's rule of five.

a. Synthetic feasibility of the chemical database

The synthetic feasibility of the compounds in any chemical library is one of the key issues and may limit their applicability when scaling to industrial development. Thus, the control of the synthetic complexity of the generated compounds seems to be a good starting point for data curation.

Tafenoquine synthesis was first stated in US4617394¹⁰⁹. The general procedure for its obtention was also related to the previous description of the synthesis of 4-methyl-5-(unsubstituted or substituted phenoxy)6-methoxy-8-(aminoalkylamino)quinolines, a previous patent (US4431807¹²³) developed also by LaMontagne *et al.* for the obtention of similar analogs unsubstituted in the second position. The general process of synthesis for the obtention of Tafenoquine is described in Figure 3.16.

Although the final synthetic route used in this study will be further deeply explained in section 3.4, the quick observation of the obtention indicated in the original drug's patent gave as a hint of the synthetic difficulty of some of the analogs. The overall synthesis consists in 15 steps. Briefly, the firsts four steps consist in the scaffold formation (**8**) and are followed by three steps for C3 derivatization (**39**). Afterwards, the nitro group in C8 is reduced to amino (**40**) to enable its derivatization. The following six steps (in the figure, colored in red), are mainly driven to add the methoxy group in C2 position via: protection of the amine in C8 (**41**), oxidation to form the N-oxide intermediate (**42**), the conversion to 2-quinolone (**43**), chlorination of C2 (**44**), the removal of the 8-phtalimido protecting group (**45**) and, finally, the nucleophilic substitution to introduce de alkoxy group in C2 (**46**). Finally, the two last steps are dedicated to the amino derivatization of C8 to introduce the aminoalkyl chain and yield compound **48**. Consequently, almost one-half of the synthetic route is entirely dedicated to derivatize the C2 position.

Therefore, in order to simplify the synthetic route from a 15-step to a final 9-step synthesis, the C2 derivatization position (R₂ in the Markush structure showed above in Figure 3.5, page 60) was fixed to those analogs containing a hydrogen atom. We hypothesized that this structural limitation would not restrain the biological activity of the molecule as many of tafenoquine's predecessors such as primaquine or other analogs which do not present any substituent in C2 has also shown schizontocidal activity in literature^{126,157-161}.

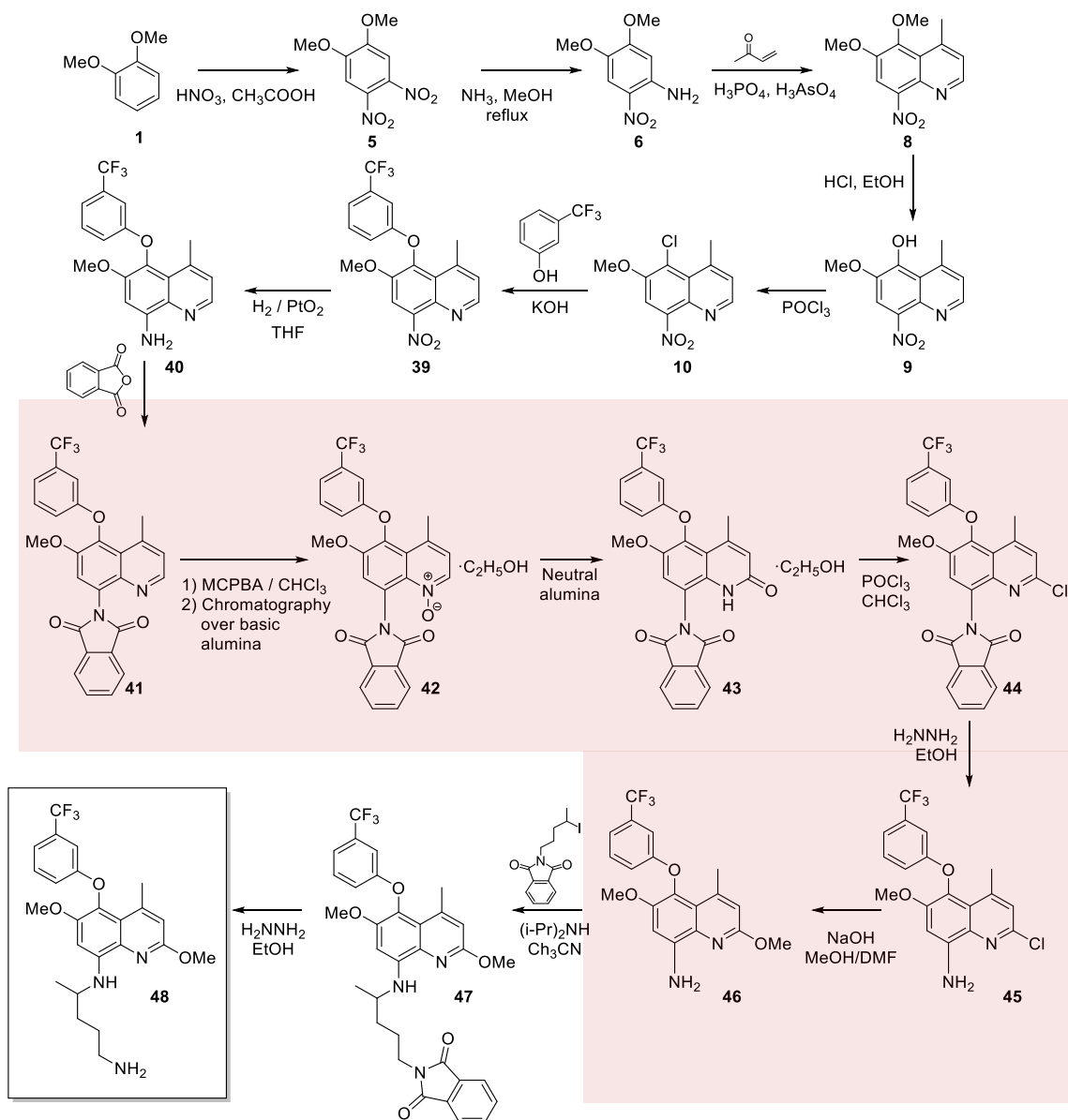


Figure 3.16. General process of Tafenoquine synthesis described in US4431807 and US4617394. In red is highlighted the needed steps for C2 derivatization.

To sum up, this first filter concerning to the synthetic feasibility of the library reduced the library of 25,472 analogs to half of the original dataset (12,736) representing still the 10 clusters and, in consequence the total coverage of Tafenoquine's chemical space.

b. Commercial availability of the reagents

The second limitation to take in account when designing a practical synthetic route is to check the commercial availability of the reagents needed. In our case, a part of veratrole – the library starting point – and different oxidant / reductive agents, solvents or catalysts to perform the different reactions, the accessibility to the reagents to form the different derivatization points was compulsory.

Taking a look to the above-mentioned Markush structure, the most diverse position that would potentially affect to the synthesis was the phenolic group in C5. All the possible combinations of R4, R5 and R6 among the five available spots of the aryl group lead to the requirement of 796 phenol analogs. The commercial availability of the usual vendors of the group (Sigma-Aldrich, Fluorochem, Abcr, Alfa Aesar and TCI) was checked programmatically using ChemSpiPy¹⁶² wrapper to access and query ChemSpider¹⁶³ database using Python. Surprisingly, amongst the above mentioned 796 possible phenol analogs, only 191 derivatives were found.

Therefore, the previous filtered dataset of 12,736 8AQ analogs was lowered to 3,056 compounds considering this commercial limitation. These were contained in 9 out of the 10 formed clusters but represent the 90.0% SC and 99.9% PC (leaving unrepresented the cluster 0 which only contains 4 molecules).

However, it should be emphasized once again that this filter was applied to simplify the synthesis process due to time limitations, some of the possible non-commercially available phenols could also be obtained by adding some side synthetic steps.

c. Lipinski rule of five

The last filter applied to the database of study was the consideration of Lipinski's rule of five which, as introduced in the Chapter I, defines four rules to measure the *druglikeness* or *druggability* of a molecule by a limited range of molecular properties¹⁶⁴. These rules predicts that poor absorption or permeation is more likely when:

1. There are more than 5 hydrogen bond donors
2. There are more than 10 hydrogen bond acceptors
3. The molecular weight is greater than 500
4. The LogP is greater than 5

In order to avoid a highly restrictive threshold, we defined this last filter to select those molecules which violated none or at least one of the rules leading to a final curated set of 2,548 tafenoquine analogs.

3.3.2. Rational selection of Tafenoquine's analogs

After data curation, the resulting chemical database declined from an original chemical space consisting in 25,472 to an accessible space described by 2,548. However, although the real accessible database was only represented by a 10% of its original data (2,548 out of 25,472) the unique chemical space appears to be better described than the known until date in literature. Remarkably, this short, filtered sample contains representative molecules of 8 out of the 10 formed clusters (leaving cluster 0 and 5 unrepresented). Therefore, the achievable exploration through rational selection in our circumstances could optimally represent the 80% SC and 95% PC. Therefore, we performed the rational selection of 8 representative compounds being these the nearest to each cluster centroid accessible compound. These selection appertain the molecules shown in Figure 3.17.

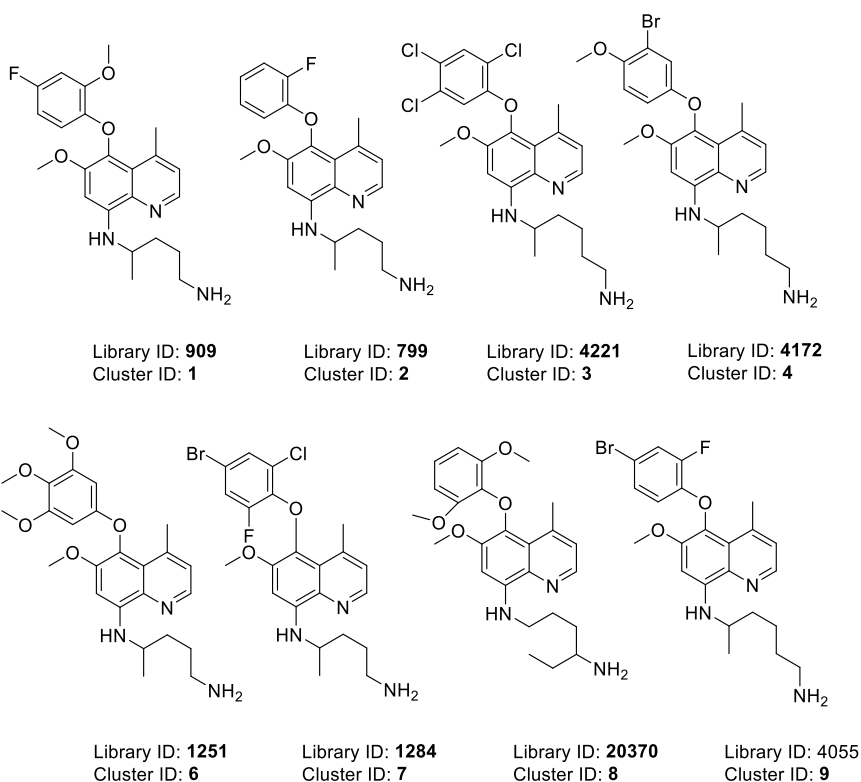


Figure 3.17. Rationally selected candidates to represent the 80% SC and 95% PC of the chemical space of Tafenoquine analogs

All things considered, as shown in Table 3.4, in this case of $k=10$ clusters, the coverage results of a random selection of 10 candidates would only improve the PC results in contrast with the BD coverages. However, it is worth noting that the rational selection of the 8 candidates has increased both SC and PC in contrast with the 58 molecules known until date.

Table 3.4 Comparison of different coverage results of the chemical space obtained for BD selection, random selection and rational selection in k=10 clusters

Selection	Subset size	SC	PC
Bibliographical Data	58	50.0 %	42.5%
Random	10	37.7%	73.2%
Rational Selection of accessible space	8	80.0 %	95.0%

In conclusion, the rational selection of a sample of eight molecules, one per each synthetically accessible cluster, would largely increase the actual knowledge of Tafenoquine's chemical space by a 30.0% SC and 47.5% PC. Indeed, there is an abysmal difference in its representativeness if comparing the seven described molecules in the drug's patents which poorly represent the 30.0% SC and 34.1% PC.

3.4. Synthesis of 8-aminoquinolines

3.4.1. General synthetic Route

As stated above, 8-aminoquinolines are very versatile compounds that have been widely used and explored as potential antimalarial candidates. The overall synthetic route used for the synthesis of the rationally selected analogs comprehend a route of 9-step synthesis that, schematically follows the derivatization pathway showed in Figure 3.18. As explained through the curation process, this route has been designed considering analogs with $R_2=H$. Thus, no derivatization point in C2 is reflected in the procedure for these 8-aminoquinolines obtention.

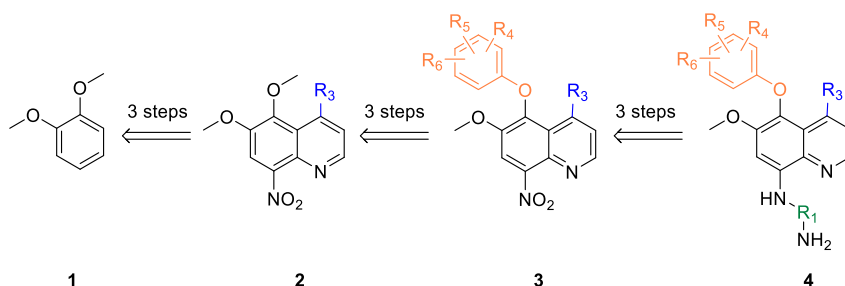


Figure 3.18 Schematic synthetic route considering the steps to obtain the intermediates with the derivatization points of the scaffold of study

The simplified route shows how every different substituent can be obtained via three-steps synthesis from the previous derivatization point. Therefore, the shared synthetic steps of the analogs

of study affected only to the three first steps and increased the complexity of the parallel synthesis. However, this limitation was overcome as all the rationally selected molecules presented a methyl group in R₃. Therefore, a large batch of product until the fifth step could be produced before the step involving the second derivatization point formation.

In the designed synthetic route, many reactions have been changed in contrast with the procedure suggested by the patent claim in order to afford a more efficient, safer and cleaner experimentation. However, the structural modifications follow the same order as suggested by the drug developers. Finally, the general synthetic route followed for the obtention of the selected molecules is shown in Figure 3.19. This will be disclosed below in the following sections.

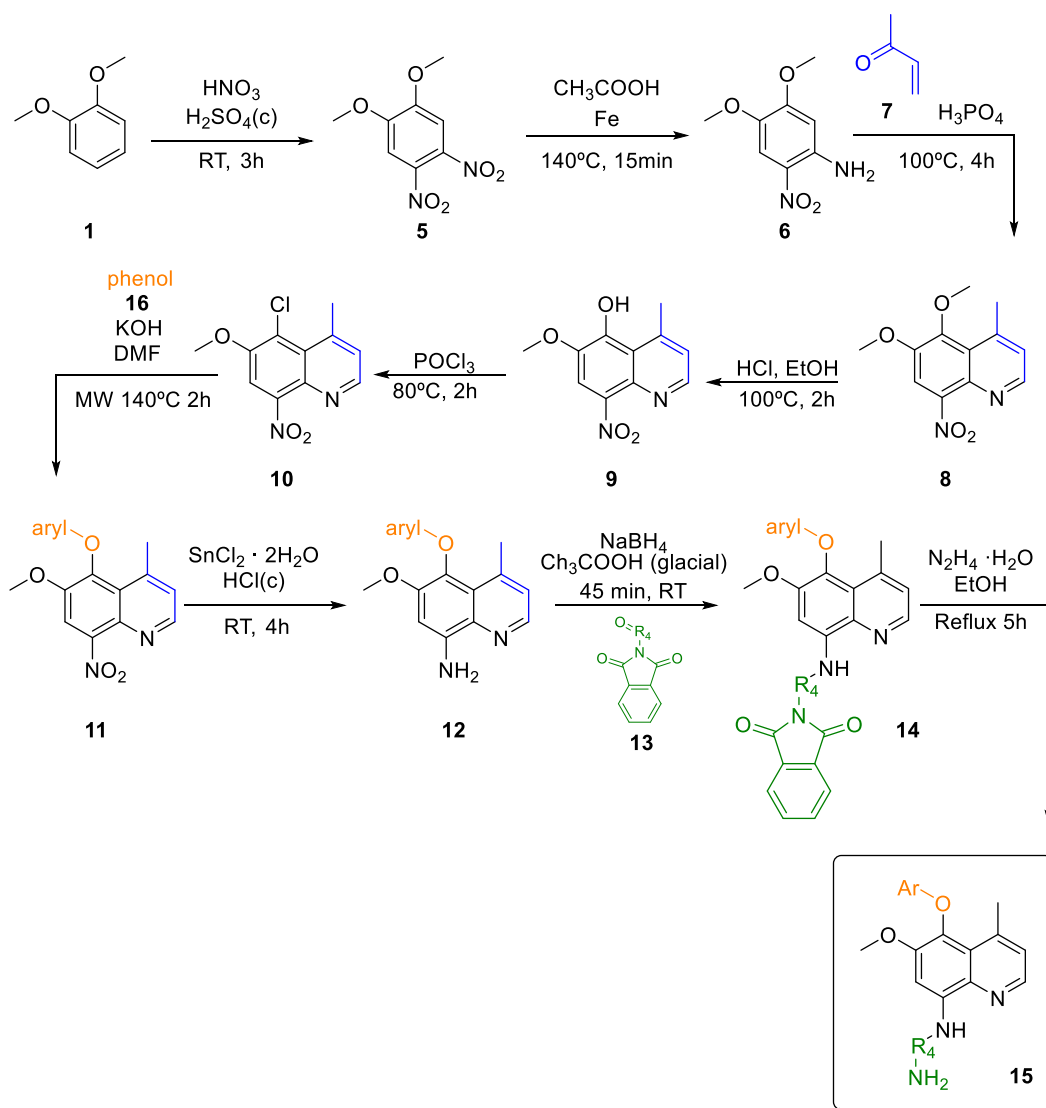


Figure 3.19 General synthetic route designed for 8-aminoquinoline obtention

3.4.2. Preparation of 5,6-dimethoxy-4-methyl-8-nitroquinoline (**8**)

The 5,6-dimethoxy-4-methyl-8-nitroquinoline (**8**) scaffold needed for the preparation of the different compounds of interest can be synthesized directly from commercial 4,5-dimethoxy-2-nitroaniline (**6**) and methyl vinyl ketone (**7**). However, as both **6** and even **5** (its precursor) intermediates trend to be expensive reagents (e.g. for both cases its cost for only 10mg is 132.0€ in Sigma-Aldrich) and it is necessary a huge quantity of departing material in order to perform so high number of steps of synthesis, we decided to start directly from veratrole (**1**) structure in order to

perform a cost-effective synthetic procedure. In contrast, the cost of **1** in Sigma-Aldrich database is 24,10€ per 100g of product.

Following an adapted procedure described by Hirth *et al.*¹⁶⁵ veratrole (**1**) reacts with 9.0 eq. of 65% nitric acid and 3.7 eq. of concentrated sulfuric acid to obtain the dinitrated analog **5** with a 98% yield. This intermediate **5** is later monoreduced adapting the procedure described by Wulfman *et al.*¹⁶⁶ where it is used powder iron as a reductant agent in acidic conditions (facilitated by glacial acetic acid) to obtain, through a highly vigorous and exothermic reaction the intermediate **6** with 87% yield (Figure 3.20).

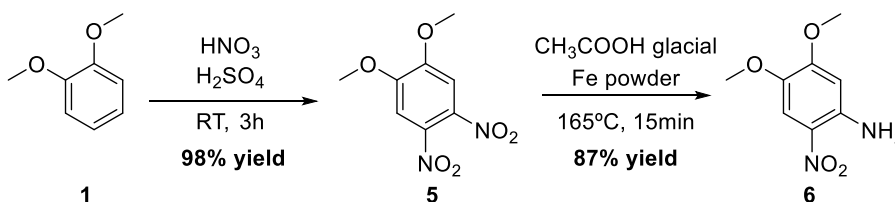


Figure 3.20. First two steps of the synthetic route to obtain 8AQ analogs

The reaction conditions and product isolation regarding the second step were modified and improved from its original procedure to raise the reaction yield. The final procedure **LMF241** was chosen as the most suitable as shown in Table 3.5.

Table 3.5. Experimental conditions tested for the obtention of intermediate 6.

Experiment No.	Conditions	Observations	Yield
LMF202	As stated in Wulfman <i>et al.</i> ¹⁶⁶ : Reaction was performed at 140 °C. Product isolation consist in two filtration: while being hot (separating iron traces) and after precipitating the filtrate in water.	Low yield, no vigorous reaction was seen. Some of them did not progress.	31 %
LMF217	Temperature reaction was raised to 165 °C to stimulate reaction activation.	Vigorous reaction occurred but isolation presented difficulties. The first filtration forced the precipitation of the product along with the iron traces due to RT conditions.	58 %
LMF241	Reaction temperature was settled to 165 °C. Work-up conditions were modified: the crude was directly precipitated in water, filtered, and the solid was transferred by dilution to a new flask with acetone separating it from the iron traces.	The work-up procedure facilitated the isolation of the product in terms of time and yield effectiveness.	87 %

Finally, the scaffold formation can be obtained through a Doebner-Miller reaction (also known as Skraup-Foebner-Bon Miller quinoline synthesis). This step affords intermediate **8** via conjugate addition of the compound **6**, a primary arylamine with an unsubstituted *ortho* position, with the α,β -unsaturated compound, methyl vinyl ketone (**7**), in presence of a protic acid which serves as a catalyzer for dehydrative cyclization such as orthophosphoric acid and using As_2O_5 as oxidant to favor the oxidative aromatization.

Although no reaction mechanism is totally confirmed until date for this kind of synthesis, an approach is shown in Figure 3.21. The aniline serves as nucleophile which attacks the β position of the 2-butenone forming the Michael adduct (**19**) which presents a keto-enol tautomerism. The intermediate then suffers a cyclization mechanism where the amine favors the electron transfer until closing the ring which through a series of deprotonations and elimination of two molecules of water will finally lead to the aromatization of the cycle by oxidation affording the quinolinic scaffold.

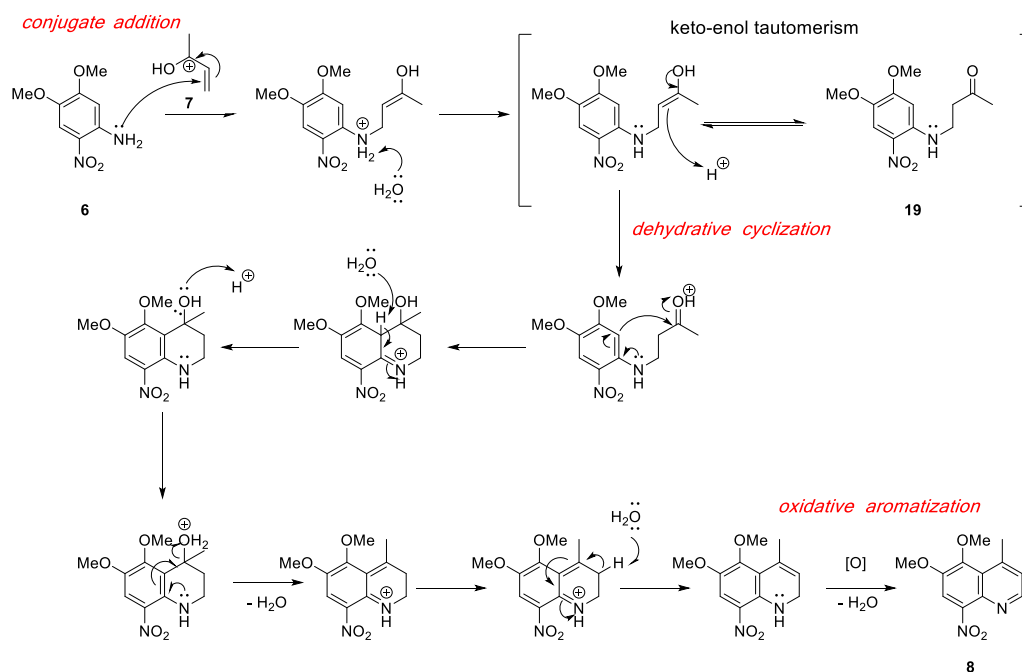


Figure 3.21 Doebner-Miller reaction mechanism to obtain intermediate **8**.

However, as expected for this sort of reactions involving α,β -unsaturated compound conjugations, the described yields in literature (30%¹²⁶ or 38%¹⁶⁷) are not as promising as desired. The high reactivity of α,β -unsaturated analogs in high temperature reactions with such acidic conditions normally involves the formation of undesired side products and cleavages, forcing the final product isolation to be driven via column chromatography. Thus, some attempts changing the reaction

conditions have been tested and are described in Table 3.6. These modifications departed from an adaptation of the methodology suggested by Jain *et al.*¹⁶⁷ for quinoline obtention. Consequently, although an improvement has been achieved, this reaction has consisted in the bottleneck step for the overall obtention of 8AQ.

Table 3.6. Experimental conditions tested for the obtention of intermediate 8.

Experiment No.	Conditions	Observations	Yield
LMF203	Adaptation from Jain <i>et al.</i> ¹⁶⁷ . It has been used 2.0 eq. of As2O5 instead of arsenic acid and 4.5 eq. of methyl vinyl ketone (7), which was added dropwise in three parts. Reaction took place at 100-110 °C for 2 hours. The subsequent work-up consisted in: crude precipitation in an ice-water batch, a charcoal treatment followed by filtration, alkalization with 30% ammonia, extraction with dichloromethane (DCM) and solvent elimination and a final residue purification by chromatographic column over silica gel (20:80 cyclohexane (Cy)-DCM, Rf 0.19) . Described yield (with another substrate): 38%.	Complex charcoal filtration and further extraction as the crude is highly dense which results in a very time-consuming work-up. An improvement of product isolation is needed.	-%
LMF203b	Modification from procedure of Lamontagne <i>et al.</i> ¹²⁶ . The previous reaction conditions are followed but residue was purified before chromatographic column with an extra step involving its reflux in toluene and further solid traces separation. Chromatographic column conditions: 50:50 toluene -DCM, Rf 0.35. Described yield: 30%	Time-consuming work-up process although problems with filtration have been overcome.	26%
LMF203d	Same procedure as LMF203b but with new chromatographic column conditions (20:80 Cy -DCM, Rf 0.19)	Improved yield due to comatographic conditions	33%
LMF203f	Same procedure as LMF203b but with new chromatographic column conditions (40:60 Cy -DCM, Rf 0.04)	Lowered yield and highly time-consuming separation column	30%
LMF245	Reaction was performed with no oxidant agent (As2O5) and the mixture was stirred for 4h at 100 °C. Work-up was simplified by direct crude precipitation in ice-water bath, alkalization with 30% ammonia, extraction with AcOEt, washings and final product was isolated column chromatography (silica gel with 30:70 Cy:DCM)	The process was simplified and the yield was the highest obtained through all the experimentation assays.	37%

Experiment **LMF245**, where no oxidant was used, was tracked during 28h at 100 °C to study if the evolution of the reaction was time dependent. Previous crude purification had permitted the isolation of a stable form (**19**) of the Michael adduct intermediate. Its characteristic triplet signal assigned to NH group (in blue) at 8.52ppm (as described in Chapter V) in contrast with the doublet formation assigned to C1 of the desired product **8** at 8.65ppm (in green) gave us insights of the reaction evolution over time as shown in Figure 3.22. At the light of the results, we concluded that

the highest conversion rate could be achieved through a 4h-reaction at 100 °C if discarding the use of any oxidant agent and using only phosphoric acid along with the desired α,β -unsaturated compound.

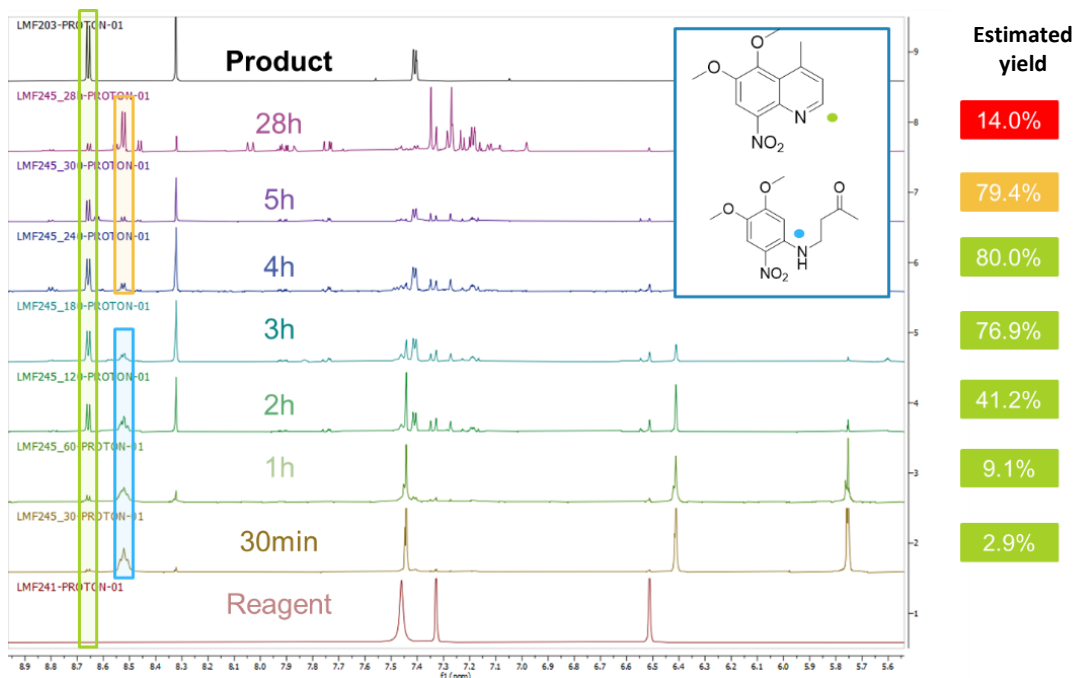


Figure 3.22 Doebner-Miller reaction crude evolution until 28h of reaction at 100 °C departing from intermediate 6 (reagent). The formation of 16 (Michael adduct) is observed in blue, the desired product 8 evolution in green and unknown signals from another side product in orange.

3.4.3. Preparation of 5-phenoxy-6-methoxy-4-methyl-8-nitroquinoline derivatives (11)

The different 5-phenoxy-6-methoxy-4-methyl-8-nitroquinoline (11) intermediates were obtained through three steps of synthesis. In order to facilitate the final nucleophilic aromatic substitution, the habilitation of a good leaving group such as a chloride on C5 was necessary. All things considered, there are other approaches to enable this substitution such as the formation of a sulfonate like tosylate in C5 or another halide like bromide or iodide.

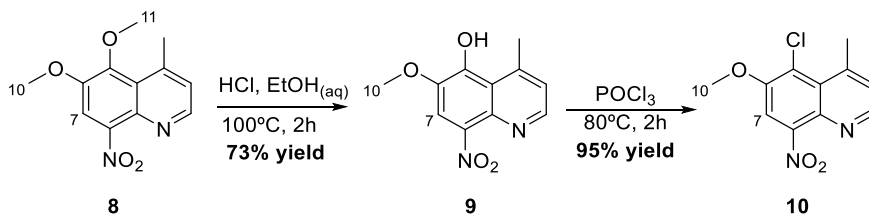


Figure 3.23 Fourth and fifth step of the general synthetic route to obtain 8AQ analogs

In this case, the methoxy group was selectively removed through an acidic cleavage (S_N2) driven by a reflux of compound **8** with hydrochloric acid in an aqueous ethanolic solution to obtain the corresponding alcohol derivative (**9**) with excellent yields following the procedure described by Carroll *et al.*¹⁶⁸. Making use of the same protocol, the intermediate **9** was further chlorinated using phosphorus (V) oxychloride as a powerful chlorodehydroxylation agent to afford compound **10** with quantitative yields. The product isolation of this last step must be carefully driven as the $POCl_3$ quenching present high risk of toxic fumes formation mainly containing hydrogen chloride among.

Above all, these two synthetic steps can be easily tracked through the 1H NMR comparison of significant signals of the three intermediates as shown in Figure 3.24. As an example, the alcohol formation can be clearly evidenced by the clearance of a methoxy group (C11) and the large C7 and slight C10 signals upfield displacement when comparing **8** and **9** spectra. Finally, the intermediate **10** may corroborate the presence of the newly incorporated halide in C5 due to the deshielded C7 and C10 signals. This last intermediate was confirmed by mass spectrometry proving the expected presence and 3:1 abundance of the $[M]^+$ and $[M+2]^+$ peaks (see Chapter 5).

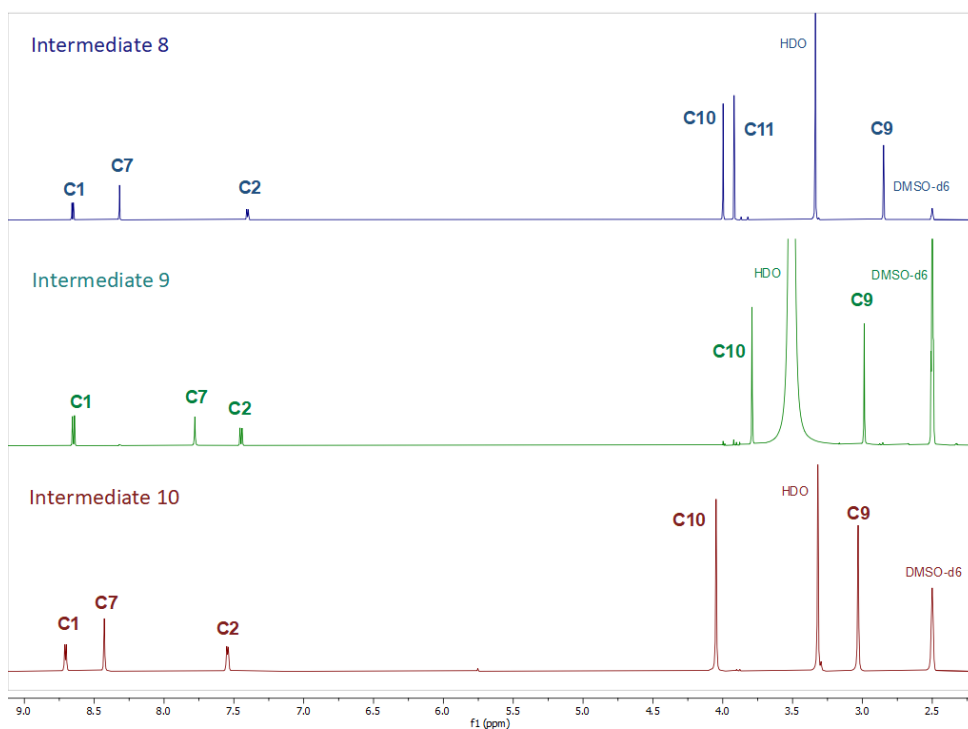


Figure 3.24 1H NMR comparison of intermediates 8, 9 and 10.

The first intermediates with different derivatization residues in C5 were finally obtained via nucleophilic aromatic substitution (S_NAr) of the chloride using different affordable commercial

phenols. In this step, a Williamson ether synthesis takes place thanks to the presence of a hydroxide ion containing base to allow the phenoxide formation which acts as a nucleophile to displace the halide.

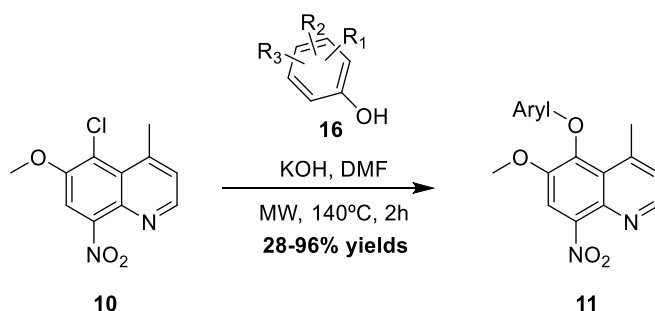
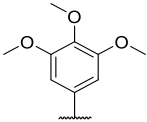
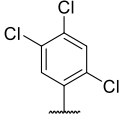

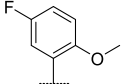
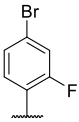
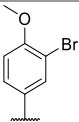
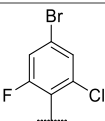
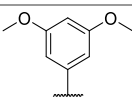
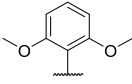


Figure 3.25 Sixth reaction step involving the nucleophilic substitution of the C5 chloride compound **10** using a phenol as nucleophile to yield **11** derivatives.

First attempts (**LMF206**) in 2-ethoxyethanol and formal heat until reflux were performed as described by Nodiff *et al.*¹²⁴ being inconclusive as the substrate **10** was reobtained. The base was then hopelessly changed to sodium hydride (strong base) deriving to the production of a mixed residue which forced its purification by column chromatography (**LMF222**). Finally, it was proved that compound **11** obtention depended on a previous isolated formation of the phenoxide with potassium hydroxide in DMF at room temperature to react with a later introduced substrate **10**. Finally, the protocol described by Chaurasiya *et al.*¹²⁹ involving conventional heating was followed and was successfully improved by the use of microwave irradiation to perform the reaction.

As shown in Table 3.7, the obtained yields were excellent ranging from 75 to 96% except for a unique case (the less nucleophilic phenoxide) out of the nine derivatives synthesized which showed a 28% yield. Nine compound **11** analogs were synthesized during this study. In contrast with the introduced rational selection, compound **11{8}** was synthesized aiming the obtention of one compound not included in the library which will be discussed below. Moreover, it was taken into consideration the cost-effectiveness of the phenols (purchased in Fluorochem¹⁶⁹) involved during the rational selection as evidenced next.

Table 3.7. 5-phenoxy-6-methoxy-4-methyl-8-nitroquinoline derivatives (**11**) prepared in this project with its corresponding yields and the commercial price of the involving phenols (**16**) for its synthesis.

Aryl structure	Name	Commercial substrate price	Yield
	11{1}	16{1}: 12£/1g	84%
	11{2}	16{2}: 36£/25g	96%
	11{3}	16{3}: 18£/50g	85%
	11{4}	16{4}: 10£/1g	91%
	11{5}	16{5}: 10£/1g	75%
	11{6}	16{6}: 26£/1g	82%
	11{7}	16{7}: 10£/1g	28%
	11{8}	16{8}: 10£/1g	89%
	11{9}	16{9}: 10£/1g	83%

We hypothesize that the low yields obtained for **11{7}** can be rationalized by the charge delocalization among the phenolate group provoked by the trihalide presence in *ortho*- and *para*-positions which leads to a highly stable structure lowering its nucleophilicity and impeding the reaction evolution. The repetitions of this experiment (**LMF302**) involved a later product purification by automated flash chromatography separation (silica column, AcOEt: Cy 1:1). Despite modifying MW

time and **11**{7} equivalents, in all cases both substrates (**10** and **11**{7}) were recuperated along with the product of interest **11**{7} so no improvement was achieved.

3.4.4. Preparation of *N*-(6-methoxy-4-methyl-5-(phenoxy)quinolin-8-yl)alkyldiamine derivatives (**15**)

Final obtention of the desired 8AQ of interest was achieved by three last steps involving the nitro group reduction, a reductive alkylation to add the corresponding alkylamino sidechain protected with a phthalimide group and a final protecting group removal by hydrazinolysis.

The seventh synthetic step consisting in the selective nitro reduction (Figure 3.26) was performed with an efficient procedure normally used for nitroarene reductions treating the intermediates **11** with 5.0 equivalents of tin (II) chloride dihydrate as reducing agent in highly acidic conditions favored by the presence of concentrated hydrochloric acid. The reaction is normally quenched with an alkaline base to remove the excess of the tin (II) chloride dihydrate which tends to firstly precipitate as the hydrated tin (II) oxide form which is finally dissolved as a stannite salt in excess of base. Nevertheless, other reductive procedures could have also been considered. To name a few, Azad *et al.*¹²⁵ obtained these analogs through Ni-Raney reduction, Chaurasiya *et al.*¹²⁹ performed a catalytic hydrogenation with 10% Pd on C (Pd/C) or the original procedure stated in the drug patent¹²³ suggested another catalytic hydrogenation using H₂/PtO₂. The procedure finally used in this study afforded excellent yields ranging from 74% to 100% with the exception of **12**{8} with a 57% and **12**{9} with a 13% yield. The last poor obtained yields are attributed to a deficient experimental performance that did not enhance the total dissolution of the corresponding **11** substrate – which clumps rupture had been normally assisted by sonication – leading to the obtention of a mixture of reactant and product which forced the purification through column chromatography (silica gel, Cy:AcOEt 1:1) and a corresponding product loss.

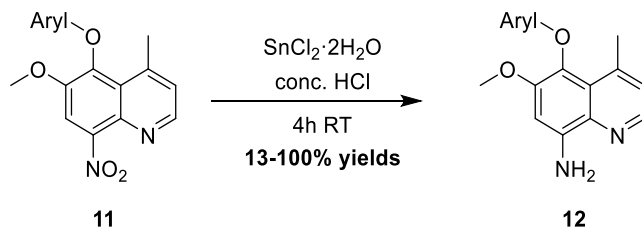


Figure 3.26. Seventh reaction step involving the C8 nitro reduction compound **11** using tin (II) chloride dihydrate in acidic conditions to yield **12** derivatives.

Finally, in order to proceed with the C8 amino derivatization, a previous preparation of the corresponding alkylamino groups was necessary. As above shown in the selected structures (see Figure 3.17, page 75), three different sidechains had to be attached (Figure 3.27).

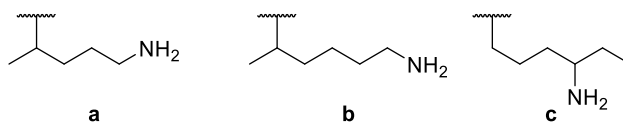


Figure 3.27. Sidechains needed to derivatize C8 position as shown in the rational selected molecules: 4 analogs present a sidechain structure, three analogs b and one analog c.

The common methodology used to attach these alkylamine chains as described in literature^{125,129,155} suggest the use of the amine-protected oxoalkane form of these analogs to react with the corresponding **12** substrates. It is crucial the protection of the free amine group in order to avoid undesired side reactions with the same reactant which could evolve forming polymeric structures leading to a substantial loss of substrate. Thus, the previous synthesis of these **13** reactants presenting an amino protecting group was required.

This was achieved by a Gabriel reaction which first step (Figure 3.28) was able to transform the oxoalkyl halide (**17**) substrates to their intermediate *N*-oxoalkylphthalimides (**13**) derivatives which, after the attachment to the main scaffold, were finally deprotected through a last synthetic step to yield the desired terminal free amine structures (**15**).

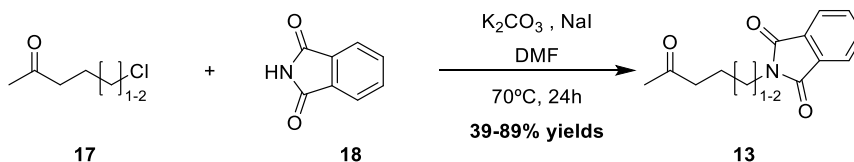


Figure 3.28. First step of Gabriel reaction to form the *N*-alkylphthalimides (**13**)

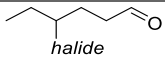
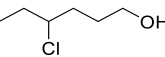
The first step for the obtention of **13** intermediates was adapted from the protocol described in WO 03/093239 patent¹⁷⁰. Intermediates **17** were treated with 1.5 equivalents of potassium carbonate and after studying the influence of the phthalimide rate by a few attempts, the amount of compound **18** was fixed to 1.1 equivalents. Finally, the comparison with other procedures gave us a sight that sodium iodide might increase the obtained yields as it acts as a catalyst in this nucleophilic substitution (S_N2). These first attempts to establish the improved conditions are resumed in Table 3.8 in a preliminary study developed to obtain 2-(4-oxopentyl)isoindoline-1,3-dione (**13{1}**).

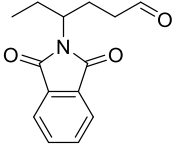
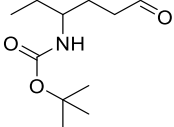
Table 3.8. Tested conditions to improve the product purity and yields for the first step of the Gabriel synthesis to produce **17** reactants.

Exp. No.	17{1} eq.	Phth. eq.	K ₂ CO ₃ eq.	Time	NaI eq.	Yield	Observations
LMF250	1.0	2.2	1.5	22h	-	21%	Conditions as described in literature ¹⁷⁰ . Crude purification needed column chromatography.
LMF250c	1.0	2.0	1.5	24h	-	51%	Crude purification needed column chromatography.
LMF254	2.0	1.0	1.5	24h	1.0	96%	Pure product obtained without purification. Better yields obtained but implies a large waste of the most expensive reagent 17{1} .
LMF255	1.0	1.1	1.5	24	1.0	40%	Best yield measured for direct obtention of the pure product without 17{1} waste.

Unfortunately, not all the three **17** necessary reagents could be purchased as only two of them presented an affordable cost. 5-Chloropentan-2-one (**17{1}**) and 6-chlorohexan-2-one (**17{2}**) could be purchased in Fluorochem for 14£/25g and 19£/1g respectively. However, 4-chlorohexanal, the necessary corresponding analog to obtain **c** sidechain above introduced in Figure 3.27 (page 87), was not commercially available. Other compounds were considered as listed in Table 3.9 even contemplating a previous one-step synthetic preparation but all direct approximations were inconclusively due to commercial unavailability or high reagent costs.

Table 3.9. Alternatives considered to overcome the commercial unavailability of 4-chlorohexanal.

Reagent	Commercial availability	Synthetic observations
 <i>halide</i>	None	Direct first step Gabriel synthesis
	700\$/1g	An additional previous oxidative step would be necessary

Reagent	Commercial availability	Synthetic observations
	None	Direct reaction with 12 for scaffold attachment
	700\$/1g	Direct reaction with 12 for scaffold attachment

At the light of the reasons explained above, the synthesis of the rationally selected molecule from cluster 8 with library ID 20370 could not be achieved in this study. Its replacement by another analog in the cluster was considered but no alternative molecule could be synthesized due to time and costs limitations.

Even though, the achievable synthesis of the seven out of the initial eight considered analogs did still increase the knowledge of the chemical space in terms of space and population coverage as discussed below.

These were finally obtained by two last synthetic steps (Figure 3.29) following the protocol described by McChesney *et al.*¹⁵⁵. Firstly, the attachment of the sidechain was performed through a reductive amination of the compounds **12** and **13** in moderate acidic conditions and using sodium borohydride as reductant agent.

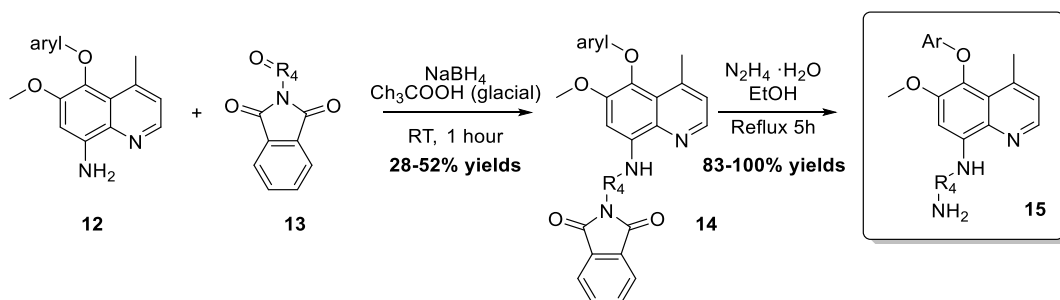


Figure 3.29. Last two steps to afford the rationally selected 8-aminoquinoline derivatives

The first reaction takes place in two parts. A first nucleophilic addition of the amine with the carbonyl group forms the imine intermediate which is afterwards reduced through a second step to the secondary amine thanks to the reductant action of the sodium borohydride. However, during this process, side-products may be formed coercing the product purification by column chromatography.

This isolation procedure of such derivatives leads to relatively poor yield values which ranges from 28% to 52% making a huge impact to the global reaction yield which will be furtherly discussed.

Finally, the deprotection of the terminal amine by the hydrazinolysis of the phthalimido group with hydrated hydrazine in ethanol afforded the rationally racemic selected compounds **15** with excellent yields (83-100%) as described in literature.

1.1.1. Global synthesis overview

To sum up, seven out of the originally eight selected Tafenoquine derivatives were finally synthesized following the 9-step synthesis methodology described above. Table 3.10 serves as a summary of the single yields obtained to synthesize the corresponding intermediate derivatives. The overall global yield measured ranges from 2% to 8% with a mean value of 5%. The limitation steps, as mentioned are the step 3 (Doebner-Miller reaction), which conditions and yield were improved, and step 8 (reductive amination).

In general, the obtained total yields are considered acceptable for a multistep process because no optimization study has been considered during the process as it was not the objective of the present work. In fact, as many other drugs in the market that are the product of several steps of chemical synthesis, the overall percent yield might be tiny, which is one factor in the huge cost of some drugs¹⁷¹.

A total of 39 molecules have been synthesized during this study. Among these, 28 novel molecules have been successfully obtained and fully characterized.

Tafenoquine. Study of 8-aminoquinolines as potential antimalarial drug candidates

Table 3.10. Overview of every single step yield and the overall yield obtained for the 9-steps synthesis of the seven 8AQ synthesized and the two 12 intermediates. In brackets is identified the intermediate obtained through the assigned synthetic step.

Product	Exp. ID	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9	Overall yield %
		yield % [5]	yield % [6]	yield % [8]	yield % [9]	yield % [10]	yield % [11]	yield % [12]	yield % [14]	yield % [15]	
15{1}	LMF270	98	87	37	73	95	84	98	52	83	8
15{2}	LMF274	98	87	37	73	95	96	79	28	93	4
15{3}	LMF269	98	87	37	73	95	85	100	33	89	5
15{4}	LMF261	98	87	37	73	95	91	79	34	94	5
15{5}	LMF273	98	87	37	73	95	75	94	31	93	4
15{6}	LMF306	98	87	37	73	95	82	95	43	90	7
15{7}	LMF308	98	87	37	73	95	28	74	43	100	2
12{8}	LMF265	98	87	37	73	95	89	57	-	-	11
12{9}	LMF264	98	87	37	73	95	83	13	-	-	2
Mean		98	87	37	73	95	79	77	38	92	5

3.5. Biological activity

Many previous studies have reported the antiplasmodial potential of 8AQ against multiple cycles of the parasite species^{172,173}. Actually, the discovery of new antimalarial drugs has been normally achieved departing from first exploratory attempts with well-established cultures models such as *P. falciparum* (SD7 and Dd2 strains) blood stages or *P. berghei* liver stages focusing on inhibition of the fast growth of in-blood asexual replication and liver schizonts but disregarding to address the hypnozoite biology¹⁷⁴. In fact, as Tafenoquine is preferably prescribed for *P. vivax* treatment, its *in vitro* study against this specie would be the choice of interest for the biological assessment of the developed analogs. However, *P. vivax* continuous *in vitro* culture is so difficult to standardize and yet has become a challenge for those research groups studying the parasite. The nature of its difficult maintenance relies in its parasitemia dynamics (in culture, the ability to re-invade new host cells is lost) and the amount of days for preserving a *P. vivax* culture¹⁷⁵. Still, at the light of the abovementioned hurdles, there are emerging protocols such as the described by Russell *et al.* for culturing *P. vivax* which forces the laboratories to be located in endemic countries in order to isolate freshly collected samples¹⁷⁶.

To sum up, it is a common pre-screening practice to use one or more of the aforementioned *P. falciparum* or *P. berghei* culture models as a fast and inexpensive method for first assessment of a compound library. Consequently, it has been determined the IC₅₀ values for *in vitro* growth inhibition assay for blood stages of *P. falciparum* with tafenoquine, its seven analogs **15**{1-7} and eight intermediates (**14**{1-7} and **12**{8}) which are not included in the original Markush combinatorial library.

Intraerythrocytic growth inhibition of strain *P. falciparum* 3D7A was determined following the detailed procedure explained in Chapter 5 which adapts the methodology from Desjardins *et al*¹⁷⁷. Asynchronous parasite cultures (ca. 70% rings) were incubated with 3-fold serial dilutions of the drugs indicated for 24h. Tritiated hypoxanthine was added and incubated at 37 °C for an additional 24 hours. This reagent is a radiochemical agent commonly used to measure whole-cell activity of antimalarial drugs in parasite studies. Parasite viability was finally measured and expressed in terms of the half maximal inhibitory drug concentration (IC₅₀) to compare potency and selectivity.

All the molecules tested, with the corresponding library identification (for the rationally selected compounds), their experimental identifier and their structure are shown in Figure 3.30. The compounds were synthesized by the method above outlined (resumed Figure 3.19, page 78) and tested by Dr. Franisco-Javier Gamo and Dr. Sonia Moliner from GSK-Global Health Medicine R&D Unit.

RATIONAL SELECTION - TAFENOQUINE ANALOGS

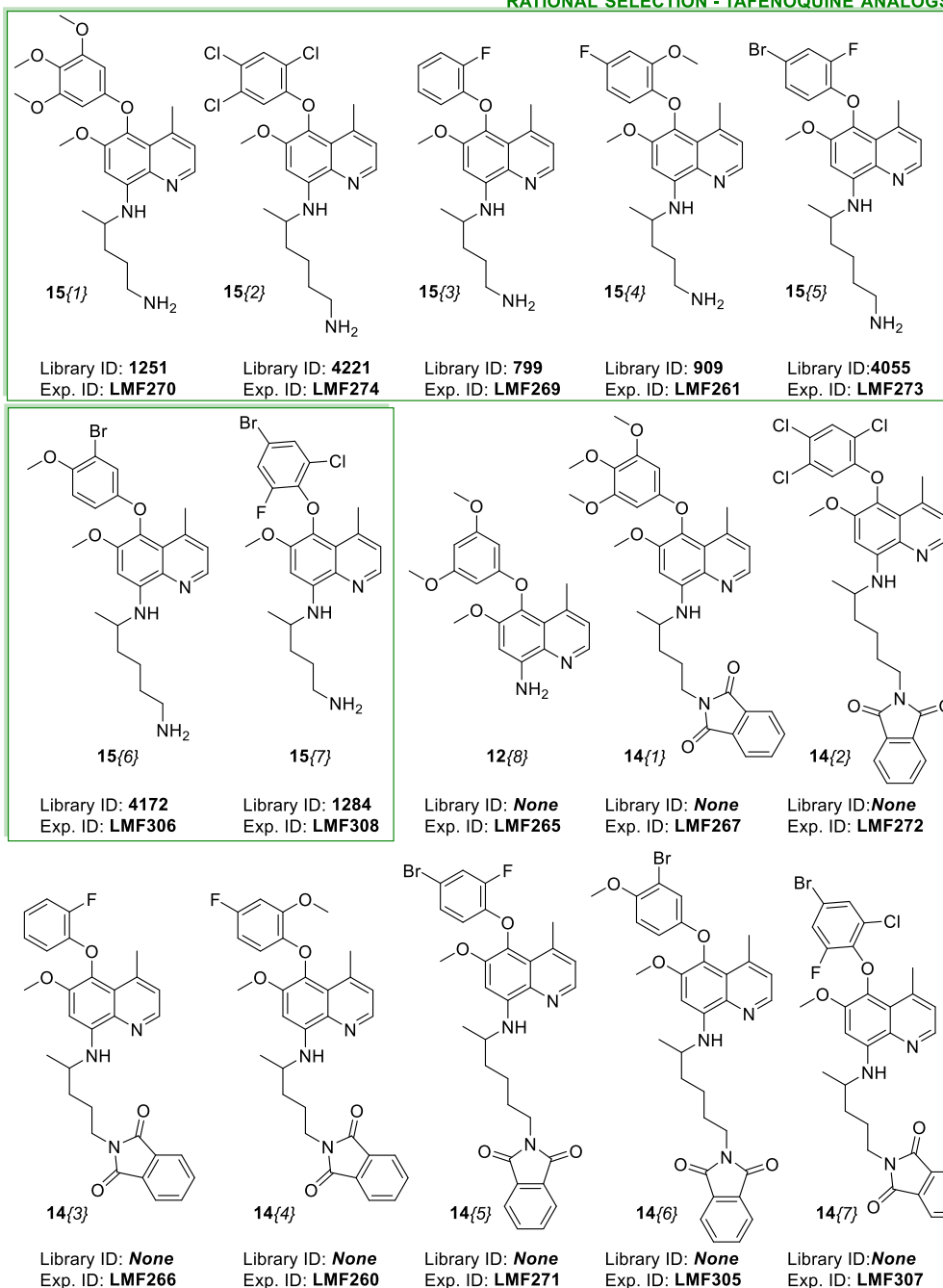


Figure 3.30. Library ID, experimental ID and structure of the molecules tested in vitro. Those molecules present in the green squares are analogs present in the combinatorial library derived from Tafenoquine's Markush structure (Tafenoquine rationally selected analogs present in the chemical space of study)

Results later shown in Figure 3.31, Figure 3.32 and Figure 3.33 are summarized in Table 3.11. Chloroquine has been tested in parallel as a standard quality control inhibitor frequently used in the assay. At present, this drug, which was once the convenient primary chemotherapeutic means of malaria control and treatment, is not further in used as this specie has developed resistance against the 4-aminoquinoline^{178,179}. Tafenoquine was tested in parallel too to contrast the synthesized compounds biological results with the original hit of the database.

Table 3.11 IC₅₀ results summary for the molecules tested. In grey are represented the control molecules; tafenoquine (the original combinatorial library hit) and chloroquine. In yellow are represented the seven rationally selected molecules. In white, the extra molecules tested not involved in the original Markush combinatorial library.

Compound	Average IC ₅₀ (μM)	Standard Deviation IC ₅₀	pIC ₅₀
Chloroquine	0.021	0.002	7.68
Tafenoquine	4.122	0.657	5.38
LMF274	4.159	0.620	5.38
LMF260	4.787	7.351	5.32
LMF306	5.126	0.371	5.29
LMF308	5.135	4.941	5.29
LMF273	6.388	0.729	5.19
LMF270	7.236	1.421	5.14
LMF271	7.964	1.771	5.10
LMF307	8.504	0.558	5.07
LMF272	8.715	2.089	5.06
LMF266	9.420	2.476	5.03
LMF269	10.243	1.094	4.99
LMF267	11.244	2.848	4.95
LMF261	11.876	1.164	4.93
LMF305	12.013	1.977	4.92
LMF265	36.626	4.941	4.44

Tafenoquine. Study of 8-aminoquinolines as potential antimalarial drug candidates

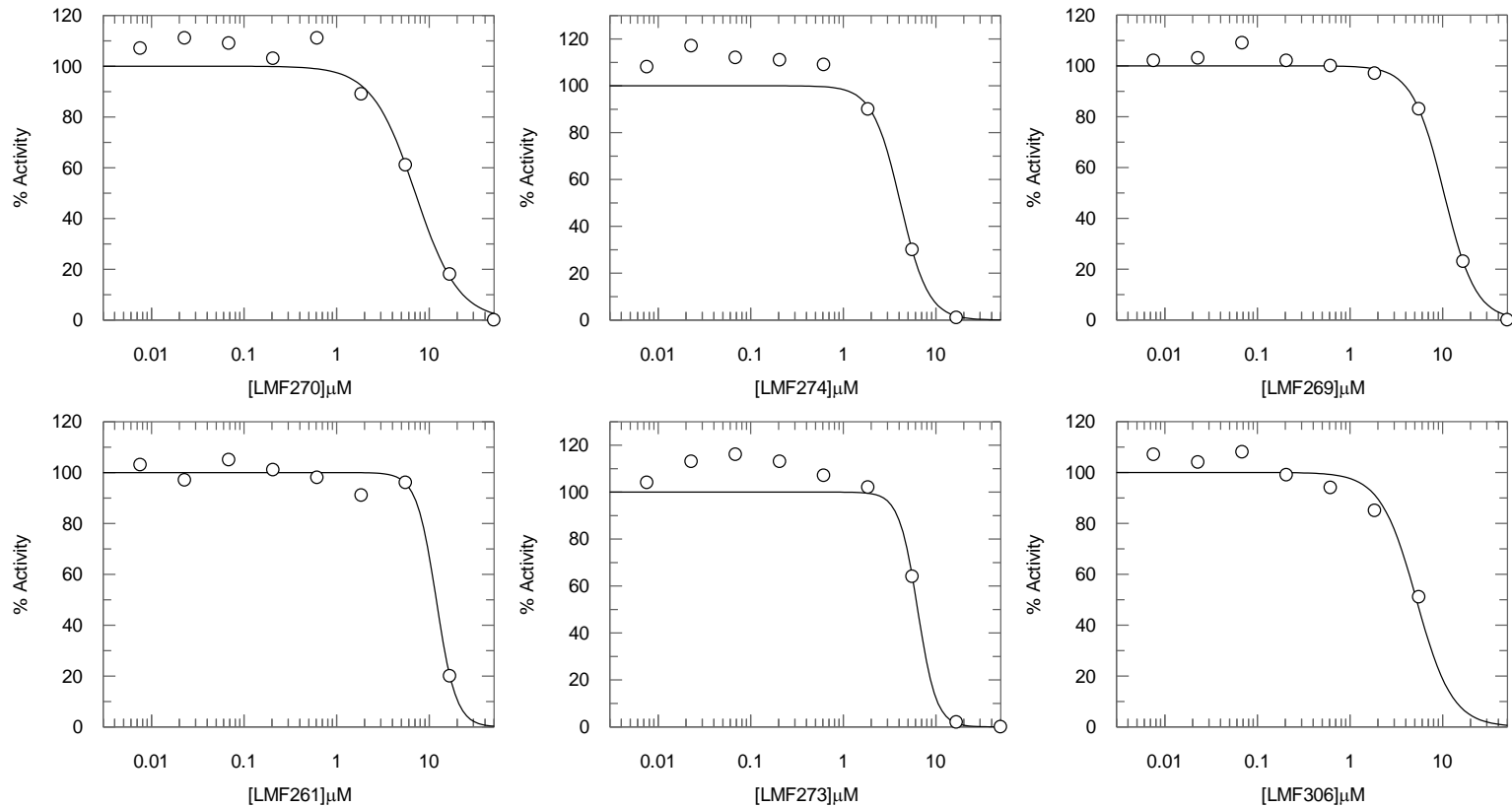


Figure 3.31 The 48-h drug susceptibility assay using strain *P. falciparum* 3D7A testing LMF270, LMF274, LMF269, LMF261, LMF273 and LMF306

Chapter 3

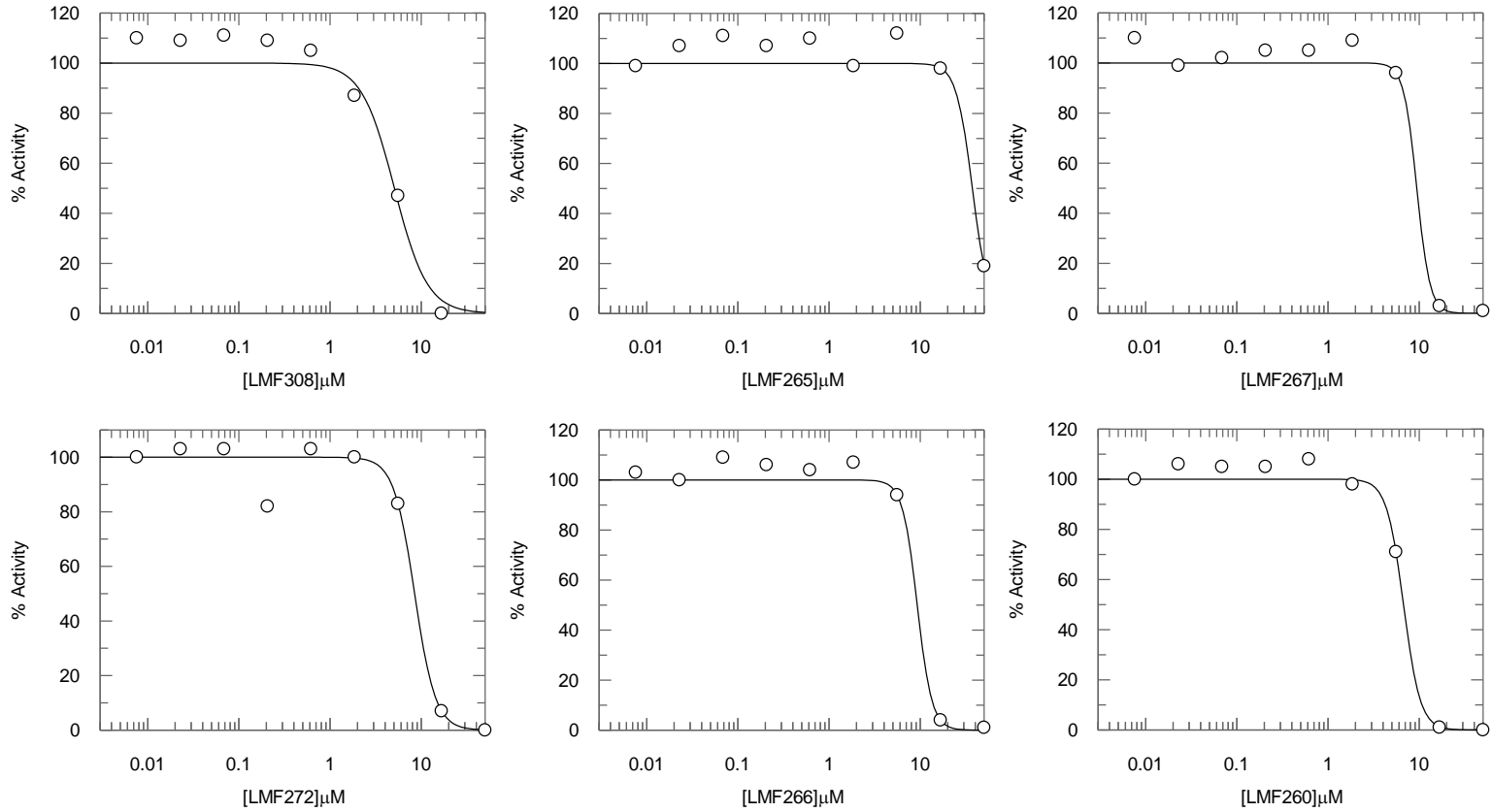


Figure 3.32 The 48-h drug susceptibility assay using strain *P. falciparum* 3D7A testing LMF308, LMF265, LMF267, LMF272, LMF266 and LMF260

Tafenoquine. Study of 8-aminoquinolines as potential antimalarial drug candidates

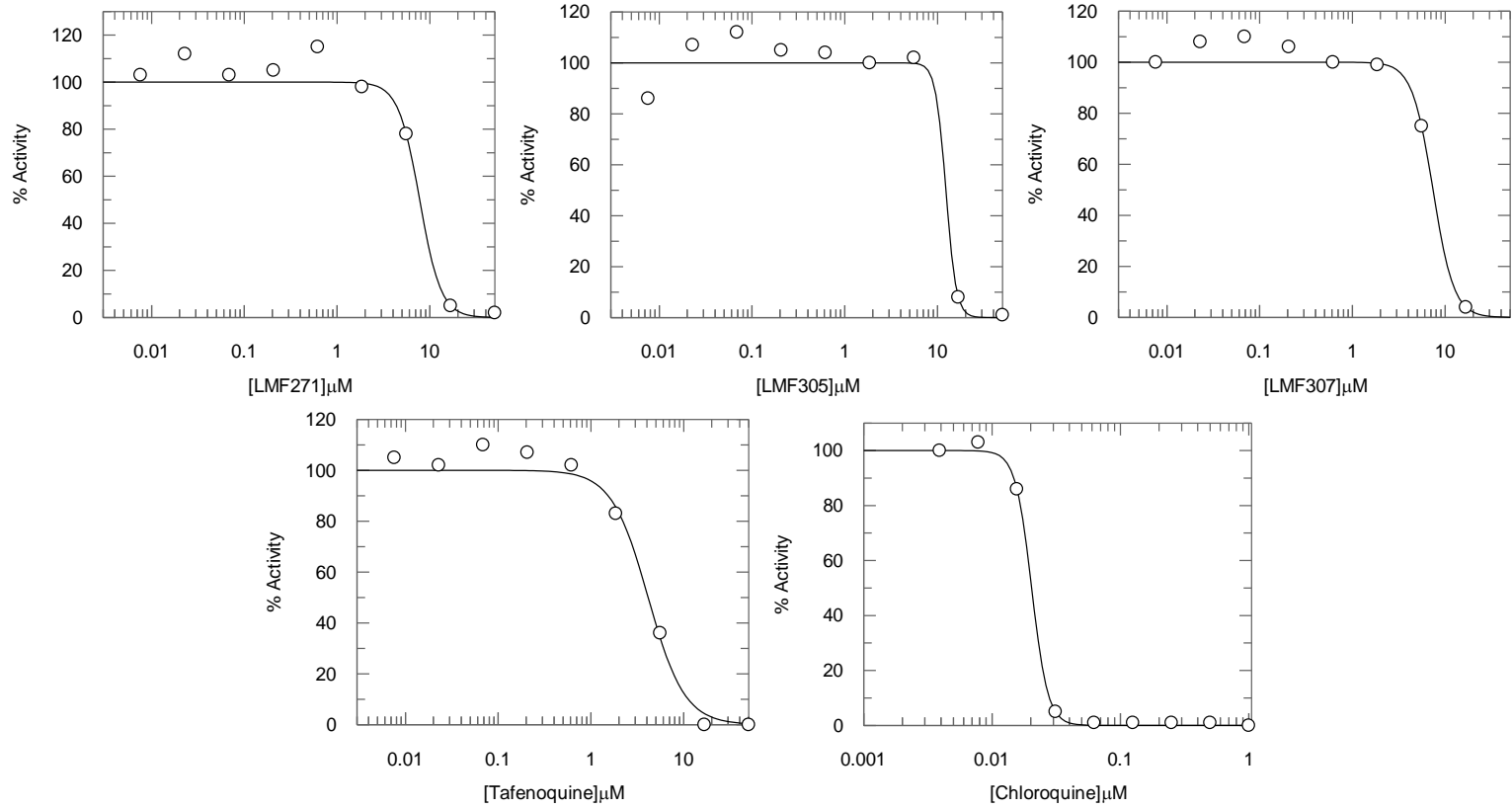


Figure 3.33 The 48-h drug susceptibility assay using strain *P. falciparum* 3D7A testing **LMF271**, **LMF305**, **LMF307**, Tafenoquine and chloroquine (latter two used as controls).

Thirteen out of the fifteen compounds tested displayed a great degree of effectiveness, with IC_{50} values below $12\mu M$. This limit is the accepted in-house threshold (three times the IC_{50} activity measured for Tafenoquine) in GSK-Global Health Medicine R&D Unit for the screening study of first *in vitro* inhibition growth trials against *P.falciparum* blood stages when it is aimed to find new leads alike to tafenoquine. These set of molecules are at present continuing with further pre-clinical assays to study their viability and, if any of them succeed these first steps, the candidates will be ulterior tested against *P.vivax* in endemic areas.

At the light of the results, it first stands out the close relationship in terms of biological inhibitory activity of **LMF274** ($IC_{50}=4.159 \pm 0.620\mu M$) compound with Tafenoquine ($IC_{50}=4.122 \pm 0.657\mu M$), the original hit. Even more, **LMF260** compound, which is not present in the original Markush combinatorial library, has also showed interesting IC_{50} measured values although its large standard deviation has discarded it as an optimal antimalarial candidate in the present results. More replicates should be considered in further studies to proof the suitability of this compound. Above all, five out of the seven candidates (**LMF274**, **LMF306**, **LMF308**, **LMF273** and **LM270**) have shown interesting biological activity results and, even more, four molecules not considered in the original Markush combinatorial library (**LMF271**, **LMF307**, **LMF272**, **LMF262**) have also shown great IC_{50} values ranging from 7.964 to $9.420\mu M$ opening the doors for the search of new 8AQ structures as an alternative for the tafenoquine's Markush-defined analogs. Indeed, the fact that the predecessors containing the phthalimide protecting group have shown interesting activity values may suggest the further study in C8 chain derivatization or elongation.

All together, regarding the aim of this research, these results provide *in vitro* confirmation of experimental antiplasmodial activity of many of the tested 8AQs providing useful information to optimally describe the behavior of the analogs depending on their chemical space location.

3.6. Results discussion

The chemical space described by the combinatorial library developed by the Markush structure of Tafenoquine reported patents^{109,155} contains 25,472 analogs. Among these, only 58 molecules have been described until date in literature poorly representing the dataset chemical diversity in terms of chemical space coverage. The database has been divided in 10 clusters using HRC algorithm using the average linkage method and Euclidean metric. This has been considered a suitable approach to homogeneously classify the data of the combinatorial library of 8AQ analogs.

The database has been curated considering the synthetic feasibility, the druggability of the analogs and the availability of the commercial reagents needed. Consequently, the final accessible chemical space attending to our time and resources limitations was lowered to 7 clusters. Therefore, seven analogs were selected with the criteria of being the nearest to centroid chemically accessible

compound. However, as summarized in Table 3.12, the final achievable representation has been able to increase the knowledge until date of the chemical space of study from a 50% to 70% SC and from 42.5 to 74.8%PC being even the random choice of 10 molecules more representative than BD database in terms of population coverage.

Table 3.12. Summary of coverages represented by the different selected datasets of k=10 clusters. Random selection is considered as the average coverage values calculated for 5,000 repetitions of cherry-picking selection of 10 molecules. In brackets are the number of analogs of each selection.

Selection	Subset size	SC	PC
Bibliographical Data	58	50.0 %	42.5%
Random	10	37.7%	73.2%
Rational Selection of accessible space	7	70.0 %	74.8%

The selected analogs were successfully synthesized with a 9-step synthetic route with overall yields ranging from 2% to 8% with a mean value of 5%. The general obtained yields are mainly limited by the third and eighth steps. A total amount of 35 molecules were synthesized during the process and, among these, 28 had not been previously described.

Biological results were obtained concerning to the IC₅₀ values measured for *in vitro* growth inhibition assay for blood stages of *P. falciparum* with the synthesized compounds. These obtained results can be extrapolated to the chemical space description and give an insight in terms of biologically active regions depiction.

As already introduced, the results concerning to the rational selection of the seven analogs serve as a sample of the biological activity of the analogs coexisting in the same cluster of the selected molecules. Therefore, the ranking of these measured IC₅₀ values from more active (lower IC₅₀ values, in green) to less active (higher IC₅₀ values, in red) and its representation on its original chemical space (described by the Markush combinatorial library of Tafenoquine analogs) could pave the way for further optimization procedures (Figure 3.34).

As expected, Tafenoquine takes part in an active cluster (cluster 7), the selected molecule representing this region (**LMF308**) showed a consistent IC₅₀ value of 5.135 μ M with only one unit of difference in contrast with the original hit. The most active region measured through our rational selection would correspond to the subset described by cluster 3, represented by **LMF274**. Nevertheless, as no large quantitative IC₅₀ differences have been observed for the results measured

for clusters 3, 4 and 7, these three could be clearly regarded as the most active regions. In contrast, clusters 2 and 1 would mainly represent the less active analogs.

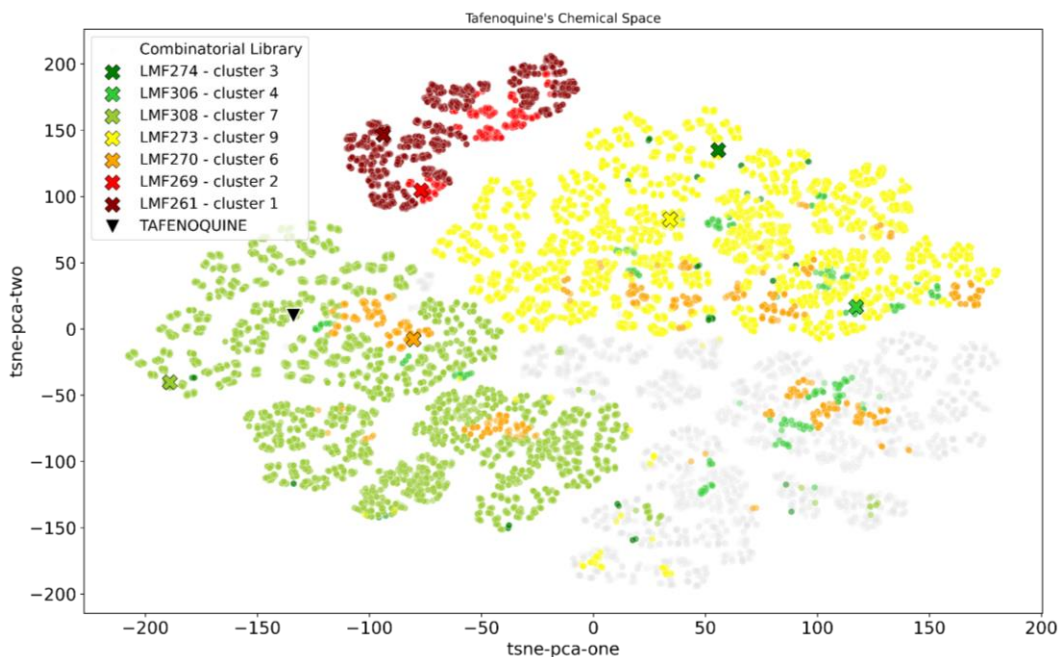


Figure 3.34. Tafennoquine's chemical space with its clusters colored in concordance with the activity measured for the rationally selected analog. In grey are represented de chemical inaccessible molecules during this study (three out of ten clusters).

Above all, by coloring the chemical space attending to the ranking of the obtained scores of the representative molecules, it has been able to update the information of the chemical space description regarding to the biological activity of its compounds. Through its fast visual inspection, it can be observed that the less active compounds are clearly located in an isolated region. Surprisingly, when compared to the bibliographical data known until date, this part of the chemical space which would not be recommended to further study matches with the most explored space by the reported literature until date (see Figure 3.35). These results then reinforce our hypothesis confirming that the actual R&D Methodology is highly focused and putting in doubt whether if the medicinal chemists are wasting time, resources and efforts in less interesting or poorly-active regions.

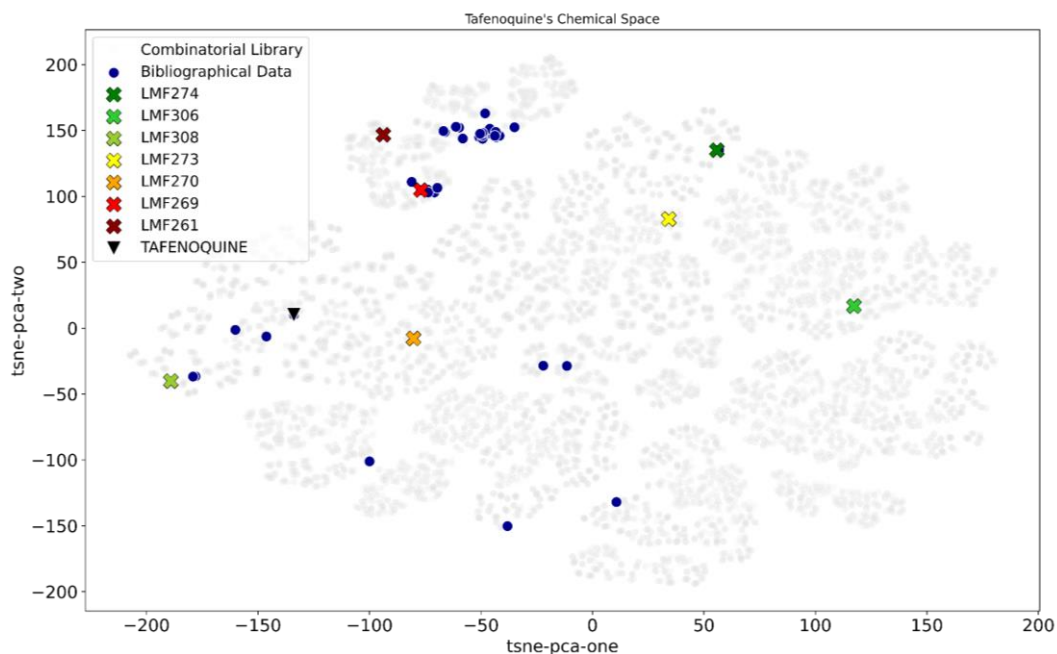


Figure 3.35 Tafenoquine's chemical space with the representation of Tafenoquine (as an upside-down triangle), the bibliographical data known until date (in blue dots) and the tested rationally selected molecules (colored crosses).

Furthermore, in this research, eight alternative molecules which are not included in the Markush structure of discussion has also been tested aiming to find new hits and enlarge the knowledge of the 8-aminoquinoline biological inhibitory activity against malaria infection. Among these, 5 compounds (**LMF260**, **LMF271**, **LMF307**, **LMF272**, **LMF266**) have showed interesting IC_{50} values ranging from 4.8 to 9.4 μM . The **LMF265** compound has been considered as non-active showing the poorest IC_{50} (36.6 μM) among all the subset tested. After comparing the structural features of the fifteen tested molecules we sustain that; (a) the **12** intermediates with a free primary amine group in C8 position do not show antimalarial activity, (b) both compounds with protected and deprotected terminal amines from the sidechain attached to C8 (intermediates **14** and the rationally selected molecules) present antimalarial activity.

To sum up, if considering proceeding with any further exploratory research, at the light of the results obtained in our study, the three most active clusters (3, 4 and 7) would be recommended to be deeply analyzed in order to increase the probabilities to find an alternative optimal hit. Finally, we suggest not only to bitterly explore these family of analogs but new compounds (following structure **14**) outside of the chemical space of study presenting phthalimide protected amines or other structural modification involving C8 chain elongation or derivatization.

3.7. Conclusions

In this study we have applied and defended an alternative methodology based in rational selection which may be more efficient in the early drug discovery stages. Through the deep exploration of the Tafenoquine chemical space described by its combinatorial library derived from the Markush structure stated in its patents, seven compounds with expected antimalarial activity have been rationally selected and synthesized being more representative (with a 70.0% SC and 74.8%PC) than the 58 reported analogs until date.

After their biological assessment, the obtained results have evidenced that there are three regions with highly active analogs and two which would be well unrecommended to be furtherly explored as they contain poor or less active analogs. Therefore, it has been confirmed our hypothesis that rational selection has proven to be a more efficient methodology of exploration suitable for the early drug discovery stages. Moreover, new compounds have been tested presenting interesting biological activity values opening new doors to the development of new antimalarial drugs.

Chapter 4: Dacomitinib. Study of substituted quinazoline derivatives as irreversible inhibitors of EGFR

4.1. Introduction to Dacomitinib

4.1.1. Non-small-cell lung cancer (NSCLC)

One of the biggest threaten to human health at present which more than the 40% of the global research is destined is cancer. Indeed, the find of a cure for the different neoplasms known until date is clearly evidenced by the latest FDA approved antineoplastic drugs (2008-2020), which englobes the 26.7% of the overall approved drugs (see Chapter 2). According to the International Agency for Research on Cancer (IARC)¹⁸⁰, the global cancer burden in 2020 estimated new 19.3 million (Figure 3.1) cases and 9.9 million related-deaths (Figure 4.2).

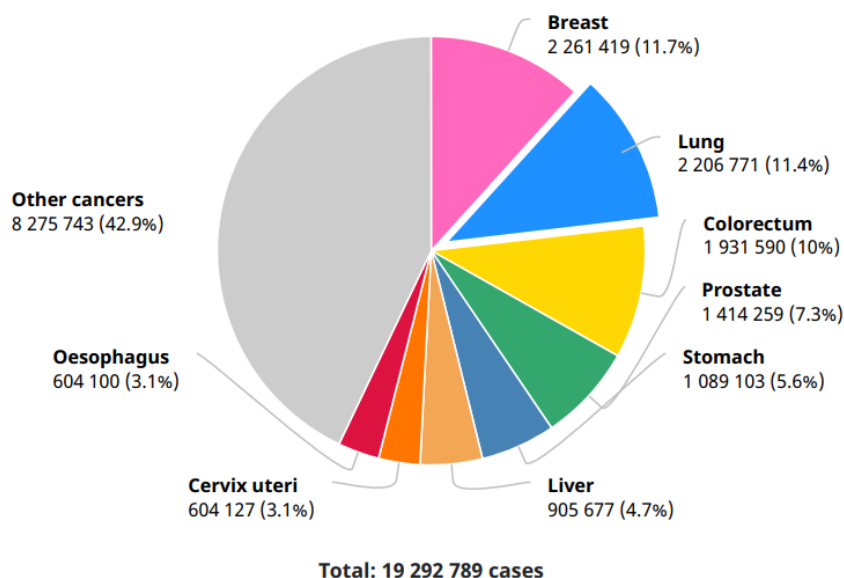


Figure 4.1. Number of new cancer cases in 2020, both sexes, all ages. Source: Globocan 2020 ¹⁸⁰

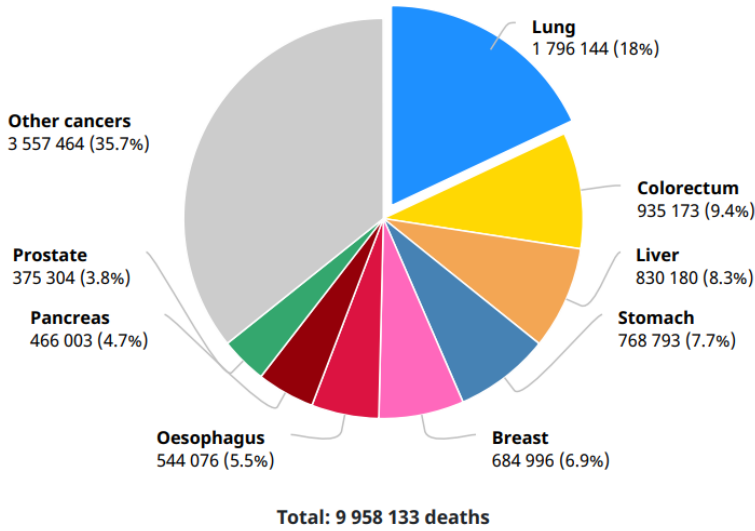


Figure 4.2. Number of cancer-related deaths in 2020, both sexes, all ages. Source: Globocan 2020 ¹⁸⁰

In terms of incidence indexes, the four most common types of cancer present worldwide are female breast, lung, colorectum and prostate cancers and account for more than four in ten (40%) of all new cases. Amongst these, lung-related cancer stands out as the deadliest with a global morbidity rate of 18% (2020). More specifically, the estimations of 2020 dictate that, in Spain, the 20.3% of the general cancer-related deaths were attributed to this type¹⁸⁰.

The world age-standardized incidence rate of lung cancer (Figure 4.3) shows that this disease is at present affecting largely in men rather than in females with 31.5 and 14.6 new cases for every 100,000 men and women respectively. Additionally, Asia and Europe are the most impacted continents by this disease.

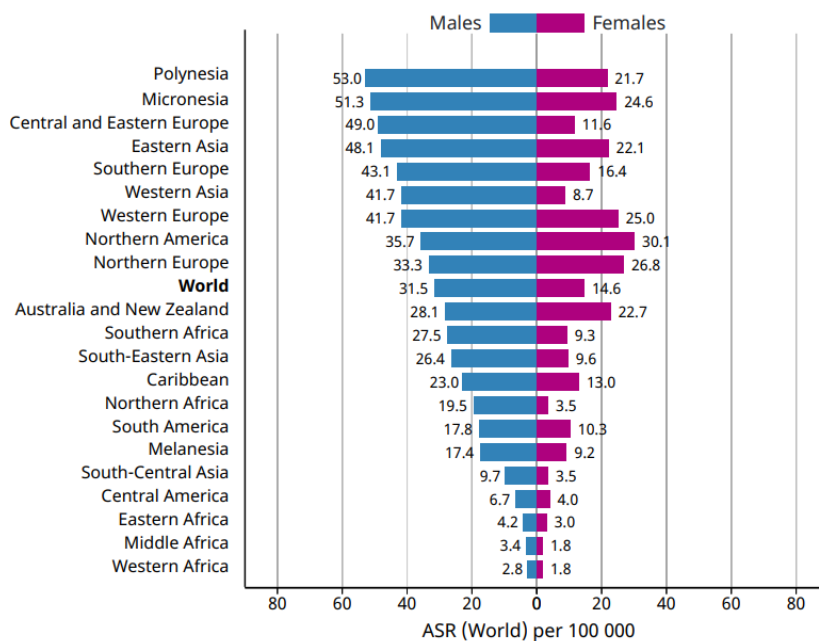


Figure 4.3. Global age standardized incidence lung cancer rates by sex. Source: Globocan 2020 ¹⁸⁰

In general terms, cancer is a multifactorial disease which arises from the transformation and of normal cells into the abnormal or tumor cells driven by a multistage process. Thus, this disease appears as a result of the progressive acquisition and accumulation of several genetic mutations¹⁸¹. These are stored on the cell genome, dysregulating the normal activity of genes that are in charge of cell growth control, genetic stability maintenance and sensitivity regulation to apoptosis¹⁸².

The World Health Organization (WHO) maintains that the genetic mutations can be produced as a result of the interaction between the genetic factors of the host and three external agents; (a) physical carcinogens like ionizing and ultraviolet radiation, (b) biological carcinogens like infections from viruses, parasites or bacteria and (c) chemical carcinogens such as tobacco smoke, alcohol, asbestos or other food and drink contaminants¹⁸³. Therefore, due to the nature of these introduced risk factors, it seems quite obvious that people’s live style plays a key role in the probability of acquiring cancer making it a preventable disease. However, cancer is not only dependent on this but also to age, gender and even geographical situation as evinced with the previous figure. As a matter of fact, the 10% of the cases are attributed to genetic defects disabling the actual capabilities to totally abolish this life-threatening disease^{184,185}.

Particularly, in the case of lung cancer, the cell growth of abnormal cells may occur in any part of the respiratory system. It typically starts in the windpipe (trachea), the main airway (bronchus)

or the lung tissue where tumors can be formed. Depending on the type of cell that the cancer started in, two primary lung cancers can take place: the small cell lung cancer (SCLC) and non-small-cell lung cancer (NSCLC). The most abundant type (around 85 out of 100 cases^{186,187}) is NSCLC. This term groups three major histological subtypes as they behave and respond to treatment in a similar way: (a) adenocarcinoma, (b) squamous cell cancer, (c) large cell carcinoma¹⁸⁸.

The cause of this cancer is primarily related to smoking. Indeed, in 2019, more than the 64% of the global cases are caused by this risk factor and, particularly in Spain the 75% of cases were attributed to it¹⁸⁹. However, other causes may be surveilled such as the exposure to radon gas, certain chemicals in the workplace or the family history of lung cancer.

4.1.2. Tyrosine kinase inhibitors; a key to treat epidermal growth factor receptor (EGFR) and its mutations in NSCL.

Through the above-mentioned process, the uncontrolled growth of new tissue (*neoplasia*) lead to the formation of a tumor which may be either malignant or benign. The first, are capable to invade surrounding tissues and move to distant locations in a process known as metastasis and its treatments generally involve the initial surgical removal of the tissue followed by radiation and/or chemotherapy^{10,183}. When the surgery is not feasible, radiation and chemotherapy become the only available options. The latter, refers to the group of drugs that are used to kill cells – desirably with selective toxicity targeting the affected biochemical pathway– and are also called antineoplastics.

Extensive molecular and genetic studies of lung cancer showed that these have multiple genetic and epigenetic alterations. In brief, it has been demonstrated that it is developed from normal epithelial cells through a multistep process involving successive genetic and epigenetic abnormalities contributing to the initiation, development and maintenance of lung cancer¹⁹⁰.

Epidermal growth factor receptor (EGFR, also known as ERBB1, shown in Figure 4.4) has been accepted among others as a prognosis marker in NSCLC. Specifically, EGFR and its mutations has been widely studied as a main biomarker or biological therapy target in lung cancer as it is known to be overexpressed in a large percentage of clinical cancers. EGFR is a cell surface receptor and such as other protein containing tyrosine kinases domains, catalyzes the transfer of phosphate groups on adenosine triphosphate (ATP) to tyrosine residues of other important proteins leading to a cascade of activated factors¹⁹¹. Specifically, this transmembrane protein regulates important tumorigenic processes including proliferation, angiogenesis, invasion and apoptosis through its overexpression and/or mutation^{192,193}. Indeed, key driver mutations of EGFR have been discovered in the specific subgroups of NSCLC and are the first target for selective treatment with small molecule inhibitors of tyrosine kinases domain.

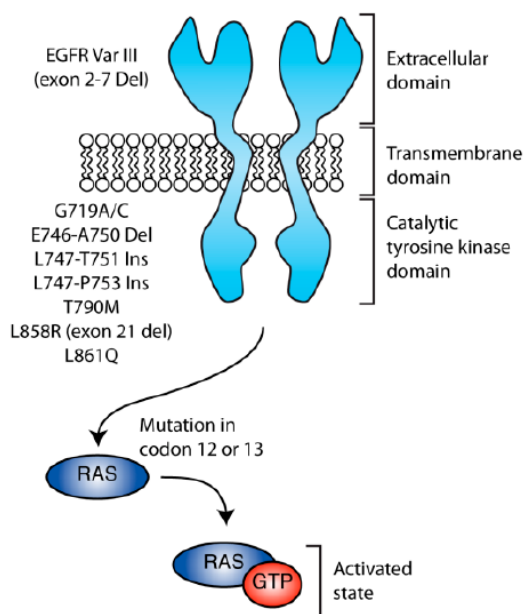


Figure 4.4. EGFR and KRAS main mutations in NSCLC. Source: Aréchaga-Ocampo *et al.*¹⁹²

As a result, there is significant interest in small molecule inhibitors of the human epidermal growth factor receptor family of tyrosine kinases for the treatment of cancer.

Most patients acquire drug resistance to first-line treatment of cancer. In these cases, through diverse mechanisms, the cancer cells develop new routes to maintain a continuous proliferation despite the presence of an EGFR inhibitor. Acquired resistance mechanisms can be originated by on-target secondary mutations, histological transformation or by the acquisition of ‘bypass’ signaling pathways. Alternatively, the causes may be external to tumor cell and be caused by the biology of its microenvironment¹⁹⁴.

The 90% of EGFR mutations are a deletion in exon 19 (del19) or a missense mutation in exon 21 (L858R). Approximately 60% of patients treated with first- or second-generation EGFR tyrosine kinase inhibitors (TKIs) develop a secondary T790M gatekeeper mutation in exon 20 that leads to acquired resistance to EGFR inhibitors. A fourth missense mutation in the catalytic tyrosine kinase domain observed within exon 18 is G719C^{195–197}. Overall, as the response rates to EGFR TKIs depends on this mutations, combination therapies of different therapies involving the three generations of TKIs along with the development of novel agents that bind and inhibit EGFR by a distinct, non-ATP competitive mechanism (Figure 4.5).

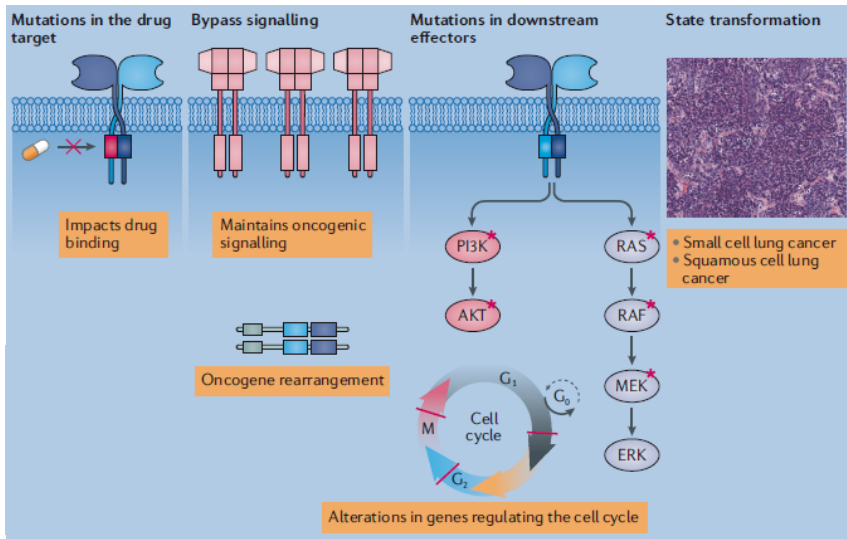


Figure 4.5 Resistance to kinase inhibitors. Source: Cohen *et al.*¹⁹⁴

Since the approval of gefitinib in 2003, the total of approved kinase inhibitors worldwide has risen to 98 drugs, being 71 of them small molecules that have been approved by the FDA. More than the 20% of these agents have been approved for the treatment of NSCLC¹⁹⁸. These illustrate how research has overcome the different outcomes caused by the appearance of new target resistant mutations by the development of three generations of EGFR tyrosine kinase inhibitors which are in present available for mutation-positive NSCLC^{199,200} (Figure 4.6):

- First-generation reversible EGFR TKIs: gefitinib and erlotinib.
- Second-generation irreversible ERBB family blockers: afatinib and dacomitinib.
- Third-generation irreversible wild-type sparing EGFR TKI (EGFR mutant-specific inhibitor): osimertinib.

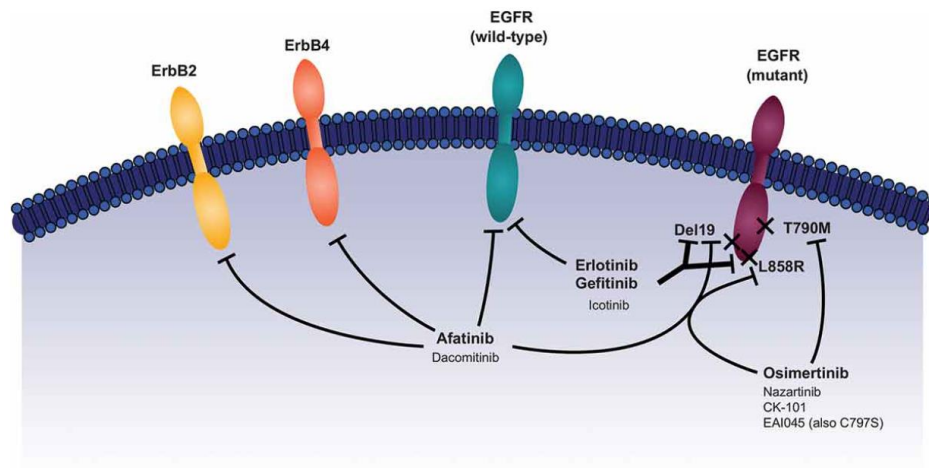


Figure 4.6 Mechanism of action of EGFR TKIs. Source: Kujtan *et al.*²⁰⁰

First and second-generation EGFR TKIs are approved as first-line therapy due the clinical trials performed which has largely demonstrated superior response rates in terms of survivalism. Additionally, the third-generation TKI are being designed to maximize the selectivity for EGFR-activating mutations such as T790M acquires resistance mutation²⁰¹. As a consequence, the present combination of first or second generation TKIs with third generation inhibitors seem to be an effective therapy to maximize overall survival rates and overcome the cancer resistance mechanisms²⁰².

4.1.3. Dacomitinib and substituted quinazoline derivatives as irreversible TKI for NSCLC treatment.

The first TKIs targeting EGFR wild type (wt) and approved for NSCLC in the early 2000s were gefitinib (2003) and erlotinib (2004)²⁰³. Later, as mutation-induced drug resistant mechanisms were developed in many cases arising T790M gatekeeper mutation or the classical EGFR activating mutations (L858R and del19), second generation covalent TKIs were developed including afatinib (2013) and dacomitinib (2018). Finally, third generation TKIs, such as osimertinib (2020), have been recently approved to selectively target activating mutations as well as T790M mutation^{204,205}.

Altogether, the large existing studies and the development and approval of the second irreversible inhibitors neratinib, dacomitinib and afatinib, all of them quinazoline derivatives, have focused the researchers attention in its interesting scaffold in order to explode it in the search of new TKIs^{206–213}. In fact, the review developed by Attwood *et al.*¹⁹⁸ confirmed that this structure is a trending approach in the chemistry of approved kinase inhibitors as it has been widely explored. Indeed, the 14% of the existing approved general kinase inhibitors in present have this moiety (Figure 4.7).

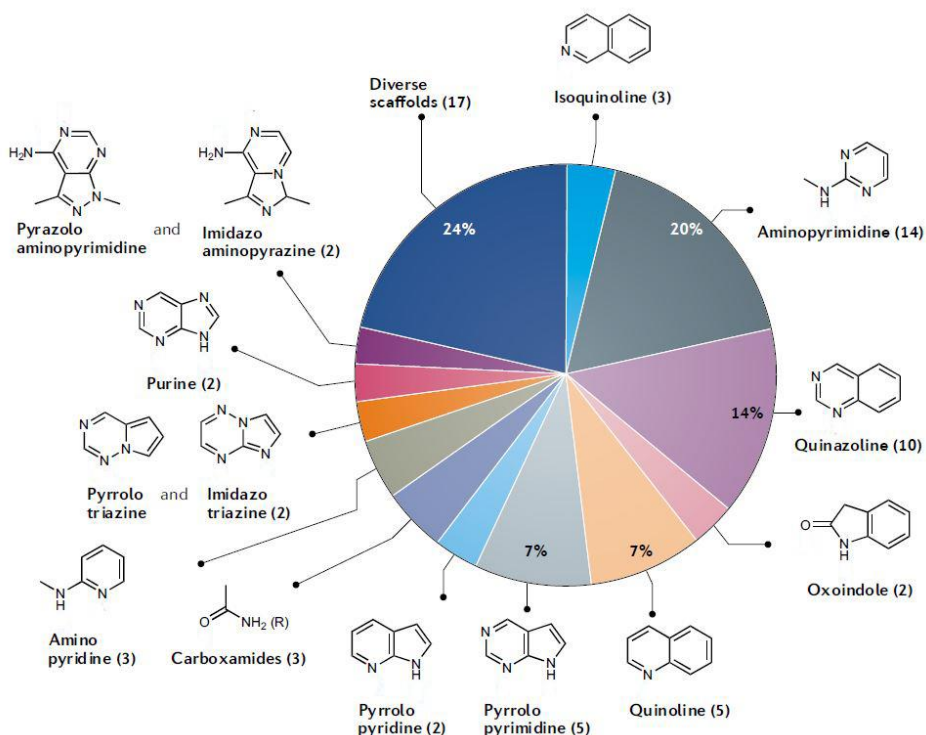


Figure 4.7 Chemical scaffolds used in FDA approved kinase inhibitors. The number of approved drugs using each moiety is provided in brackets. Adapted from Attwood *et al.*¹⁹⁸

Amongst these (Figure 4.9), Dacomitinib (originally termed as PF00299804²¹⁴) stands out as a second-generation TKI developed by Pfizer Inc. which has recently been approved by the FDA in 2018. This active principal ingredient is the actual drug of interest for the treatment of NSCLC presenting EGFR with del19 and L858R mutations. In contrast with first-generation TKI, this inhibitor it is considered a TKI with more attractive pharmacokinetic properties, lower plasma clearance, greater bioavailability, larger volume of distribution and longer half-life. Indeed, from a detailed multicenter, randomized, open-label and active controlled trial, it was demonstrated a significant improvement in progression-free survival in contrast with gefitinib (14.7 vs. 9.2 months respectively)^{215–217}. Finally, this quinazoline derivative is a potent inhibitor of EGFR-activating mutations as well as the EGFR T790M resistance mutation both *in vitro* and *in vivo*.

Above all, the main difference among Dacomitinib and its predecessors (gefitinib and erlotinib) is that this has a potential efficacy as an irreversible pan-erbB inhibitor irreversibly inhibits tyrosine autophosphorylation of EGFR (also known as HER1 or erbB1), HER2 (erbB2) and HER4 (erbB4) as previously showed in Figure 4.6. Its irreversible inhibition is achieved via covalent bonding to cysteine residues of the catalytic domains of the erbB receptors. Specifically, its electrophilic moiety

is subject to a nucleophilic attack by Cys-797 of EGFR (Figure 4.8) performing a covalent bond between ligand and receptor leading to the inhibition of the enzyme via this irreversible binding²¹⁸. As a result, C797S is a common mutation observed as mechanism acquired by EGFR to resist dacomitinib²¹⁹.

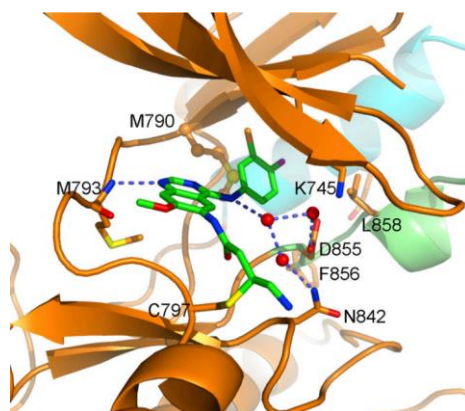


Figure 4.8 Close-in view of dacomitinib recognition in EGFR. Illustrated interactions: The covalent bond between Cys-797; the nitrogen from the quinazoline ring makes a single hydrogen bonding contact with the backbone nitrogen of Met-793; the aryl group of the aniline accessing the end of the ATP-binding pocket, in van der Waals contact with Leu-788, Met-790, Thr-854 and Lys-745; the aniline nitrogen involved in water-mediated interactions with Asn-842 and Asp-855. Source: Gajiwala *et al.*²¹⁸

The original efficacy comparison of these molecules containing the same main scaffold is shown in Figure 4.9. These results, shown as IC₅₀ values which were measured for *in vitro* kinase assay of wild-type erbB proteins, were obtained by Engelman *et al.*²¹⁴ while studying Dacomitinib.

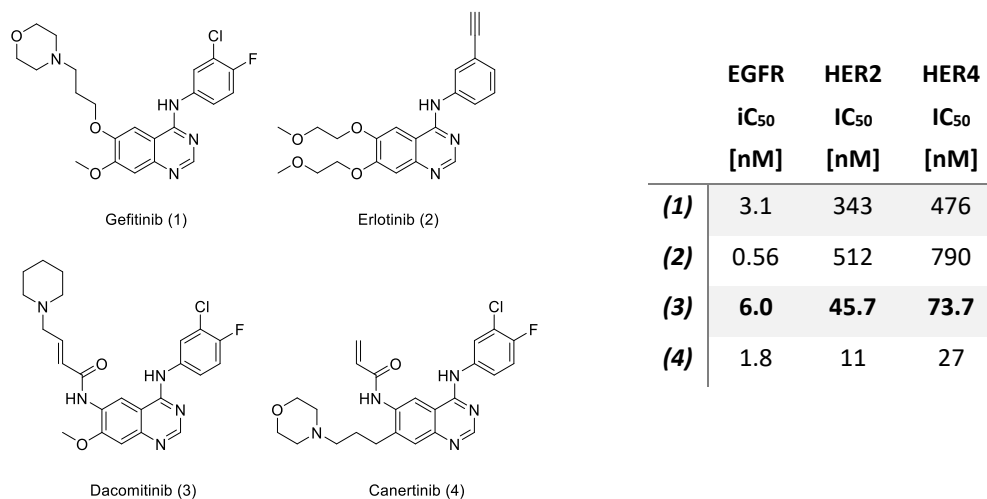


Figure 4.9 Main quinazoline structures and efficacy comparison in terms of IC₅₀ values of each drug using an *in vitro* isolated kinase assay against wild type erbB receptors. Values obtained by Engelman *et al.*²¹⁴.

The results shown in this original study showed Canertinib as a potential pan-erbB inhibitor. Indeed, this molecule is still considered as an experimental drug candidate as it reached the initial phase I of clinical testing conducted in NSCLC patients. However, its development carried out by Pfizer Inc. was discontinued in 2015 due to skin toxicity and the observation of thrombocytopenia²²⁰.

Additional results in EGFR mutant cell lines showing outstanding values suggested that Dacomitinib is effective against a T790M mutation, L858R and 19 deletion activating mutation. These evidences positioned dacomitinib as the actual oral, once-daily, pan-erbB inhibitor indicated as first-line treatment for patients with metastatic NSCLC with EGFR del19 and L858R mutations at present²²¹.

At the light of this scenario, Dacomitinib seems to be a promising candidate in the present discussion as a second practical case of study as it has been recently approved by the FDA (2018) and is an emerging TKI of interest which belongs to a presumably largely explored family of quinazoline analogs. Thus, it is expected that its developed chemical space may be better explored than the previous practical study introduced in Chapter 4.

4.2. Chemical space assessment of Dacomitinib's analogs

4.2.1. Dacomitinib library of analogs and its chemical space

The Markush structure containing Dacomitinib structure can be found in the drug's patent (US20050250761²²²) in their concluding claims section.

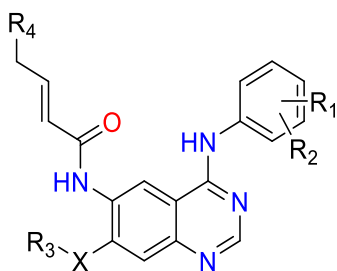


Figure 4.10 Markush structure of Dacomitinib analogs

The complete combinatorial library of anilinoquinazoline analogs depends on three main points: (a) the aryl structure in C4, (b) the chain in C7 position and (c) the enlargement of the α,β unsaturated carbonyl sidechain in C6. However, when looking closely to each substituent definition, a main drawback was found relating to unprecise terminology:

1. **R1 and R2:** The patent establishes that R1 can be any halide from F, Br, Cl and I and R2 the same set or halogens plus hydrogen. Therefore, this point of substitution,

considering all the possible positions available in the five available in the ring and discarding repetitions would lead to 86 possible substructures in C4.

2. **X and R3:** The patent claim defines X as O, S or N. Nevertheless, the description of R3 may trigger more difficulties. As extracted from the text it can be a selection from:

a. "*C₁-C₃ straight or branched alkyl optionally substituted by one or more halogens*". In other words, 4 unhalogenated –in SMILES; C, CC, CCC, C(C)C– and 28 halogenated derivatives – CZ, CCZ, C(Z)C, CCCZ, C(Z)CC, CC(Z)C, C(C)CZ being Z =F, Cl, Br and I –. In total, 32 substructures.

b. "*(CH₂)_n – Het*" being *n* an integer from 1 to 4 and Het consist in 9 different heterocycle: morpholine, piperidine, piperazine, piperazine-N(C₁-C₃ alkyl) (4 inherent possible combinations considering isopropyl and unbranched alkyls), pyrrolidine or imidazole. In total, this group would consist in 36 combinations (4x9) if only considering that the attachment to heterocycle is to its NH group.

In total, 272 possible substructures ((32+36)x4) could be considered for C7 derivatization point

3. **R4:** "*(CH₂)_m – Het*" being *m* an integer from 1 to 3 and Het consist in 12 different heterocycle: morpholine, piperidine, piperazine, piperazine-N(C₁-C₃ alkyl) (4 inherent possible combinations as mentioned before), pyrrolidine, imidazole, azepane, 3,4-dihydro-2*H*-pyridine or 3,6-dihydro-2*H*-pyridine. However, the complexity of this substitution point relies in the possibility of presence of 0 to 3 substitution within each heterocyclic moiety with 18 possible functional groups: C₁-C₃ alkyl (4 combinations), halogen (F, Cl, Br, I), OH, NH₂, NH(C₁-C₃ alkyl) (4 combinations) or N(C₁-C₃ alkyl)₂ (4 combinations). As an example of these vague indetermination for the heterocycle substitution and assuming once again that the attachment to the heterocycle is to its NH, for the unique case of piperidine, the possible combinations would lead 187,410 possible substructures as it presents five aliphatic carbons that could show two groups attached. Specifically, these substructures are enumerated without distinguishing the chiral conformation of them.

The total amount of 1,854,474 distinguishable substructures indorsed in R4 classified by its main ring are listed below:

- a. Morpholine derivatives: 94,503
b. Piperidine derivatives: 187,410

- c. Piperazine derivatives: 540,336 (162,324 + 4 x 94,503)
- d. Pyrrolidine derivatives: 94,503
- e. Imidazole derivatives: 20,577
- f. Azepane derivatives: 322,545
- g. 3,4-dihydro-2H-pyridine derivatives: 297,300
- h. 3,6-dihydro-2H-pyridine derivatives: 297,300

Hence, at the light of the ambiguous description of R4 that would lead to an enormous description of the combinatorial library of Dacomitinib involving more than 10 billion analogs (43,379,855,808=86x272x1,854,474), a new approach was considered for this second case of study. In this case, expecting to better explore a synthetic feasible fraction of the gigantic library described above, the rational study of the drug derivatives would be applied to the combinatorial library of the already described analogs by the inventors. In other terms, the claim 5 of the abovementioned patent contains 51 described compounds that have already been synthesized and therefore protected by Pfizer Inc. The combinatorial library of these known analogs would build the explored Markush structure by the drug developers containing the subfraction of the chemical space explored until date. The fragment combination of these 51 analogs present in the claim 5 of the patent resulted in a database of 16,530 analogs as shown in **¡Error! No se encuentra el origen de la referencia..** This is the result of the combination of the 58 possible anilines in C4, 19 substituents in C6 and 15 sidechains in C7 (58x19x15).

Table 4.1 Fragments used to build the 16,530 compounds library of Dacomitinib analogs.

R1	R2	X-R3		R4	
-H	-H	-O-C	-S-CCF	-CN1CCCCC1	-Cn1cncc1
-F	-F	-S-C	-O-CC(F)(F)F	-CN1CCC(F)CC1	-CN1CCCC1
-Br	-Br	-N(H)-C	-O-C(F)CF	-CN1CCCC(F)C1	-CCN1CCCCC1
-Cl	-Cl	-O-CC	-O-CC(F)F	-CN1C(F)CCCC1	-CCN1CCC(F)CC1
		-O-CCC	-O-CCCN1CCOCC1	-CN1CCOCC1	-CCN1CCCC(F)C1
		-O-C(C)C	-O-CCN1CCCCC1	-CN1CCCCC1	-CCN1C(F)CCCC1
		-O-C(F)(F)F		-CN1C=CCCC1	-CCN1CCOCC1
		-O-CF		-CN1CC=CCC1	-CCN1CCCCC1
		-O-CCF		-CN1CCNCC1	-CCCN1CCCCC1
				-CN1CCN(C)CC1	

4.2.2. Choice of a suitable clustering methodology

In order to perform a further rational selection of a tractable number of representative molecules to be synthesized in the laboratory, the first assessment of the chemical space was performed with the proper choice of a suitable clustering methodology. In this case, the chemical space was firstly assessed attending to the classification of the data in $\sqrt{N} = 129$ clusters. As with Tafenoquine's study, it was studied the data distribution for the same previously reported (see Chapter 3) nine clustering and two partitioning algorithms. Binning and optimum variance binning (OV BIN) algorithms were able to classify the data in 102 and 101 occupied bins respectively.

Consistent with the previous case, to contrast all the classificatory methodologies, the centroid of each cluster/bin was picked leading a selection of 129 candidates for each case – with the exception of the 102 and 101 compounds selection for binning and OV binning partitioning methodologies– and the values of space and population coverages were calculated when overlapping these subselections on the other clusters resulting from the different algorithms. The obtained results are shown through two heatmap plots showed in Figure 3.7 and Figure 3.8.

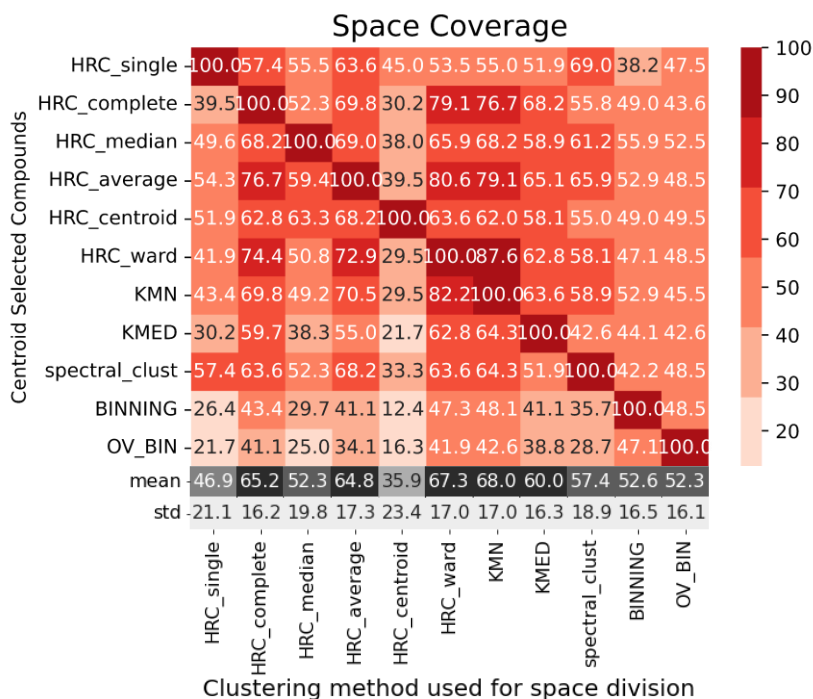


Figure 4.11 Space coverage heatmap for the clustering and partitioning methodologies assessing Dacomitinib's chemical space in 129 groups.

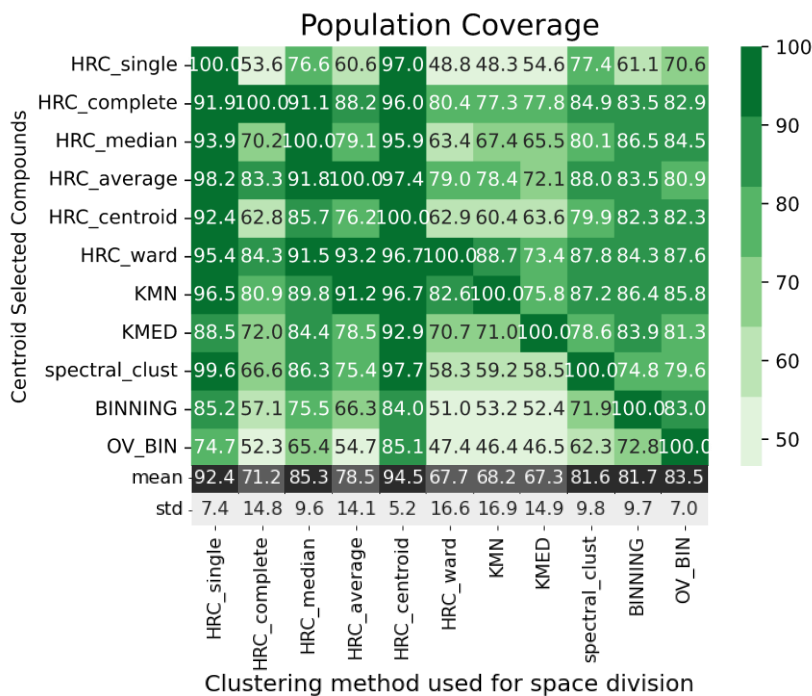


Figure 4.12. Population coverage heatmap for the clustering and partitioning methodologies assessing Dacomitinib's chemical space in 129 groups.

Firstly, as already demonstrated with the previous study, HRC single and HRC centroid were discarded as these clustering methodologies trend to present large, overpopulated clusters leading to unbalanced population distribution. This is evidenced by the lowest SC (with a mean of 46.9% and 35.9% respectively) and largest PC (with a mean of 92.4% and 94.5% respectively) measured in contrast with the other classificatory approaches. Indeed, cluster 115 in HRC single contains 4,173 analogs (the 25% of the database) and the cluster 97 in HRC centroid contains 8,408 compounds (the 51% of the database).

Another point of view to demonstrate this first indication can be reflected in the population distribution inside the different clustering and partitioning methodologies performed. These are revealed via boxplot representation (shown in Figure 3.9 and Figure 3.10). As anticipated, HRC single and HRC centroid are the two most notorious algorithms showing a larger overcrowded clusters as outliers (showed in Figure 3.9 as flying circles). However, this new perspective reveals that also HRC median, spectral clustering and both binning methodologies lacks of uniformity in its population distribution. These three classificatory methods were secondly discarded considering the aim of avoiding overcrowded cluster formation.

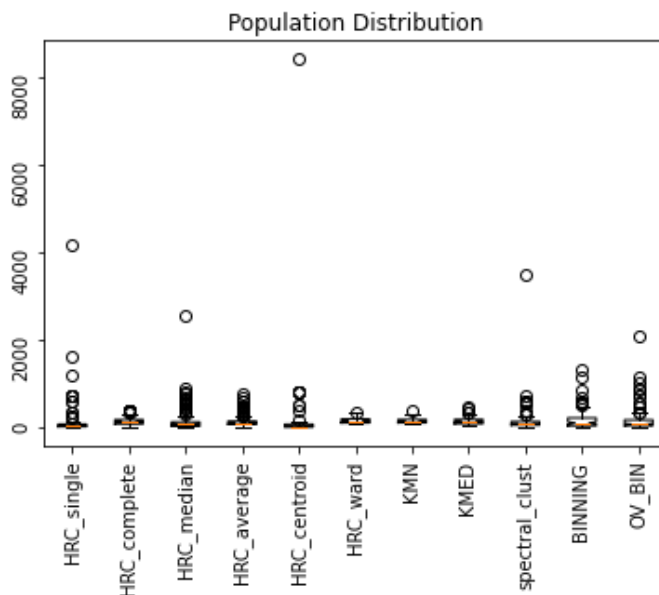


Figure 4.13 Population distribution of Dacomitinib database of analogs for k=129 considering outliers

After these two first clustering subset rejections, HRC complete, HRC ward, HRC average, KMN, KMED and binning were left for a further assessment. Figure 3.10 shows that the four clustering approaches show a balanced distribution. Therefore, the use of any of these methodologies could be suitable for the performance of a rational selection. Lastly, for Dacomitinib's chemical space assessment, KMN was chosen in contrast to the first study (which explored the chemical space with a HRC algorithm) to apply this highly commonly used algorithm as an example of stochastic approach. However, in order to ensure the repeatability of the study and save the classified data for further chemical space explorations, the cluster indexes were saved and are available in Annex VIII.

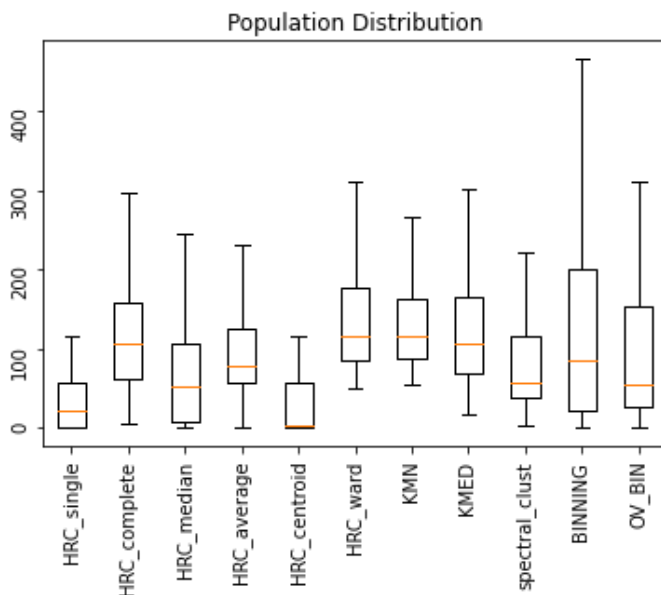


Figure 4.14 Population distribution of Dacomitinib database of analogs for k=129 without outliers

4.2.3. The explored space known until the date

After choosing KMN clustering as the clustering methodology of choice to assess this second study concerning to Dacomitinib's chemical space, research of the compounds described in bibliography was conducted to discuss the explored space until date. This was again accomplished by inspecting programmatically the publicly available information in PubChem¹⁵⁶ when crosschecking the database of study.

A total amount of 60 coincident molecules were found in bibliography. Unfortunately, for all cases, the coincident molecules found in this source do not present any publicly available biological data as they are all related to the already cited Dacomitinib's or other related patents (such as US20130274275 or US20100190977). As this set contained the 51 original patent molecules but the rest of the molecules (9) were not found explicitly defined in the mentioned patents, BD was defined as the original bibliographical dataset of 51 compounds found in the 5th claim. The sorted list of this bibliographic data (BD) with the Library ID of the database and corresponding CID of PubChem per each compound can be found in Annex VII. Therefore, it is expected that all the set of molecules has been tested to treat NSCLC as it is prescribed for its original hit. Furthermore, the same patents leaves the door open for further applications such as it defends that these novel compounds *"are useful in methods of treating, preventing or inhibiting proliferative diseases, including cancer, atherosclerosis, restenosis, endometriosis and psoriasis"*²²².

When representing the BD found in the drug's chemical space in the first three principal components (Figure 3.11 which explains the 55.4% of the variance) it can be easily perceive that the data known until date is largely spread over the chemical space. In this case, it is assumed a large representativeness of the BD as the chemical space itself has been built up from the fragment combination of this set of molecules. Still, there are some unexplored regions that, in the case of being synthetically accessible, would be interesting to be deeper explored.

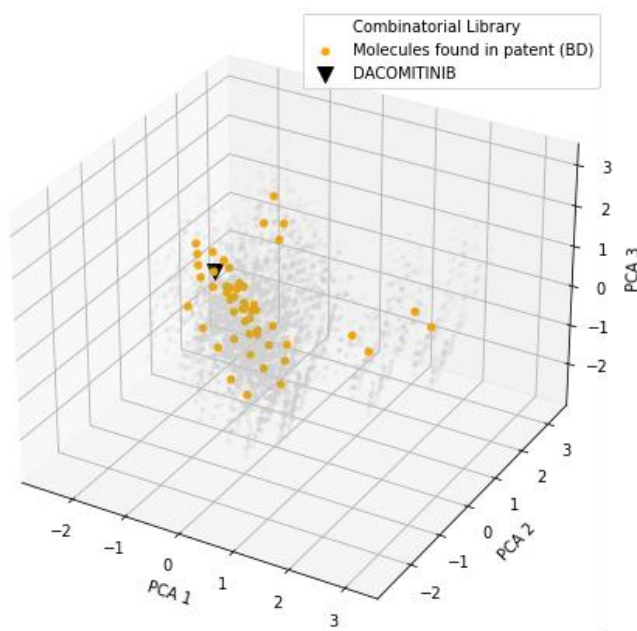


Figure 4.15 First three principal component plot displaying BD in Dacomitinib's chemical space

Actually, for the standard size of $\sqrt{N} = 129$ clusters calculated via KMN, the overall subset of 51 analogs (N_{BD}) would represent the 23.3% SC and the 30.5% PC. In contrast, a random choice of N_{BD} analogs (calculated as the mean of 5,000 repetitions) would better represent the chemical space with 31.8% SC and 36.9% PC and a rational selection of 51 analogs could maximally cover the 39.5% SC and 57.2% PC (as the 51 firsts most populated clusters contain a total sum of 9461 analogs). In contrast with the previous study with antimalarial drugs (Chapter 3), this database shows a significant representativeness through its BD sample as already evidenced by the visual inspection of its chemical space. As expected, the same trend occurs when clustering the space in N_{BD} clusters. The BD data covers the 47.1% SC and 66.7% PC while a cherry-picking approach could achieve to cover a mean value of 57.4% SC and 69.6% PC. Still, even the BD dataset cannot achieve to cover all the chemical space of its derived combinatorial library leaving more than the half of it unexplored. A better

exploration would be suggested by picking 51 rationally selected samples to elucidate the biological behavior of the other half of the chemical space.

4.3. Rational Selection analysis

However, attending to the time and resources of the ongoing research, the rational selection of the same amount of BD analogs was not reasonably attainable. Instead, the chemical space was divided into 10 clusters using KMN algorithm to later on describe the database in an optimal subsequent representation of ten molecules if the further data curation grants it.

This clustering algorithm led to the classification of the data as shown in in Figure 3.15. However, part of this chemical space might be inaccessible considering factors that may impede the synthesis of its inner analogs.

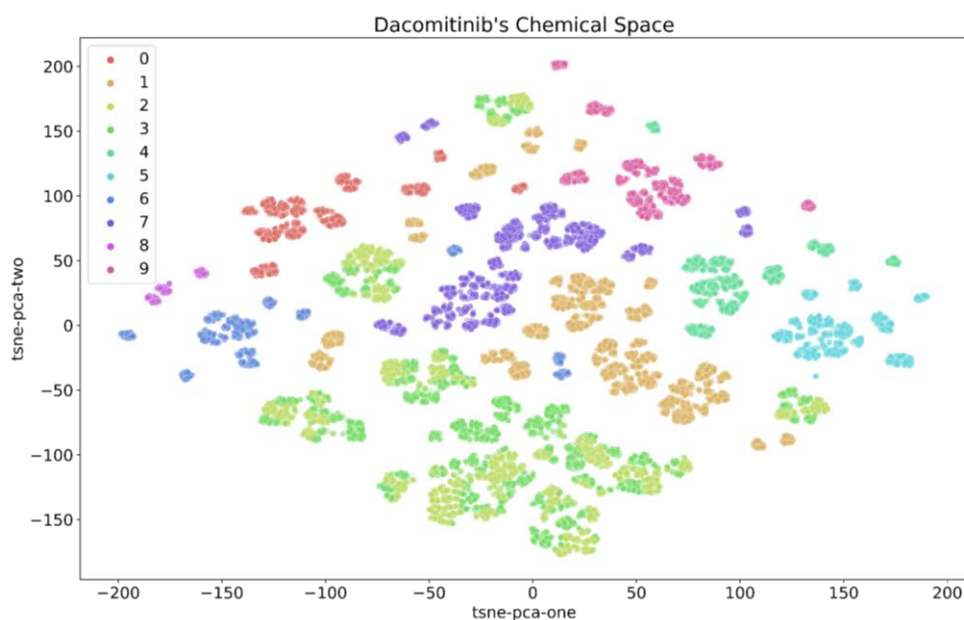


Figure 4.16 t-SNE representation of Dacomitinib's chemical space after KMN clustering ($k=10$)

4.3.1. Database curation; shaping the accessible chemical space.

For Dacomitinib's chemical space, only two factors have been considered in order to shape the real accessible space considering our time and resources limitations and the drug-likeness of the derivatives: Lipinski's rule of five and the commercial availability of the reagents.

In contrast with the previous discussed case, the synthetic feasibility has been not questioned with this database as this has been built up from a previous subset of already synthesized

analogs (the 51 explicitly described molecules). Its synthesis consisting in a 8-step route, which will be further discussed in section 4.4, did not present any limitation which could restrict the synthesis of Dacomitinib's dataset of analogs.

a. Lipinski rule of five

The first filter applied to the database of study was the consideration of Lipinski's rule of five in order to only ponder those parts of Dacomitinib's chemical space containing *drug-like* molecules by its molecular properties¹⁶⁴. As a reminder, these rules predicts that poor absorption or permeation is more likely when:

1. There are more than 5 hydrogen donors
2. There are more than 10 hydrogen bond acceptors
3. The molecular weight is greater than 500
4. The logP is greater than 5

In order to avoid a highly restrictive threshold, we defined this first filter to select those molecules which violated none or at least one of the rules leading to a curated set of 4,895 Dacomitinib analogs.

b. Commercial availability of the reagents

The second restriction considered after designing the synthetic route consisted in checking the commercial availability of the reagents needed. All the solvents, acids, bases, reductive or oxidant agents were available in the laboratory of the research group. Therefore, apart from the first two departure reagents to form the scaffold – 2-amino-4-fluorobenzoic acid and formamide – the commercial availability reagents needed to derivatize the different substitution points had to be examined. This was checked programmatically using ChemSpiPy wrapper to access and query ChemSpider¹⁶³ database of commercial vendors using Python. In particular, this was effectuated by filtering the usual vendors of the group (Sigma-Aldrich, Fluorochem, Abcr, Alfa Aesar, BLDpharm and TCI).

Taking a look to the above-mentioned Markush structure, three positions (in C4, C6 and C7) have to be substituted (Figure 4.17):

- a) Derivatization of C4 of general structure **20** depends on the nucleophilic substitution of 58 possible anilines (**21**) which are the result of all the possible combinations of the hydrogen or halides of R1 and R2 in the different five ring positions. 57 of them are commercially available.

- b) Derivatization of C6 depends on two group of structures. The first compound with general structure **22**, is used to introduce the α,β -unsaturated ketone (with *E* conformation) terminated with a bromide to ensure the final substitution with compounds **23**. Three possible carboxylic acids depending on the number of methylene unities (from one to three) would need to be purchased. Only 4-bromocrotonic acid ($n=1$) is commercially available by BLDpharm at a reasonable prize (88€/10g). Nor the acid form neither its ester analogs for $n=2,3$ were found in the suppliers' websites. Finally, only ten out of the twelve possible heterocycles possible as compound **23** were found in the mentioned purchase databases. Therefore, 10 out of the 19 possible structures contemplated for C6 substitution are commercially available.
- c) To derivatize C7 position, two thiols (**24**), twelve alcohols (**25**) and methylamine (**26**) would be necessary. Among these set of 15 reagents, 10 can be bought through the above-mentioned vendors websites.

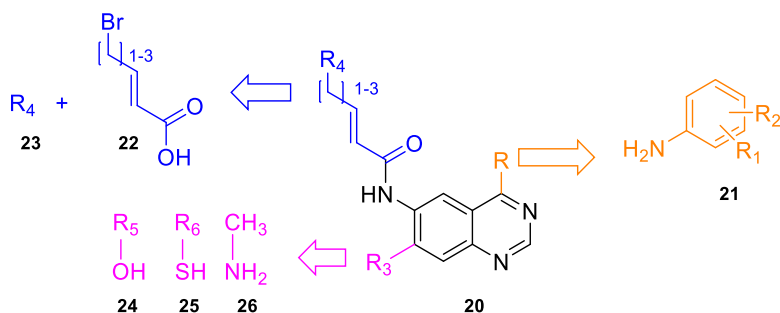


Figure 4.17 Simplified scheme for the need of the different reagent comprehension for the three derivatization points of the Markush structure.

The subset of molecules that could be synthesized attending to the restrictions of commercial availability would be uniquely formed by 5,700 analogs ($57 \times 10 \times 10$).

Therefore, the previous filtered dataset of 4,895 quinazoline analogs (after Lipinski's rule of five filtration) was finally lowered to 3,373 compounds considering this commercial limitation.

In conclusion, the final accessible space defined by the 16,530 analogs of the combinatorial database of the 51 Dacomitinib reported derivatives was finally lowered to 3,737 analogs. However, it is necessary to emphasize that this represents even a portion of the subspace generated by the derivatives found in the claim 5 of the drug's patents as the vastness of the real chemical space by the Markush structure would be computationally unattainable and unrealistic. An overview of the final filtered subset can be depicted in Figure 4.18.

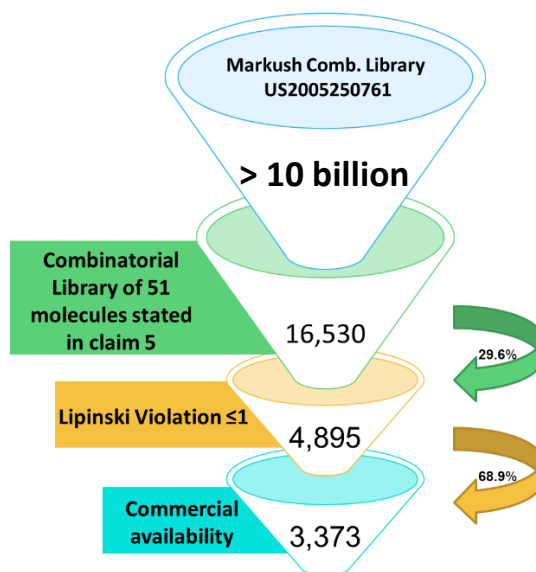


Figure 4.18 Filtration review departing from the original combinatorial database of Dacomitinib's Markush structure until the final accessible space .

4.3.2. Rational selection of Dacomitinib's analogs

Although final accessible space only contains the 20.4% of the original data (3,373 compounds out of the overall combinatorial set of 16,530), the representativeness of this filtered space in the chemical space of study is surprisingly large. Indeed, this subset of accessible molecules contains derivatives representing 8 out of the 10 clusters formed by KMN (leaving clusters 0 and 9 unrepresented). Thus, the maximum representation achievable by a rational selection could explain the 80.0% SC and 87.4% PC.

In contrast with the BD dataset, at the light of the t-SNE representation shown in Figure 4.19, this accessible space might seem poorly representative in order to perform a rational selection. However, the BD set of 51 molecules does also represent 8 out of the 10 clusters of the chemical space. Specifically, cluster 0 and 8 are not represented by the BD selection. Although the accessible space cannot achieve to represent neither the cluster 0, the 8th cluster present synthetically feasible analogs. Therefore, with the selection of 8 accessible analogs, it would not be increased the knowledge of this chemical space in terms of SC but, at least, a new region would be explored.

The rational selection of 8 representative compounds (later shown in Figure 3.17.) was performed, being these the nearest to each cluster centroid accessible compound. The visual comparison of the rational selected subset with BD and the accessible chemical space is shown in Figure 4.19.

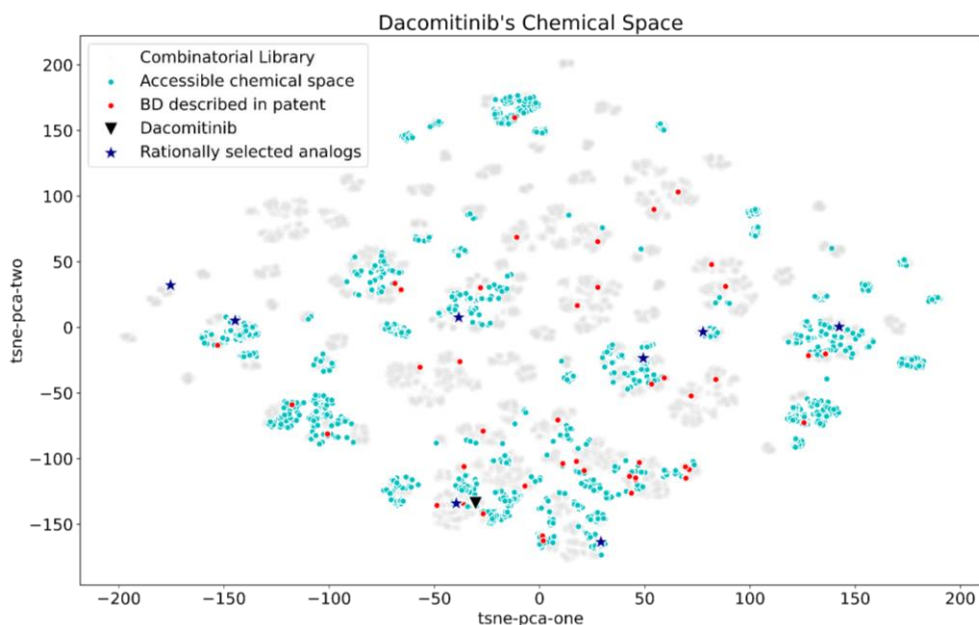


Figure 4.19 t-SNE plots representing Dacomitinib's chemical space (in grey), the accessible space defined by the filtered 3,373 analogs (in blue), the BD (in red), the original hit (as a black triangle) and the rationally selected analogs (dark blue stars).

All things considered, as shown in Table 4.2, in this case of $k=10$ clusters, neither the coverage results of a random nor rational selection of 10 candidates would improve the SC and PC results obtained by BD. This is a sign proving that we are focusing on a chemical subspace that appears to be highly explored as expected for its combinatorial nature which is BD-dependent. However, with this methodology, if having been previously implemented, the same chemical space would have been explored with the synthesis of much fewer candidates (only 8 in comparison with the 51 described) saving time, costs and resources to the original developers. Once again, this case is a new evidence of the main advantages of the suggested methodology of this thesis.

Table 4.2 Comparison of different coverage results of the chemical space obtained for BD selection, random selection and rational selection in $k=10$ clusters

Selection	Subset size	SC	PC
Bibliographical Data	51	80.0 %	92.6%
Random	10	57.8%	82.5%
Rational Selection of accessible space	8	80.0 %	87.4%

Hitherto, as not much public data is available concerning the biological activity of the BD analogs, its space description is still to be defined. In fact, only the inhibitory activity of seven molecules (analogs with library ID 8255, 8351, 8399, 8423, 8591, 9095, 16026) representing merely three clusters (1, 3 and 7) is specified in the proper patent describing poorly the 30% SC and 49.8 % PC. Thus, apart from describing all the accessible chemical space through the rationally selected samples, the aim of this study is focused in describing the inhibitory behavior of these analogs against EGFR, some of its main mutations, HER2 and HER4 and give insights for further lead optimization purposes.

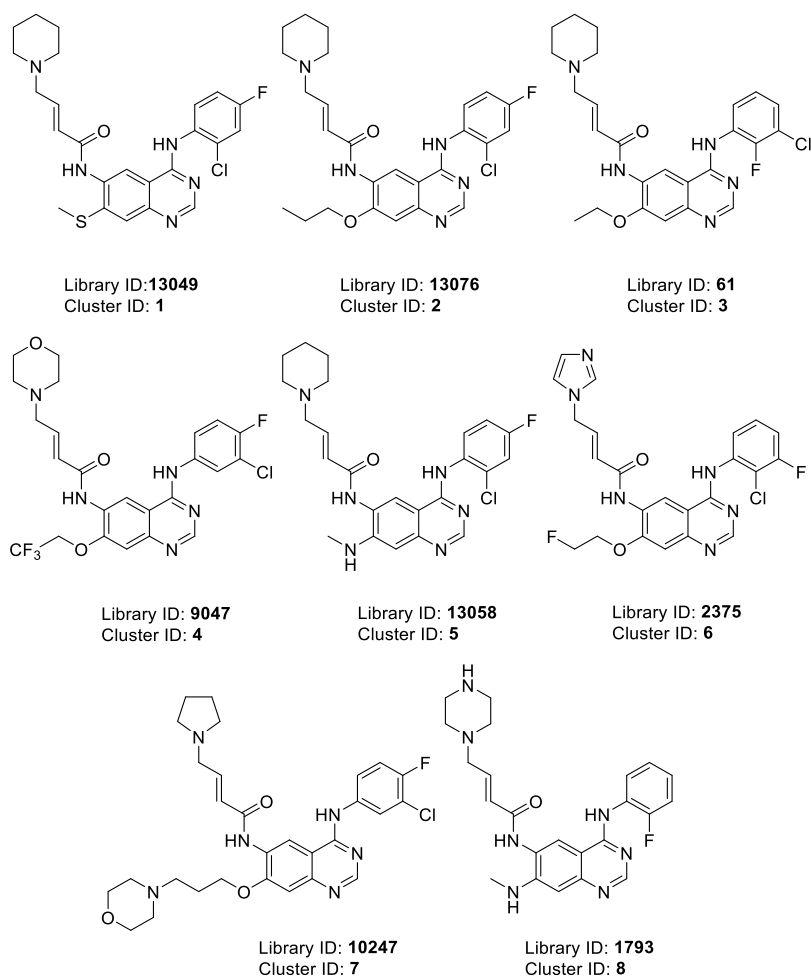


Figure 4.20. Rationally selected candidates to represent the 80% SC and 87% PC of the chemical space of Dacomitinib analogs

4.4. Synthesis of quinazolines

4.4.1. General synthetic Route

The general procedure for Dacomitinib derivatives obtention described in the drug's patent involves 8 steps of synthesis. However, after the study and synthesis of the rationally selected analogs, the general synthesis of this quinazoline derivatives has been achieved by alternative methodologies involving only 6 steps as it will later be disclosed. The overall synthetic route used for the synthesis of the rationally selected analogs (considering also the two additional steps described by the patent) schematically follows the derivatization pathway showed in Figure 3.18.

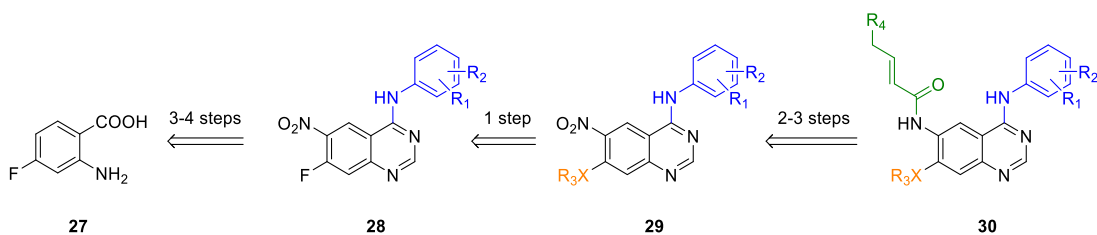


Figure 4.21 Schematic synthetic route considering the steps to obtain the intermediates with the derivatization points of the scaffold of study

The simplified route shows that there can be 3 or 4 initial shared synthetic steps until the obtention of the first derivatization point in C4. Afterwards, through a single step, the derivatives with substituted C7 positions are synthesized and, finally, the final Dacomitinib derivatives can be obtained through 2 to 3 synthetic steps depending on the complexity of the product isolation as explained below in the following sections.

Finally, the general 6-step synthetic route followed for the obtention of the selected molecules is shown in Figure 3.19. The conditions of the fourth step are not specified as they depend on the nature of the compound R₃X as it can be an alcohol (**24**), sodium methanethiolate (**25**) or methylamine (**26**) to afford the desired substitution with the fluorine in C7. This route will be deeply described below in the following sections.

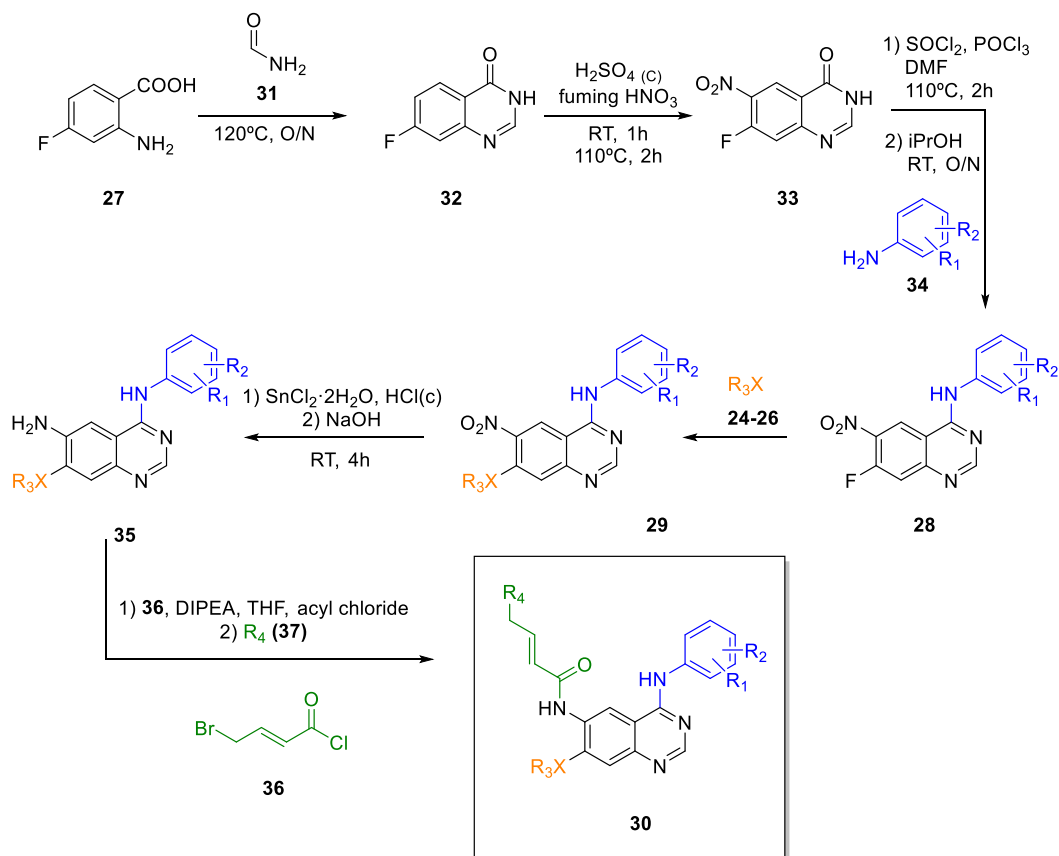


Figure 4.22 General synthetic route designed for the quinazoline derivatives obtention

4.4.2. Preparation of 7-fluoro-6-nitroquinazolin-4(3H)-one (33)

The 7-fluoro-6-nitroquinazolin-4(3H)-one intermediate (**33**) is the last shared intermediate after first derivatization of the quinazoline scaffold. This compound can be obtained through two steps of synthesis departing from the commercially available 2-amino-4-fluorobenzoic acid (**27**, 128€/100g in Fluorochem).

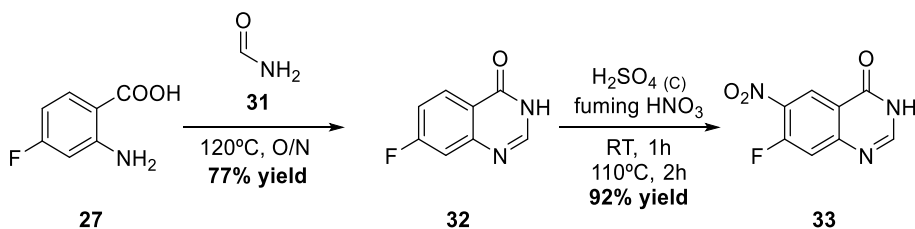


Figure 4.23. First two steps of the synthetic route to obtain quinazoline analogs

The first step synthesis consists in the classical Niementowski quinazolinone synthesis which affords 3*H*-quinazolin-4-ones by the condensation of an anthranilic acid with an amide. This synthetic step to obtain compound **32** was performed adapting the procedure described by Huan *et al.*²²³, which reacts 1 eq. of compound **27** with an excess of formamide (10 eq.) serving the latter as solvent. The mixture, after left reacting overnight at 120 °C is then poured into ice-cold water to allow the precipitation of compound **32** which is isolated from the excess of formamide due to its different solubility in water. The original protocol establishes that 3h of reaction are enough for total conversion but a previous follow-up of the reaction with TLC revealed that more than 5h are needed in order to obtain the pure product **32** without traces of **27** which would co-precipitate with the desired product during the work-up.

The obtention of product **32** can be easily proved by its ¹H-NMR spectrum (Figure 4.24) which evidences the ring formation due to the characteristic singlet of C2 and the NH in present in downfield along with the three proton signals corresponding to C5, C6 and C8 which multiplicity are altered (being *dd*, *td*, and *dd* respectively) due to their coupling with the fluorine isotope (¹⁹F).

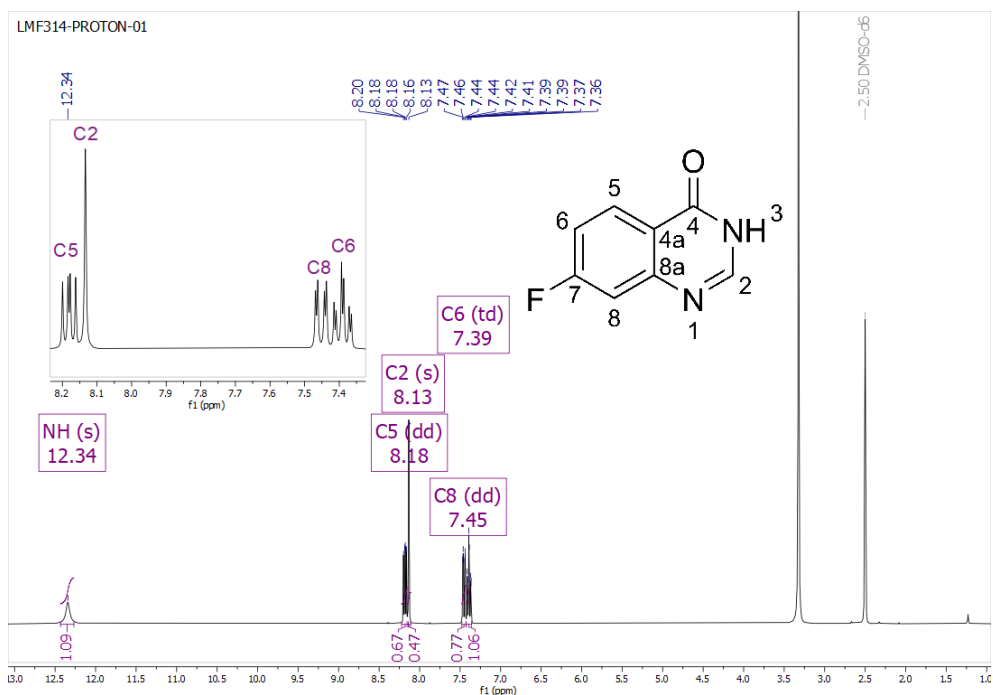


Figure 4.24 ¹H-NMR spectrum of intermediate **32**.

Secondly, the C6 position of the quinazolinone undergoes a nitration through an electrophilic aromatic substitution (*S*_EAr) in order to later permit the derivatization in this position. This nitration was performed following the procedure described by Hansel *et al.*²²⁴ which describes the obtention

of compound **33** through the reaction of 1 eq. of the intermediate **32** in a mixture of concentrated sulfuric acid and fuming nitric acid at high temperatures. This acidic mixture undergoes the formation of the nitronium ion which acts as an electrophile nitrating which replaces the hydrogen in C6 position.

At first, the execution of this reaction did improperly undergo yielding the departing material (**32**). This occurred after using 65% nitric acid as the presence of water in the reaction system impedes the proper reaction evolution to nitronium ion formation (product) as shown in Figure 4.25. This shifted equilibrium can be explained by the Le Châtelier's principle.

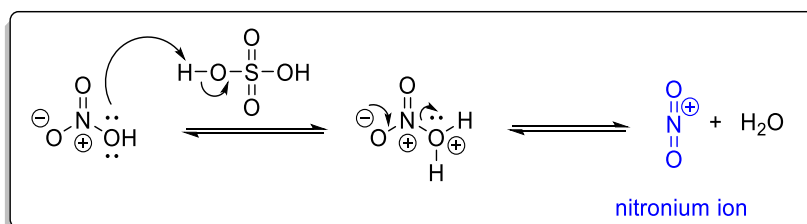


Figure 4.25 Reaction mechanism for nitronium ion formation.

Therefore, the use of fuming nitric acid (>90%) is a must for this synthetic step. This can be purchased although it is an expensive reagent (e.g. 466€/1L in Sigma-Aldrich). Thus, it was freshly prepared in the laboratory before the abovementioned reaction through its isolation via distillation while reacting sodium nitrite (1.0 eq) with concentrated sulfuric acid (1.3 eq.) at 100 °C.

The nitration placement of intermediate **33** is confirmed by the disappearance of the proton in C6 in its ¹H-NMR spectrum with the subsequent simplification of the C2, C5 and C8 system shown by two doublets (C5 and C8 coupling with ¹⁹F) and one singlet (the isolated C2). However, the most revealing evidence of the nitro addition is seen in the IR spectrum which evidences the appearance of the two bands corresponding to the characteristic asymmetric (1572 cm⁻¹) and symmetric (1326 cm⁻¹) N-O stretches (Figure 4.26).

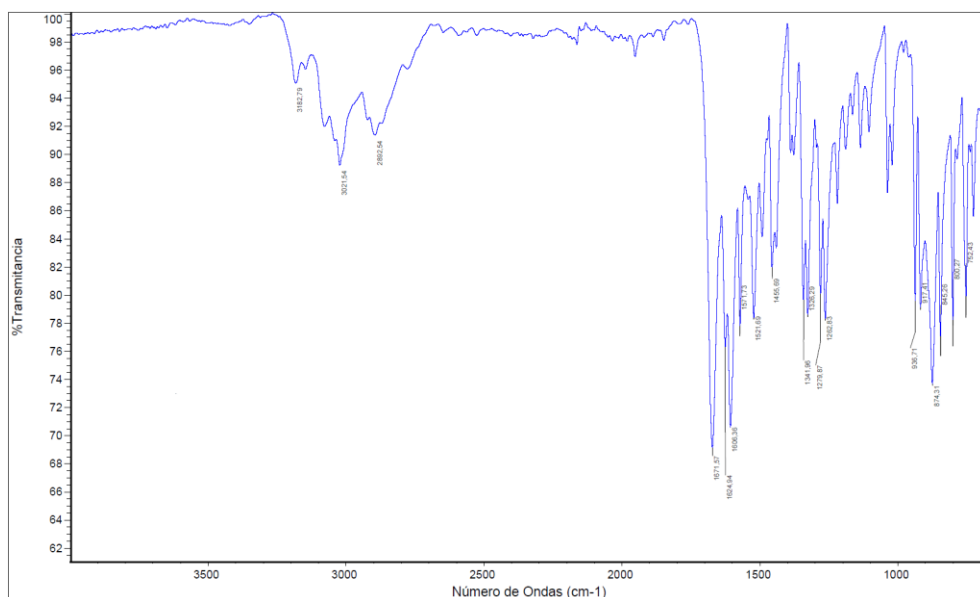


Figure 4.26 IR spectrum of intermediate 33. The bands at 1572 and 1326 cm^{-1} correspond to the N-O asymmetric and symmetric stretching demonstrating the successful nitration of the 32 intermediate.

When this synthetic route was designed and performed (in September 2021) neither the intermediates **32** nor the compound **33** were found at a reasonable price in the vendors websites. However, at the end of this thesis a new product **33** has been found at the affordable cost of 67€/100g in BLDpharm (updated in July 2022). Thus, if considering repeating the synthesis of the rationally selected analogs or other derivatives, it would be recommended to depart directly from the commercially available reagent **33** and then reduce the synthesis to a final 4-steps route.

4.4.3. Preparation of 7-fluoro-N-phenyl-6-nitroquinazolin-4-amines (**28**)

The following step to derivatize the C4 position of the quinazoline involves two reactions: a first C4 chlorination and a further nucleophilic substitution ($\text{S}_{\text{N}}2$) of the halide by different anilines (**34**).

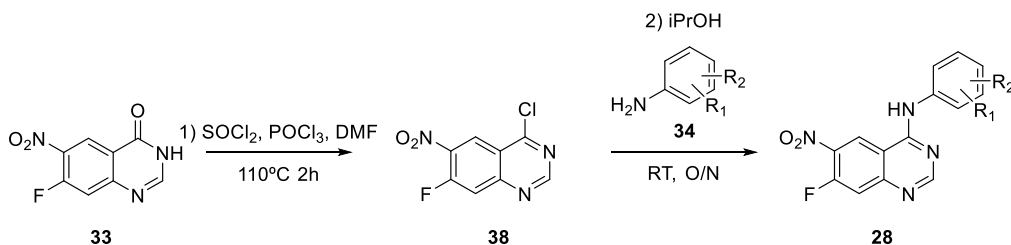



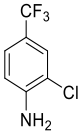
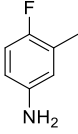
Figure 4.27 . Third step (involving two reactions) of the synthetic route to obtain quinazoline analogs

Following the procedure described by Hansen *et al.*²²⁴, the quinazolinone (**33**) was treated with two chlorinating agents, thionyl chloride (24.5 eq.) and phosphoryl chloride (3.44 eq.), along with dimethylformamide (DMF, catalytic amounts) resulting in the desired 4-chloroquinazoline (**38**). This reaction involves conventional heat and the further removal *under vacuo* of the reagents in excess. As an improvement in the synthetic route (an already seen strategy with other procedures and conditions described by Zhao *et al.*²²⁵), the chlorinated intermediate (**38**) is used further on without purification to proceed with the derivatization of C4 position. The described yield for the first reaction is 74%.

Thus, the **38** intermediate undergoes directly a nucleophilic substitution driven by its reaction with different anilines (**34**) at room temperature using isopropanol (*i*PrOH) as solvent to afford the precipitation of the desired **28** derivative. The yields obtained for this step involving the two reactions described ranges from 61 to 99% depending on the aniline used as substrate. Seven intermediate **28** analogs were synthesized during this study. In contrast with the introduced rational selection, compounds **28{6}** and **28{7}** were synthesized in the search of the obtention of two compounds not included in the library which will be discussed below. Moreover, it was previously taken consideration during the rational selection process the cost-effectiveness of the anilines involved (purchased in Fluorochem¹⁶⁹) as evidenced in Table 3.7.

Table 4.3. 7-fluoro-*N*-phenyl-6-nitroquinazolin-4-amines derivatives (**28**) prepared in this project with its corresponding yields and the commercial price of the involving anilines (**34**) for their synthesis.

Aniline structure	Name	Commercial substrate price	Yield
	34{1}	16{1} : 10€/1g	97%
	34{2}	16{2} : 10€/1g	99%
	34{3}	16{3} : 21€/100g	77%
	34{4}	16{4} : 10€/1g	91%

Aniline structure	Name	Commercial substrate price	Yield
	34{5}	16{5} : 38£/100g	96%
	34{6}	16{6} : 14£/25g	61%
	34{7}	16{7} : 10£/1g	82%

At the light of the anilines shown in the table above, it can be easily seen that all the derivatives contain at least one additional fluorine atom (a trifluoromethyl group in the case of **34{6}**) which resulted in the obtention of the **28** derivatives with complex NMR spectrum due to their ^1H - ^{19}F and ^{13}C - ^{19}F couplings. The ^1H -NMR (Figure 4.28) and ^{13}C -NMR (Figure 4.29) spectra of compound **28{5}** may serve as an explanatory example. At the light of the ^1H -NMR spectrum, the successful substitution of the aniline is evidenced by the appearance of a broad labile signal in 11.82 ppm corresponding to the secondary amine linking the aryl group (7.55-7.33 ppm) to the quinazoline structure (evidenced by the three main protons with multiplicity *d*, *s*, *d*, respectively). However, the multiplicity of the protons of the aniline ring shows an altered multiplicity of its protons due to the presence of the fluorine atom in C11.

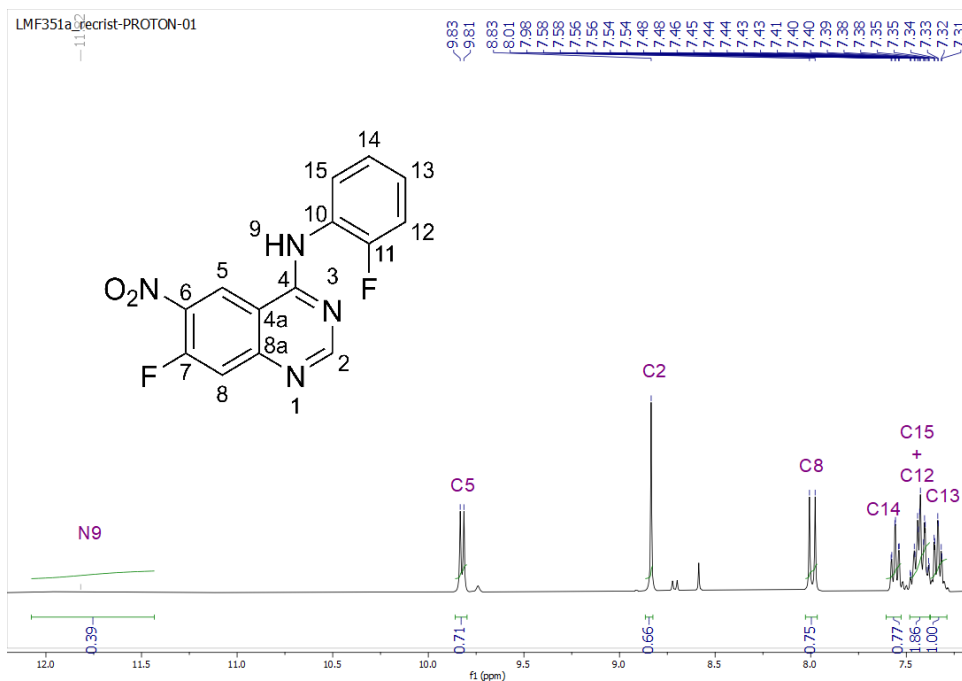


Figure 4.28 Zoom in the aromatic region of the $^1\text{H-NMR}$ spectrum of intermediate **28(5)**.

The proper identification of these latter protons can be easily elucidated with the HSQC spectrum. Depending on the coupling constant ($J_{\text{C-F}}$) of each carbon signal, each location relative to the fluorine atom can be predicted (the nearest to the halide, the greatest the $J_{\text{C-F}}$ value). Furthermore, the difference between C7 and C11 assignment, which are both present at downfield, is also proved by the $J_{\text{C-F}}$ and the HMBC spectrum. The fluorine atom in the quinazoline typically presents a $J_{\text{C-F}} \approx 267\text{Hz}$ and fluorine atoms present in phenyl rings shows $J_{\text{C-F}} \approx 245\text{Hz}$.

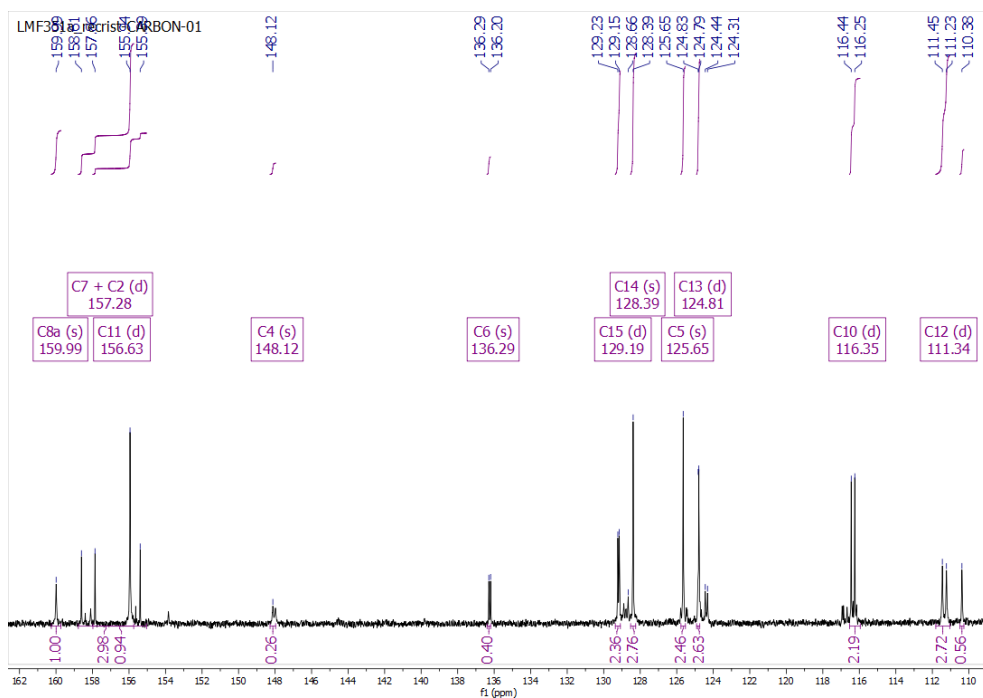


Figure 4.29 Zoom of the region of ^{13}C -NMR spectrum with the intermediate **28{5}** signals.

Additionally, the structure composition of compound **28{5}** composition was confirmed by HRMS.

4.4.4. Preparation of 7-substituted-*N*-phenyl-6-nitroquinazolin-4-amines (**29**)

The second substituted position is C7 considering the synthetic route. In order to derivatize this point with the stated substructures, three possible different nucleophilic substitutions of the fluorine can take place depending on the nature of the reagent leading to the attachment of a methylthiol group **29{1}**, seven alkoxy chains **29{2,3,4,9,6,7,10}** or methylamine, **29{5,8}**.

In order to obtain the thiolate intermediate **29{1}** an adaptation of the procedure described by Vasu *et al.*²²⁶ was carried out (Figure 4.30). In this case, the 7-fluoro quinazoline derivative (**28{1}**, 1.00 eq.) was treated with sodium methanethiolate (**25**, 4.00 eq.) in DMF at 100 °C for 4 hours to afford, after product isolation, the desired thiolate intermediate **29{1}** with excellent yields.

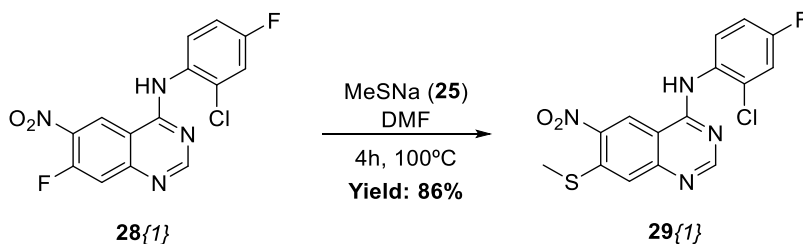


Figure 4.30 C7 derivatization with methanethiolate.

Furthermore, to introduce different alkoxy groups in C7, two approaches were performed depending on the nature of the alcohol reagent (**25**). Both of them involved the transformation of the diverse alcohols to its alkoxide forms with the use of different bases depending on the reaction conditions.

When the alcohol can be used both as reagent and solvent (method A in Figure 4.31), an excess of it was used along with 10.0 eq. of sodium hydroxide until **28** was totally dissolved and was refluxed at the alcohol's boiling point temperature for 30 minutes. The product isolation involved its precipitation in a sodium bicarbonate saturated solution, filtration and solvent removal under pressure. When using 1-propanol (**25{1}**) or ethanol (**25{2}**), the obtained products did not need further purification steps. However, after using 2,2,2-trifluoroethanol (**25{3}**) with this methodology, the obtention of the intermediate **29{4}** required its purification by column chromatography due to the formation of different side-products or product decomposition evidencing the need of a different methodology.

Thus, alternatively, the procedure described by Shao *et al.*²²⁷ was used when the alcohol reagent consisted in a substituted and limited alcohol (not commonly used as solvent). In this case (method B in Figure 4.31), dimethyl sulfoxide (DMSO) was used as the solvent where 1.0 eq. of the starting product **28** reacted with 1.5 eq. of the alcohol **25{4,5}** in basic conditions favored by the presence of 3.0 eq. of potassium tert-butoxide. The complete product transformation took place after 30 minutes of reaction at 25 °C and this was isolated after its precipitation by the reaction quenching in water, filtration, washing and its dryness in vacuo.

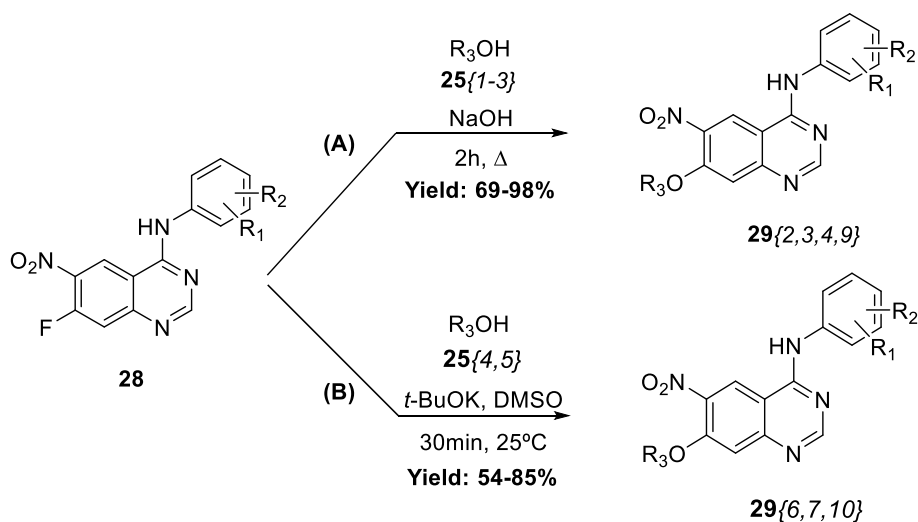


Figure 4.31 Two methods used for the C7 derivatization with different alcohols

Finally, for the obtention of the **29{5}** and **29{8}** intermediates, the corresponding **28** substrates were directly reacted with an excess of methylamine which serves both as base and nucleophilic agent. The same procedure as method A for the fluorine substitution with different alcohols was followed. In this case, isopropanol (*i*PrOH) was used as solvent and no base was used rather than the same reagent **28**. The reflux temperature was settled to 83 °C (isopropanol's boiling point).

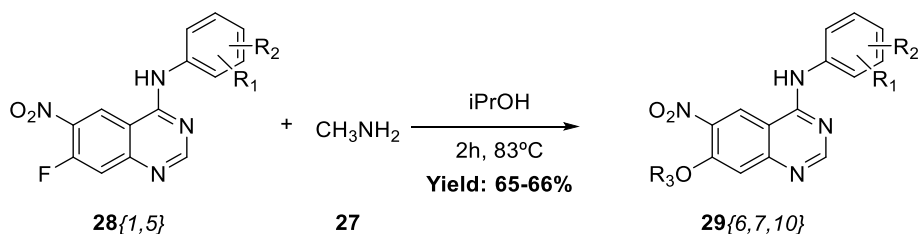


Figure 4.32 C7 derivatization with methylamine

An example of this different derivatization of C7 can be shown by the $^1\text{H-NMR}$ spectra differentiation of the intermediates **29{1}**, **29{2}** and **29{5}**. These compounds share the same substrate **28{2}** so its main differentiation can be evidenced by the signals of the newly chain attached to C7 position as highlighted by different colors in Figure 4.33. Furthermore, the disappearance of the downfield doublet signal with $J_{\text{C-F}} \approx 267\text{Hz}$ in the $^{13}\text{C-NMR}$ spectra of all the analogs confirms the disappearance of the fluorine atom in C7. In all cases, the structures were confirmed by mono and bidimensional NMR spectroscopy observing the expected signals and IR spectroscopy was used to

identify the presence or absence of the main groups. The compounds compositions were confirmed by HRMS.

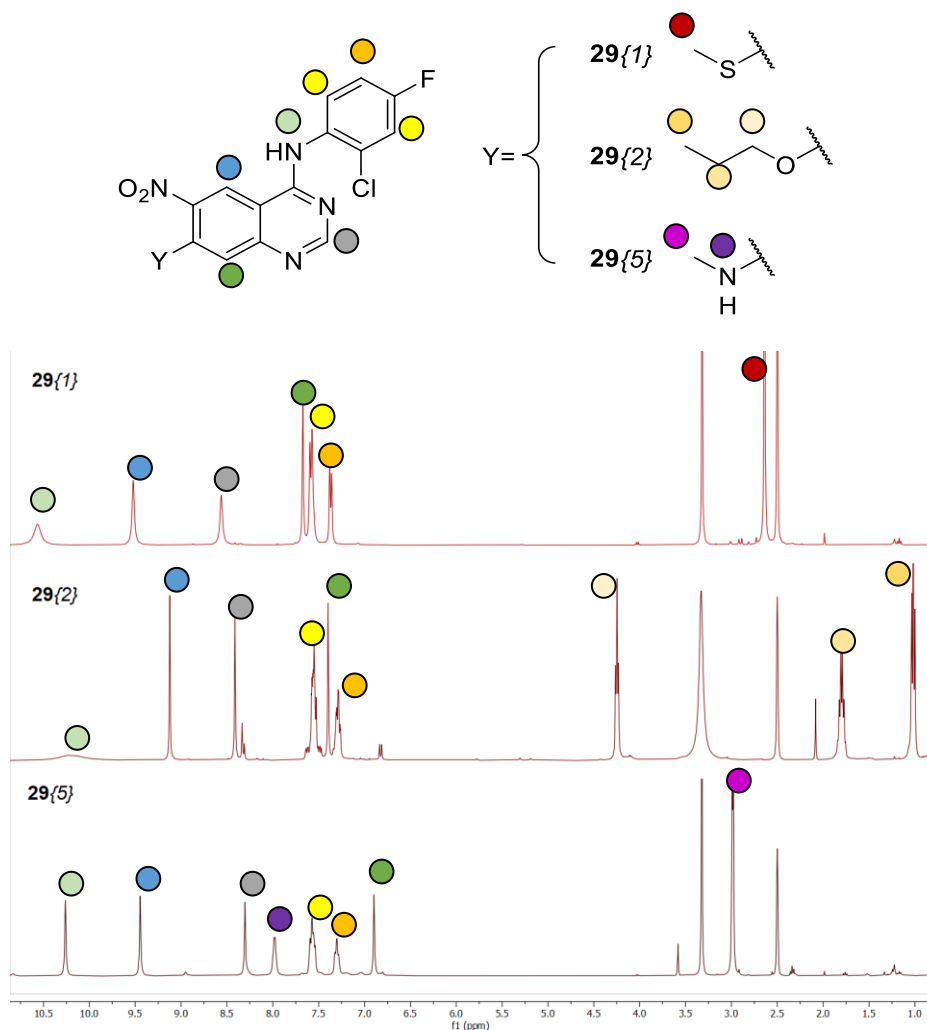


Figure 4.33 ¹H-NMR spectrum of compounds 29{1}, 29{2} and 29{5} sharing the same scaffold and aniline attached to C4.

4.4.5. Preparation of dacomitinib analogs (30)

The last derivatization of the quinazoline scaffold relies on the C6 position of the aromatic ring. In order to achieve the final obtention of the dacomitinib analogs two or three steps can be considered depending on the methodology used for the final introduction of the α,β -unsaturated chain presenting a heterocyclic fragment in its end.

Firstly, the reduction of the nitro group present in **29** to a primary amine is needed in order to permit the next formation of the amide bond desired. To reduce this nitro and obtain the diverse 7-substituted-*N*⁴-phenyl-6-nitroquinazolin-4,6-diamines, the classic and highly effective reduction with tin (II) chloride dihydrate was performed. The nitroarenes derivatives (**29**) were treated with 5.0 equivalents of tin (II) chloride dihydrate as reducing agent in highly acidic conditions favored by the presence of concentrated hydrochloric acid. The reaction was quenched with a 2M sodium hydroxide solution to remove the excess of the tin (II) chloride. The desired diamine analogs (**35**) were obtained with excellent yields ranging from 80% to 99%.

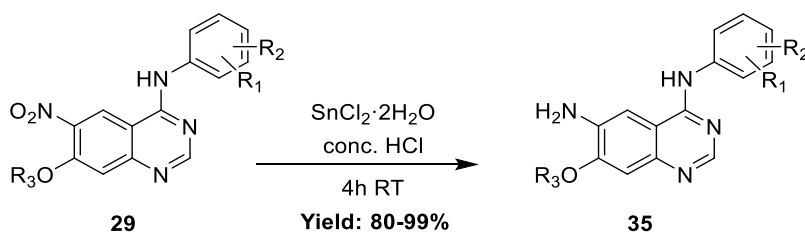


Figure 4.34 Fifth step of the synthetic route to obtain quinazoline analogs

The effectiveness of this synthetic step is easily evidenced by the examination of its IR spectrum of the molecule which shows the characteristic bands of the newly formed primary amine. To name an example, Figure 4.35 shows the intermediate **35**{3} IR spectrum which shows the presence of the following distinctive bands: the asymmetric and symmetric N-H stretches at 3250 and 3154 cm^{-1} respectively (A, B) properly attributed to primary amines, the scissoring N-H bending at 1509 cm^{-1} (C), the C-N stretching at 1261 cm^{-1} (D) and the wagging N-H bending in 756 cm^{-1} . In contrast with the bands shown for its predecessor **29**{3}, the typical asymmetric and symmetric stretches of the nitro group have disappeared (1571 and 1326 cm^{-1}).

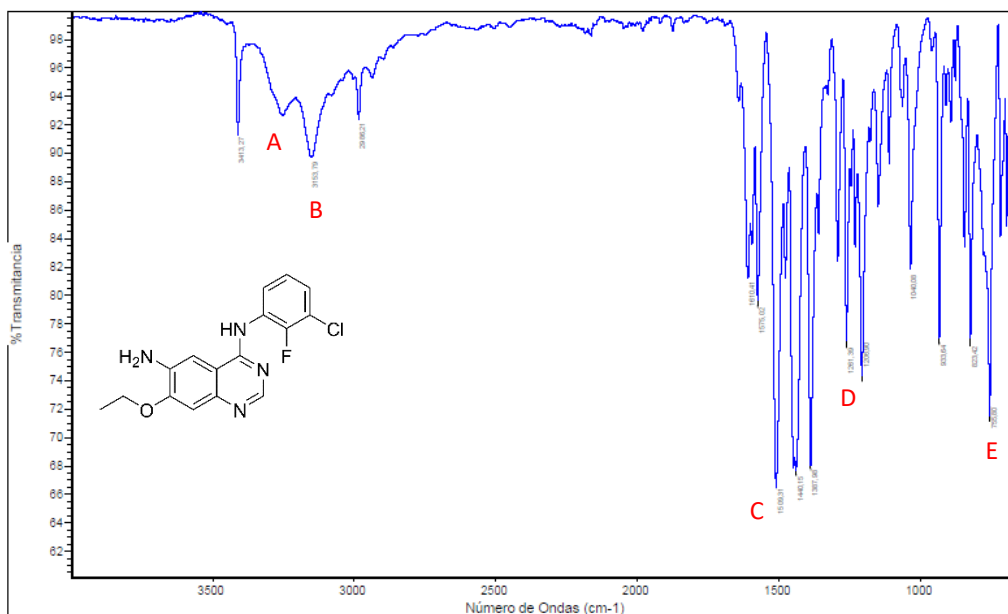


Figure 4.35 IR spectrum of compound **35(3)** with the highlighted bands at 3250 (A), 3154 (B), 1509 (C), 1261 (D) and 756 cm^{-1} (E).

Last of all, the final obtention of the Dacomitinib analogs was achieved through two procedures involving one or even an additional step as shown in Figure 4.36)

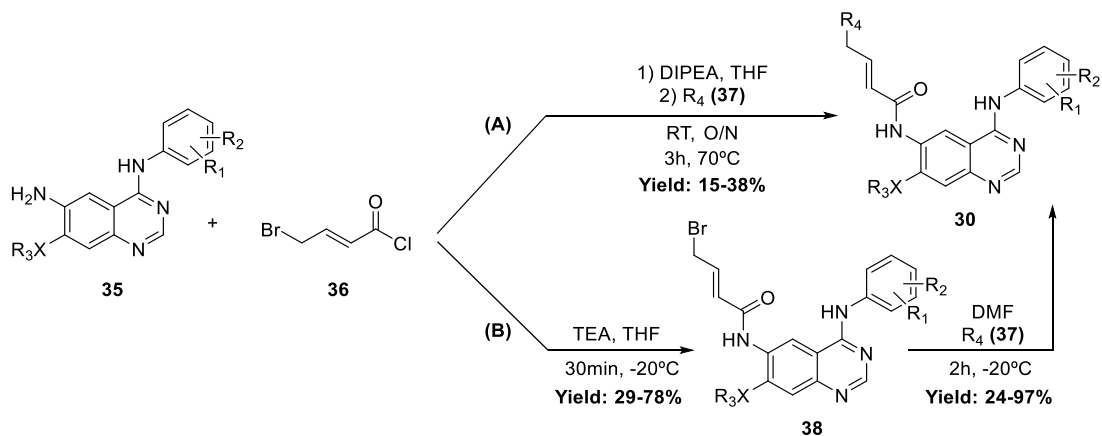


Figure 4.36 Two methods applied for the final obtention of Dacomitinib analogs. Method A is adapted from the drug's patent protocol²²² and method B from Smail *et al.*²⁰⁸

Firstly, adapting the methodology stated in the drug's patent²²² (method A in Figure 4.36) acyl chloride (**36**) was freshly prepared to react with the desired **35** intermediate. This was set to react the previous day in an overnight reaction consisting in the 4-bromocrotonic acid substrate (2.0 eq.) with oxalyl chloride (2.2 eq.) in dichloromethane. Later, 1.0 eq. of the corresponding **35** intermediate

was reacted with the acyl chloride (**36**) and a base media favored by *N,N*-diisopropylethylamine (DIPEA) in tetrahydrofuran (THF) during 2 hours at room temperature to allow the amide bond formation through nucleophilic substitution. Later, after the attachment of the **36** to **35** was expected to have been successfully completed, 2.2 eq. the desired heterocycle (**37**) were added and the reaction was let to stir at room temperature overnight. The following day, an additional amount of 5.7 eq. of compound **37** were added and the mixture was heated to 70 °C for 3 hours. The cited procedure suggests the reaction quenching with water. However, this work-up step did show inconveniences involving the formation of thick slurries which were difficult to redissolve in the subsequent extractions affecting the final reaction yields. Thus, the protocol was adapted skipping this isolation step and directly removing the solvent *under vacuo*. The crude was extracted with ethyl acetate, washed with water and brine and dried over anhydrous sodium sulphate yielding a solid which was finally purified automated flash chromatography (silica gel 0-20% DCM: MeOH) to yield the desired product **30**.

This first methodology was applied to obtain four Dacomitinib derivatives **30**{3,4,5,6,9} which presented poor yields ranging from 15% to 38%. Apart from wasting a large amount of **35** substrate with this reaction step, it involved a difficult separation of the product through a chromatographic column. As shown in Figure 4.37, the isolation of the product **39**{8} through this methodology (NotebookID: **LMF377**) involved a tedious process of almost 3 hours which separated the desired product (colored in yellow) with still some impurities. Thus, the combination of the poor yield measured values, the complexity in the product purification and the unsuccessful obtention of the analog **39**{8} through this methodology forced the search of an alternative approach to obtain the rest of compounds.

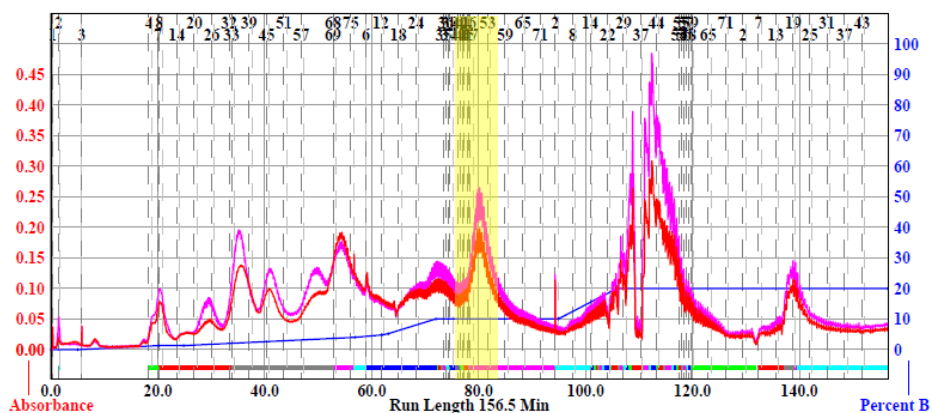


Figure 4.37 Automated flash chromatography chromatogram to isolate compound **30**{8} (fraction colored in yellow) through methodology A. Conditions: silica gel, DCM:MeOH.

Other methodologies were attempted such as the Sheehan and Hess approach²²⁸ for peptide synthesis, involving the use of coupling reagents such as 1-ethyl-3-carbodiimide (EDC) or dicyclohexylcarbodiimide (DCC) with 4-dimethylaminopyridine (DMAP) to directly react the *N*-Het-bromocrotonic acid structure with the **35** intermediates. However, the attempts were unsuccessful leading to complex crude mixtures where no product was able to be identified.

Finally, the rest of dacomitinib analogs were successfully obtained following the methodology described by Smaill *et al.*²⁰⁸ (method B in Figure 4.36) which involved one further step for the intermediate isolation of the compound **38**. In this case, 1.31 eq. of the acyl chloride (**36**) were freshly prepared by a 3-hour reaction at room temperature and directly dissolved in 3 ml anhydrous THF. The reaction of the desired **35** was driven in a basic media of 30 ml THF with 2.5 eq. of triethylamine (TEA) at -20 °C (in a salted ice-water bath). The solution containing compound **36**, was poured dropwise for 30 minutes and the reaction was let to stir at -20 °C for an additional 30 minutes. The crude work-up, similarly, to described in the method A, involved three extractions with ethyl acetate, a washing with brine, its dryness with anhydrous sodium sulphate and solvent removal *under vacuo*. Finally, intermediate **38** was obtained through a previous purification through automated flash chromatography with gentler conditions (silica gel 0-2% DCM: MeOH). As a comparative example with the chromatogram shown for the obtention of **30{8}**, the purification of the intermediate **38{8}** presented a cleaner column involving 90 minutes of separation with a slope from 0 to 4% of dichloromethane-methanol mixture used as mobile phase. Indeed, the time of purification could be improved and reduced at the light of the shape of mixture separation. However, no further attempts were performed for the optimization of the column conditions. Overall, through this additional step, the complexity of the product purification was lowered and measured yields ranged from 29% to 78%.

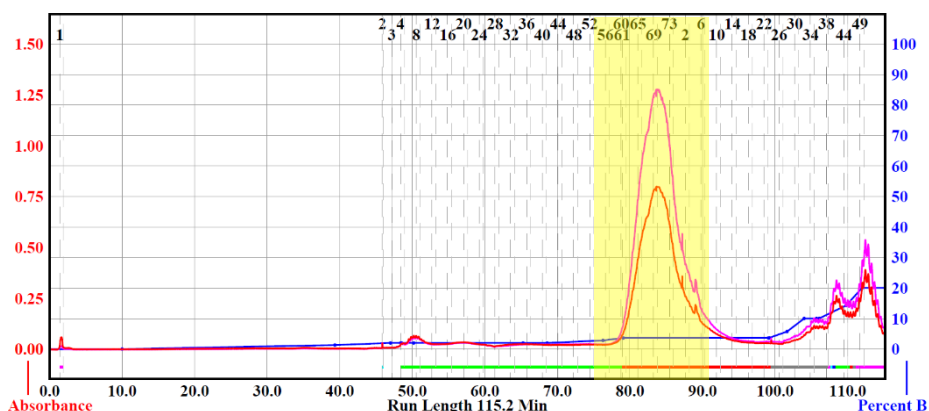


Figure 4.38 Automated flash chromatography chromatogram to isolate compound **38{8}** (fraction colored in yellow) through methodology A. Conditions: silica gel, DCM:MeOH.

Finally, the four resting dacomitinib analogs were obtained through a second nucleophilic substitution reacting 1.0 eq. of the isolated **38** intermediates with 6.8 eq. of the corresponding **37** heterocycle in 5 ml of DMF (the procedure stated by Smail *et al.* suggests the use of dimethylacetamide) at -20 °C for 2 hours, after the complete reaction was checked by TLC. This was finally quenched with water, poured into saturated NaHCO₃ solution, and extracted twice with ethyl acetate. The combined organic fractions were washed with brine, dried and the solvent was removed under reduced pressure to finally afford, without further purification steps, the desired product **30**. The yields measured for this final step ranged from 24% to 97%.

Consequently, by the use of this alternative methodology B, the general yield measured (as the product of the two steps yields) ranges from 11% to 55% slightly increasing the effectiveness of this in contrast with the methodology A which showed 15%-38% yields. Overall, the advantages of this additional procedure implicates the use of a neater synthesis with less time-consuming and easier work-up and purification steps.

The discussion of the NMR spectra of the compounds obtained by this additional methodology can evidence the structure modifications observed in each synthetical step. Next, an example of the final analog **30{7}** and its intermediate **38{7}** is given with their ¹H-NMR spectra (Figure 4.39) and ¹³C-NMR spectra (Figure 4.40).

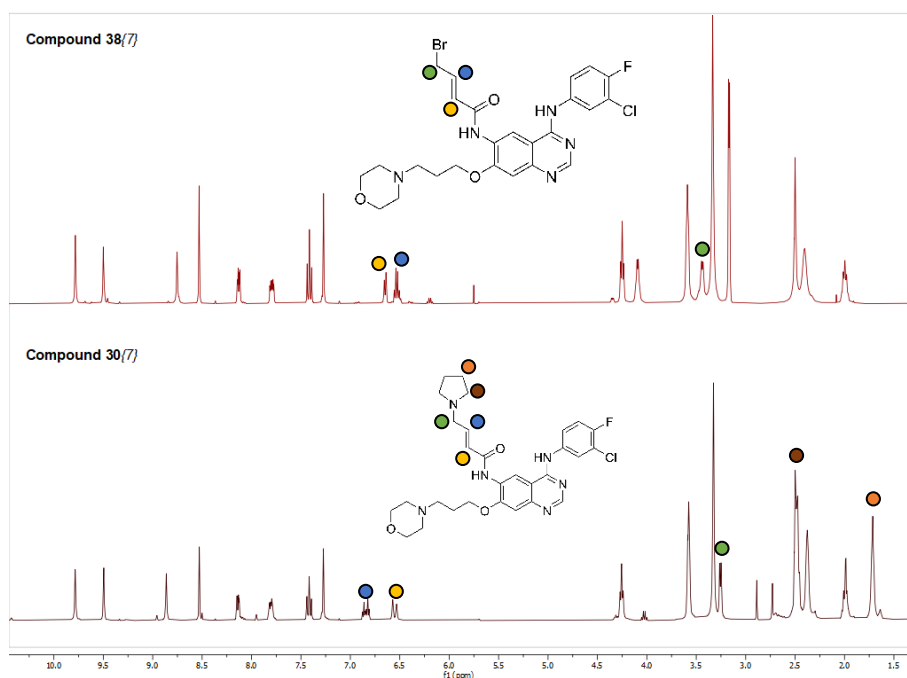


Figure 4.39 ¹H-NMR spectrum of the intermediate **38{7}** and the final product **30{7}**.

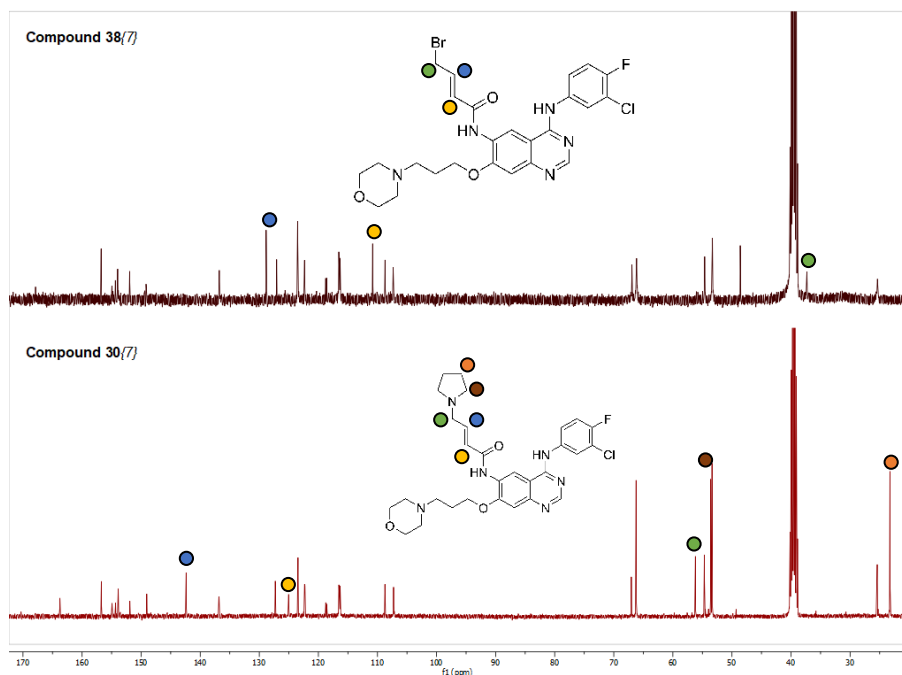


Figure 4.40 ¹³C-NMR spectrum of the intermediate **38(7)** and the final product **30(7)**.

Firstly, the expected amide bond formation was confirmed by the appearance of the two characteristic protons of the unsaturation: the α -H is assigned to the doublet in downfield (6.65 ppm) and the β -H to the signal at 6.53 ppm. Furthermore, the doublet in upfield (6.44 ppm) confirmed the attached methylene to the bromide. The further nucleophilic substitution of the bromide by the pyrrolidine was easily confirmed both by the shifted position of the α -H and β -H signals in the ¹H-NMR spectrum (6.55 ppm and 6.84 ppm respectively) and their deshielded locations (along with the methylene signal) of their corresponding carbons in the ¹³C-NMR spectrum (shifting from 128.8 ppm and 110.8 ppm to 142.4 ppm and 125.0 ppm respectively). The appearance of the two signals of the pyrrolidine group in both spectra confirmed the overall structure. Furthermore, the HMBC spectrum of product **30(7)** (Figure 4.41) showed a correlation between the proton signal of the methylene of the chain (in green) with the ¹³C the double bond (blue and yellow) and the first pyrrolidine methylenic group (in dark red). Plus, the constant coupling ($J \approx 15$ Hz) of both proton signals taking part in the double bond confirms the *E* configuration of the structure.

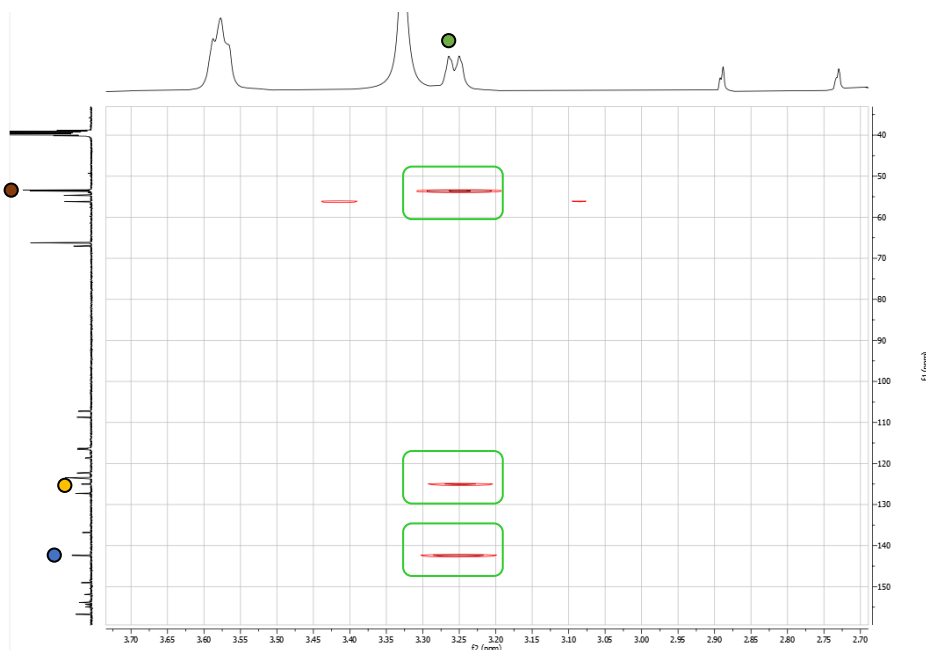


Figure 4.41 HMBC of compound **30{7}** demonstrating the bromide substitution by the desired heterocycle.

4.4.6. Global synthesis overview

The final eight Dacomitinib derivatives were finally synthesized following a 6-step synthesis for four of them (**30{3,4,5,6}**) and an additional analog outside of the Markush chemical library **30{9}**) and a 7-step synthesis for the rest of the rational selection (**30{1,2,7,8}**). The larger synthetic route consisting in seven steps offered a cleaner and slightly more effective alternative in terms of measured yields and time expended during product purification. The overall global yield measured ranges from 6% to 29% with a mean value of 11% (Table 4.4). The limitation steps are mainly present in the final derivatization of C6, which involves the attachment of a reactive α,β -unsaturated chain that affords the formation of undesired side products that forces the product purification via column chromatography.

Altogether, the measured total yields are considered acceptable for a multistep process as no optimization study has been implemented during the process of product obtention.

Finally, a total amount of 44 molecules have been synthesized (also considering the acyl chloride **36** used in the final step). Even more, among these, 35 previously undescribed molecules have been synthesized and fully characterized.

Dacomitinib: Study of substituted quinazoline derivatives as irreversible inhibitors of EGFR

Table 4.4. Overview of every single step yield and the overall yield obtained for the synthesis of the nine quinazolines synthesized and the **35** intermediate. In brackets is identified the intermediate obtained through the assigned synthetic step.

Product	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 6	Step 7	Overall
	yield % [32]	yield % [33]	yield % [28]	yield % [29]	yield % [35]	yield % Method A [30]	yield % Method B [38]	yield % Method B [30]	
30 {1}	77	92	97	86	92	-	44	24	6
30 {2}	77	92	97	80	95	-	78	71	29
30 {3}	77	92	99	98	95	27	-	-	18
30 {4}	77	92	77	69	94	16	-	-	6
30 {5}	77	92	97	65	99	15	-	-	7
30 {6}	77	92	92	54	80	38	-	-	11
30 {7}	77	92	77	85	82	-	29	97	11
30 {8}	77	92	96	66	95	-	53	26	6
30 {9}	77	92	61	88	97	31	-	-	11
35 {10}	77	92	82	80	91	-	-	-	42
Mean	77	92	88	77	92	25	51	55	11

4.5. Biological activity

Following the synthesis of this family of quinazoline compounds, it was next aimed to assess their biological activity in terms of tyrosine kinase inhibition for their further purpose to treat NSCLC. The eight rationally selected analogs **30{1-8}** and three extra compounds not considered in the Markush combinatorial library were tested (shown in Figure 4.42). The additional structures were added to assess the influence of the following structural modifications in the inhibitory activity of the scaffold (highlighted in the subsequent figures): the use of a different aniline in C4 (**30{9}**), the removal of the nitrogenated heterocycle in the C6 sidechain (**38{2}**) and both the use of a different aniline in C4 and the replacement of the sidechain containing the cleavage point for the covalent bond between ligand and receptor by a free amine (**35{10}**).

To study the inhibitory activity of this set of molecules, nine enzymes of interest – among all the diverse offered by the external company responsible of the *in vitro* testing – were chosen as target receptors attending to the described inhibitory behavior of the original hit (explained in Section 4.1):

- Wild-type EGFR, HER2 and HER4 (as an irreversible ERBB blocker)²¹⁴
- EGFR L858R²²¹ and EGFR T790M²¹⁹ the commonly known acquired mutations as a mechanism of resistance to first and second generation TKIs.
- The main Dacomitinib and Osemirtinib acquired resistance mutation C797S^{219,229,230}
- The G719S acquired mutation^{195–197,231}
- And the double and triple mutated kinases T90M/L858R and T790M/C797S/L858R respectively.

The *in vitro* study of this family of compounds (the rationally selected molecules and the set of structures with the abovementioned modifications) and the original hit (Dacomitinib) as tyrosine kinase inhibitors was performed in the proper isolated enzymes through the external company Reaction Biology Corp.²³². The Determination of the effect of the 12 compounds on the kinase activity of the 9 protein kinases *in vitro* was performed with a radiometric protein kinase assay (³³PanQinase® Activity Assay). The kinase inhibition profile was determined by measuring the residual activity values at two concentrations each in singlicate in the 9 protein kinase assays.

Results found in literature²¹⁴ of Dacomitinib *in vitro* kinase inhibitory assay of wild-type EGFR proof its IC₅₀ value is at 1.8nM. Furthermore, complementary studies developed by Zhang *et al.*²⁰⁹ involving quinazoline analogs similar to the Dacomitinib derivatives of study have shown IC₅₀ values

ranging from 2 nM to 6 nM. Consequently, the synthesized compounds were tested at the following final assay concentrations, in singlicate: 2 nM and 10 nM.

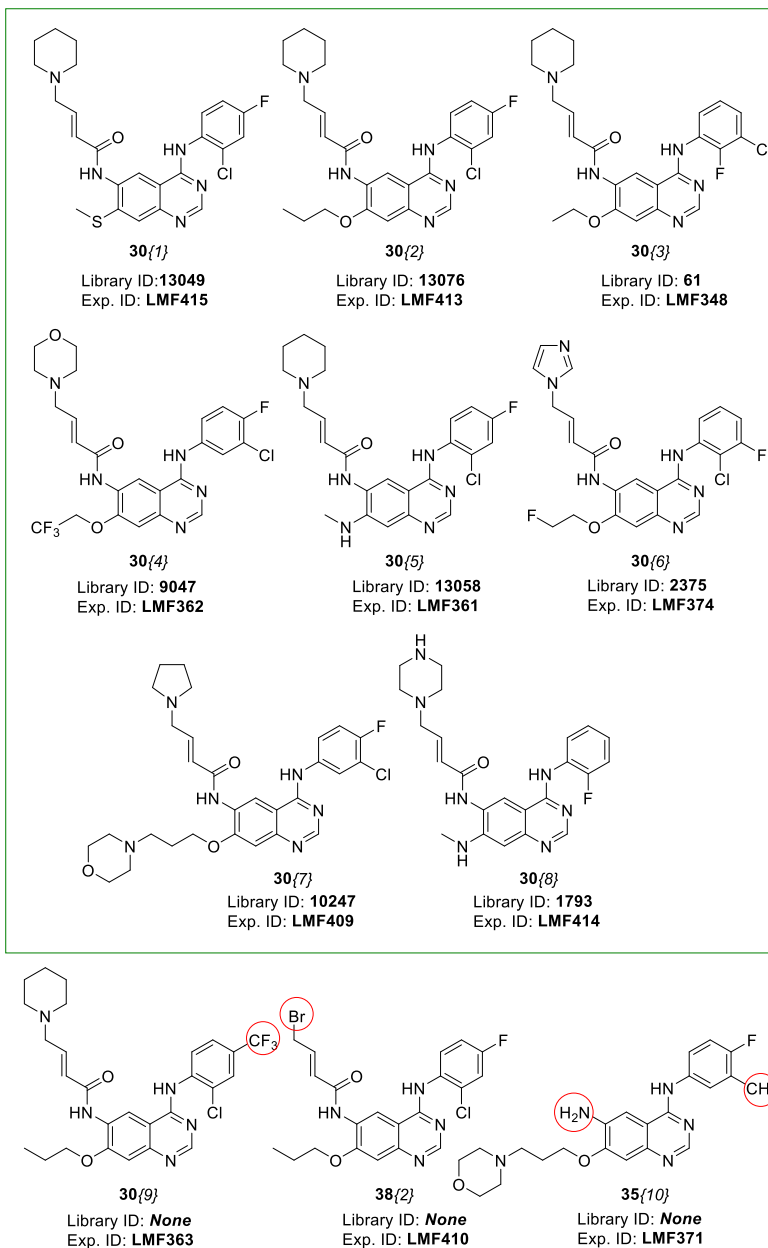


Figure 4.42. Library ID, experimental ID and structure of the molecules tested *in vitro*. Those molecules present in the green squares are analogs present in the combinatorial library derived from Dacomitinib's Markush structure (Dacomitinib rationally selected analogs present in the chemical space of study). In red circles, the structural modifications studied to assess the characteristics of the alternative molecules.

The general overview of the results obtained for the 14 molecules tested in the 9 kinases are summarized in Table 3.11. In general terms, these family of analogs, as expected, appear to be promising, since all of them have shown interesting inhibitory activity against wild-type EGFR at 10nM. A few assumptions can be formed by a first general glimpse of the results.

Firstly, the candidate **30{3}** has shown highly similar values to the original hit at both concentrations (2nM and 10nM) demonstrating that its IC₅₀ might fit with Dacomitinib's. Furthermore, analogous highly active results have been measured for the isolated G7196 and L858R EGFR mutations. In contrast, Dacomitinib have shown better inhibitory activity against HER2 and HER4.

The less active molecule is the compound **30{9}**, which presence of the trifluoromethyl group in the aniline moiety in C4 may have affected negatively to the general interaction of this ligand with the different receptors of study. Still, it has shown inhibitory activity at 10nM against wild-type EGFR, G719S and L858R.

Surprisingly, the compound **35{10}**, did show some inhibitory activity against 7 of the 9 kinases testes. The absence of the α,β -unsaturated sidechain in C6 was expected to affect detrimentally the inhibitory activity of the molecule due to the unviability of the covalent bond formation. However, the results have shown that the ligand may have established other non-covalent interactions with the different receptors. Still, this should ideally be deeper analyzed and proved computationally through a modelling study.

Plus, the results shown by the compound **38{2}**, which do not extremely differ from the dataset of **35** analogs but show slightly larger residual values, confirms that either the presence of the terminal heterocycle in C6 sidechain makes an additive contribution to the overall inhibitory activity of the molecule, or the presence of a terminal bromide affects negatively.

Finally, although singlicates at two concentrations have been tested in order to provide a lower and upper limit to rank the different inhibitory behavior of the molecules against the 9 kinases, the upper limit (10nM) has shown more variety of results. Therefore, these latter results will be more extensively disclosed in the following section.

Dacomitinib: Study of substituted quinazoline derivatives as irreversible inhibitors of EGFR

Table 4.5 Results summary of the enzymatic radiometric assay performed by Reaction Biology Corp. Values indicate the residual kinase activity (%) of the 9 kinases tested after treatment with the indicated compounds. Dacomitinib is used as positive control as it is the original hit of the study. In yellow are indicated those results with values ranging 25-50%, in green those with values below 25%.

	Compound ID	30{1}		30{2}		30{3}		30{4}		30{5}		30{6}		30{7}		30{8}		30{9}		38{2}		35{10}		Daco.	
	Assay conc. (nM)	2	10	2	10	2	10	2	10	2	10	2	10	2	10	2	10	2	10	2	10	2	10	2	10
1	EGFR	81	37	74	32	45	7	59	13	60	22	57	17	77	22	72	31	84	24	80	41	67	22	34	5
2	EGFR C797S	88	69	84	73	63	14	76	42	69	52	62	30	94	42	75	57	95	75	88	62	81	54	63	17
3	EGFR G719S	73	28	45	29	41	5	58	14	54	16	55	15	66	16	65	21	79	17	63	29	57	26	35	6
4	EGFR L858R	69	40	71	37	45	6	54	14	57	22	49	24	66	23	61	35	75	22	76	52	71	49	37	6
5	EGFR T790M	79	36	84	60	85	37	69	31	68	23	87	68	82	40	74	32	88	79	89	40	77	33	79	30
6	EGFR T790M/C797S/L858R	84	50	84	58	81	47	76	39	75	40	82	40	89	57	78	40	90	53	87	53	80	41	81	46
7	EGFR T790M/L858R	84	47	60	44	80	41	75	37	78	37	74	39	87	45	75	36	76	51	90	49	76	38	75	40
8	HER2	87	87	85	71	104	51	103	72	94	72	93	62	89	59	81	60	96	69	98	81	92	64	77	45
9	HER4	69	36	61	26	65	26	53	26	57	29	66	26	83	27	55	24	74	28	68	31	61	23	52	14

4.6. Results discussion

Dacomitinib's chemical space described by the combinatorial library developed by the Markush structure present in its patent²²² is computationally unfeasible as if fully enumerated it would involve a total amount of more than 10 billion analogs. An alternative approach has been performed with this family of TKI developing a new database of derivatives considering those 51 structures explicitly described in the drug's patent. Thus, a new library of 16,530 analogs has been created by the fragment combination of the developed molecules by the drug inventors.

Among these, 60 molecules have been described until date in literature. These, as expected, richly represent this chemical subspace described by the 16,530 analogs as its nature mainly relies on the fragment combination of the 85% bibliographical database (51 out of 60). However, as there is a lack of publicly available information concerning the inhibitory activity of these described analogs the aim of this study has been redefined to describe and rank the different biological behavior hidden in the chemical space regions through a rationally selected set of molecules.

The database has been divided in 10 clusters using KMN clustering both aiming to still explore undiscovered regions with a tiny selection of analogs and considering the time and cost limitations of the present research. This has been considered a suitable method to homogeneously classify the data of the combinatorial library of quinazoline analogs. Next, the database has been curated considering the drug-likeness of the analogs (via Lipinski's rule of five) and the availability of the commercial reagents needed. The final accessible chemical space was lowered to 8 clusters after data curation. Thus, 8 analogs were rationally selected by picking the nearest to centroid chemically accessible compound. Hence, the represented chemical space with this selection consists in 80.0% SC and 87.4% PC, similar to the bibliographical data (80.0% SC and 92.6% PC).

The selected analogs were successfully obtained by a 6-step and 7-step designed synthetic routes with overall yields ranging from 6% to 29% with a mean value of 11%. The general obtained yields are mainly limited by the last synthetic step. Overall, during the synthetic process, 44 molecules have been obtained (accounting Dacomitinib analogs and intermediates) and, among these, 35 had not been previously described.

The *in vitro* inhibitory activity of the 8 rationally selected analogs and 3 additional compounds against have been tested at 2nM and 10nM concentrations through a radiometric protein kinase assay (³³PanQinase® Activity Assay) against 9 kinases (wild-type EGFR, 6 mutated EGFR, HER2 and HER4) have been tested. The 10nM concentration results where higher inhibitory activity have been measured (resulting in lower residual activity values) serve as a guide map to better evaluate the compounds efficacy.

The obtained results when testing the inhibition of wild-type EGFR activity surprisingly evidence that all compounds show a residual activity below 50%. Even more, the compound **30{3}** shows a close measured value to the original hit which proofs the hypothesis of the study that their similar biological behavior is related to their chemical similarity as they both take part in the same cluster 3. It should be pointed out that the aim of this implemented process was not rather to optimize the hit area but to significantly represent the chemical space described by the dataset of analogs and no optimization process. Still, compound **30{3}** appears to be an interesting lead candidate.

Table 4.6 Results from the enzymatic radiometric assay of the 12 molecules tested in wild-type EGFR. Values indicate the residual activity in % for the specified concentration. In yellow are indicated those results with values ranging 25-50%, in green those with values below 25%.

Cluster	Exp. ID	ID	10nM
3	Dacomitinib	Dacomitinib	5
3	LMF348	30{3}	7
4	LMF362	30{4}	13
6	LMF374	30{6}	17
7	LMF409	30{7}	22
5	LMF361	30{5}	22
	LMF371	35{10}	22
	LMF363	30{19}	24
8	LMF414	30{18}	31
2	LMF413	30{2}	32
1	LMF415	30{1}	37
	LMF410	38{2}	41

The ranking of these results from more active (lower residual activity, in green) to less active (larger residual activity, in red) and its translation to the original chemical space is shown in Figure 4.43. At a first sight, it can be seen that the three most active clusters (3, 4, 6) are spread in the lower regions of the described space.

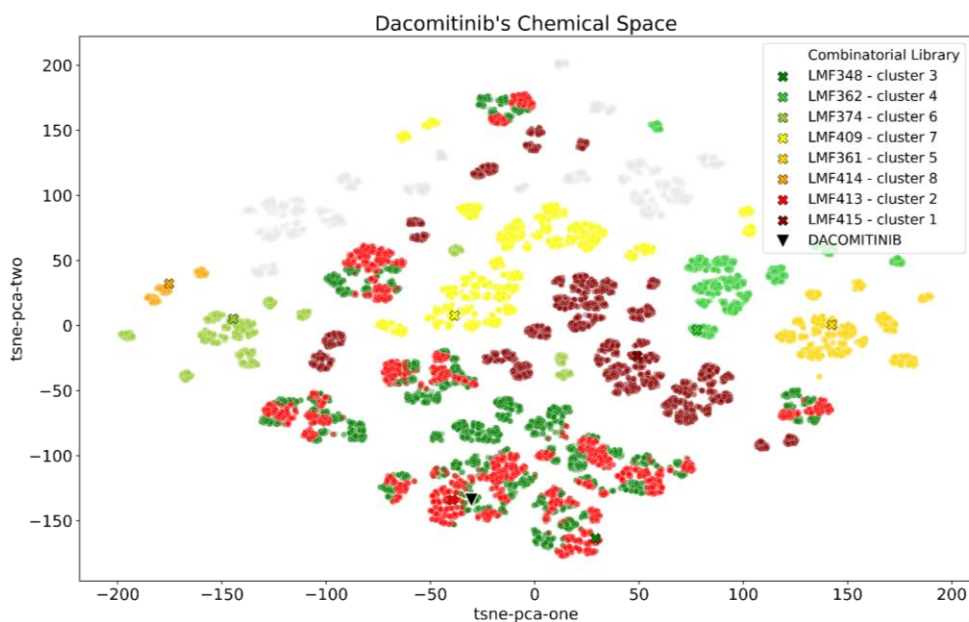


Figure 4.43. Dacomitinib's chemical space with its clusters colored in concordance with the residual activity measured for the rationally selected analogs in EGFR inhibition. In grey are represented the chemical inaccessible molecules during this study (two out of ten clusters).

The following tables (Figure 4.7, Figure 4.8 and Figure 4.9) summarize the ranked results for the tested molecules in the L858R, G719S and C797S respectively mutated EGFR kinase. In all cases, there is an observed trend placing the cluster 3 in the top of the ranking followed by clusters 4, 6, 7 and 5. The set of compounds have shown great activity values for the first two mutations and slightly worst results for C797S which was expected as this widespread mutation is a resistance mechanism commonly induced by covalent inhibitors in order to avoid its interaction with the Cys-797 cleavage. Still, compound **30**₃ appears to be a great TKI as it shows close values to Dacomitinib's.

Table 4.7 Results from the enzymatic radiometric assay of the 12 molecules tested in EGFR L858R.

EGFR L858R	Cluster	Compound	10nM
	3	Dacomitinib	6
	3	30{3}	6
	4	30{4}	14
	5	30{5}	22
		30{9}	22
	7	30{7}	23
	6	30{6}	24
	8	30{8}	35
	2	30{2}	37
	1	30{1}	40
		35{10}	49
		38{2}	52

Table 4.8 Results from the enzymatic radiometric assay of the 12 molecules tested in EGFR G719S.

EGFR G719S	Cluster	Compound	10nM
	3	Dacomitinib	5
	3	30{3}	6
	4	30{4}	14
	6	35{10}	15
	7	30{8}	16
	5	30{6}	16
		30{7}	17
	8	30{2}	21
		30{5}	26
	1	30{9}	28
	2	30{1}	29
		38{2}	29

Table 4.9 Results from the enzymatic radiometric assay of the 12 molecules tested in EGFR C797S.

EGFR C797S	Cluster	Compound	10nM
	3	30{3}	14
	3	Dacomitinib	17
	6	30{6}	30
	4	30{4}	42
	7	30{7}	42
	5	30{5}	52
		35{10}	54
	8	30{8}	57
		38{2}	62
	1	30{1}	69
	2	30{2}	73
		30{9}	75

One of the most interesting, mutated EGFR enzymes to be improved in terms of drug efficacy is T790M as it is the most common acquired resistance mechanism to first- and second-generation EGFR-targeted TKIs. As a second-generation TKI, Dacomitinib is also affected by this mutation. Therefore, an alternative molecule showing a higher inhibitory value for this mutation or double or

triple mutated EGFR showing combined mutations along with T790M (such as T90M/L858R and T790M/C797S/L858R) would be an interesting optimization of the actual hit.

In this case, at the light of the results shown in the following tables (Table 4.10, Table 4.11 and Table 4.12), the inhibitory activity of the candidates has not improved Dacomitinib's except for **30{5}**. However, for the double and triple mutated cases, no great difference was observed between the candidates.

Table 4.10 Results from the enzymatic radiometric assay of the 12 molecules tested in EGFR T790M.

EGFR T790M	Cluster	Compound	10nM
	5	30{5}	23
	3	Dacomitinib	30
	4	30{4}	31
	8	30{8}	32
		35{10}	33
	1	30{1}	36
	3	30{3}	37
	7	30{7}	40
		38{2}	40
	2	30{2}	60
	6	30{6}	68
		30{9}	79

Table 4.11 Results from the enzymatic radiometric assay of the 12 molecules tested in EGFR T790M/L858R.

EGFR T790M/L858R	Cluster	Compound	10nM
	8	30{8}	36
	4	30{4}	37
	5	30{5}	37
		35{10}	38
	6	30{6}	39
	3	Dacomitinib	40
	3	30{3}	41
	2	30{2}	44
	7	30{7}	45
	1	30{1}	47
		38{2}	49
		30{9}	51

Table 4.12 Results from the enzymatic radiometric assay of the 12 molecules tested in EGFR T790M/L858R.

EGFR T790M/C797S/L858R	Cluster	Compound	10nM
	4	30{4}	39
	5	30{5}	40
	8	30{8}	40
	6	30{6}	40
		35{10}	41
	3	Dacomitinib	46
	3	30{3}	47
	1	30{1}	50

EGFR T790M/C797S/L858R	Cluster	Compound	10nM
		30{9}	53
		38{2}	53
	7	30{7}	57
	2	30{2}	58

Finally, the results obtained for the inhibition of HER2 and HER4 are consistent with the measured values for Dacomitinib. On the one hand, HER2 was not highly inhibited by the set of samples (with measured residual values from 45% to 87%). On the other hand, the inhibition *in vitro* tests of HER4 show residual values ranging from 14% (Dacomitinib) until 36%.

To sum up, the **30{3}** analog has shown great activity values against the receptors tested, similar to Dacomitinib. The fact that they are both part of the same cluster could leave the door open for further optimization steps involving the deep exploration of this cluster in order to possibly find more potential candidates with optimal activity values. The Euclidean distance between Dacomitinib and compound **30{3}** in the 13-PCs dimensional space is of 3.075. This is a mid-value considering that the minimum in-cluster distance to the hit is 0.362 and the maximum 7.537. Thus, after re-clustering the cluster 3 containing 3,120 Dacomitinib analogs in 10, a resulting subclassification (shown in Figure 4.44) would place the Dacomitinib molecule in cluster 8 (in dark pink) and the **30{3}** hit in cluster 3 (in green). Therefore, if it was aimed to continue this study in further optimization steps, it would be suggested to rationally select a representative compound of each of the eight unrepresented clusters in order to synthesize them and test their *in vitro* inhibitory activity against the same targets performed in this study.

As an open suggestion that would need a further assessment considering its synthetic feasibility, Figure 4.45 is given to summarize the closest-to-centroid analogs selected for the eight remaining clusters in case any future optimization studies will be performed.

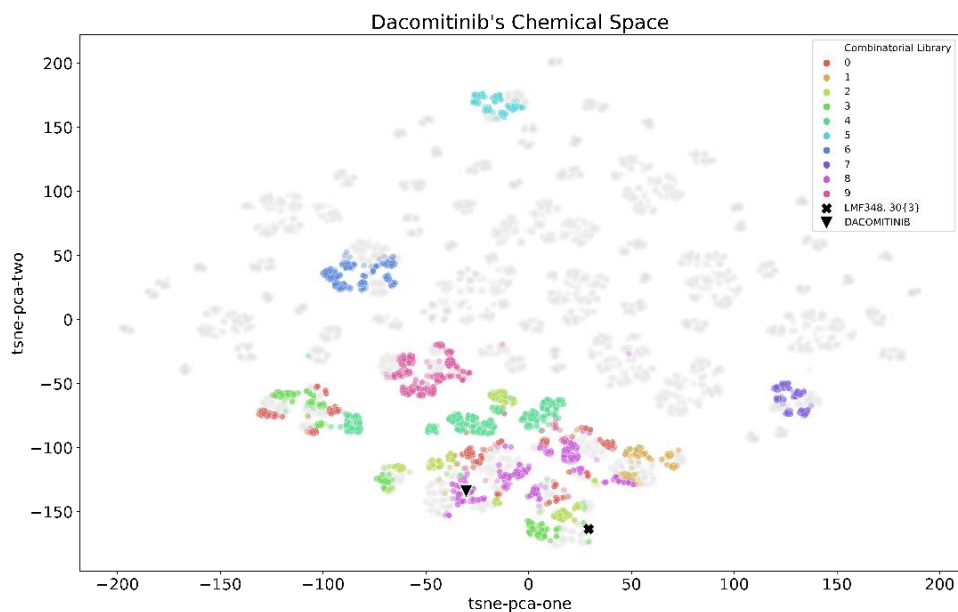


Figure 4.44. Dacomitinib's chemical space with the original $k=3$ cluster re-clustered in 10. In grey are other discarded molecules of the dataset (due to chemical inaccessibility and worst biological activities measured).

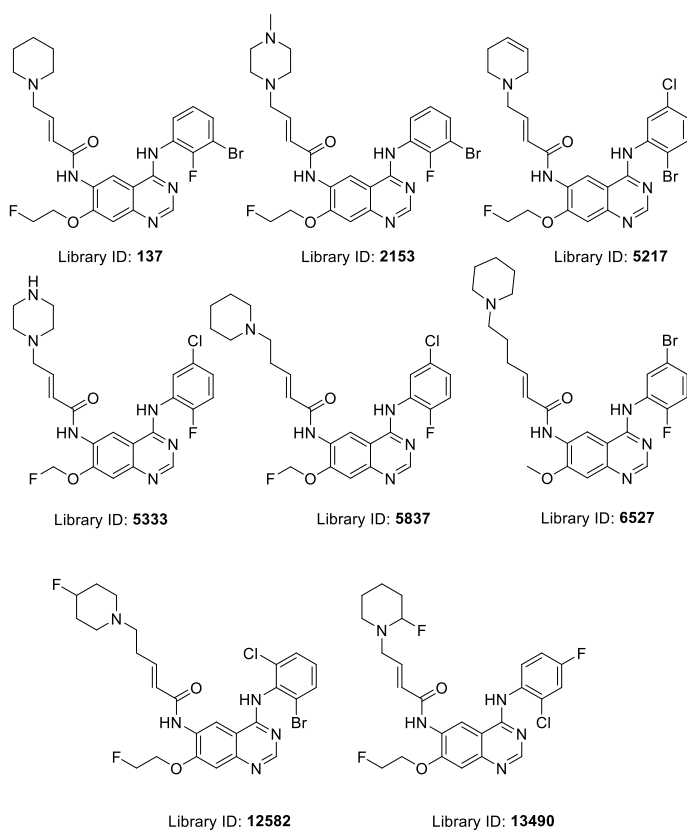


Figure 4.45. Rationally selected compounds selected and recommended to proceed with this study.

4.7. Conclusions

As Dacomitinib's in-patent Markush structure describes a chemical space computationally unattainable with more than 10 billion analogs, an alternative approach applying rational selection has been performed in its study. Instead, the combinatorial library derived from the 51 described analogs in the drug's patent has been developed shaping a new alternative chemical space of analogs. As expected, the bibliographical data found in literature (60 molecules) largely represent this chemical database with 80.0% SC and 92.6% PC. However, the publicly available data concerning to the biological activity of these analogs is extremely short. Thus, a rational selection of the eight synthetically feasible analogs (representing 80.0% SC and 87.4% PC) was performed in order to better describe the inhibitory *in vitro* behavior in the described chemical space. A total amount of 44 molecules were synthesized during the process and, among these, 35 had not been previously described.

The *in vitro* biological assessment of the 8 rational selected analogs and three additional compounds with some structural modifications in contrast with the original combinatorial library were tested with a radiometric protein kinase assay against EGFR (wild-type and 6 of its mutations), HER2 and HER4. All the molecules showed some inhibitory activity at 10nM for all cases except for HER2. Surprisingly, the candidates not considered in the original combinatorial library did also show interesting results. In fact, even the compound **35**{10}, which does not present the typical C6 residue involved in the covalent interaction with the tyrosine kinases, showed some inhibitory activity against 7 of the 9 kinases tested. Compound **30**{3} stands out as the most interesting candidate for further studies as it has shown analogous inhibitory trend compared to Dacomitinib. This showed low residual inhibitory values for wild-type EGFR, the mutations C797S, G719S, L858R, HER4 and intermediate values for T790M, T790M/L858R, T790M/C797S/L858R. In consequence, as this analog and the original hit are part of the same cluster (3), it would be suggested to deeply explore this region in an additional optimization step in the search for a lead candidate. This could be performed through further re-clustering steps in order to probably find more active hits. Above all, this exploratory strategy has successfully find a new possible lead candidate without aiming to optimize the original hit – which may have been obtained as a result of a previous hit-to-lead optimization process – .

5.1. Computational methods; introduction to PyLINS

Due to its extensive libraries support, syntax simplicity and its late popularity in the scientific community, Python 3.7.3 was chosen as the most suitable programming language to develop this project. The described strategy and all the involved developed and applied modules in this thesis involving python coding have been collected in a package named PyLINS. This is an updated version of the previously developed software (in C+ coding) in the group called Pralins (Program for Rational Analysis of Libraries in silico) developed by R. Pascual *et al.*⁹⁰ and contains new functions, clustering and partitioning algorithms.



Figure 5.1 PyLINS logo.

The PyLINS.py script is attached in the Annex IX of the thesis with an exemplifying Jupyter Notebook (PyLINS.ipynb) file.

The general and recommended methodology followed to obtain the different computational results shown in this project were obtained following the schematic representation shown in Figure 5.2. In order to enable the comprehension of this developed strategy, the next sections will be introduced and described following this workchart.

- The correlation of the introduced concepts and methods with PyLINS scheme and nomenclature will be highlighted from now and on in this format during the Section 5.1.

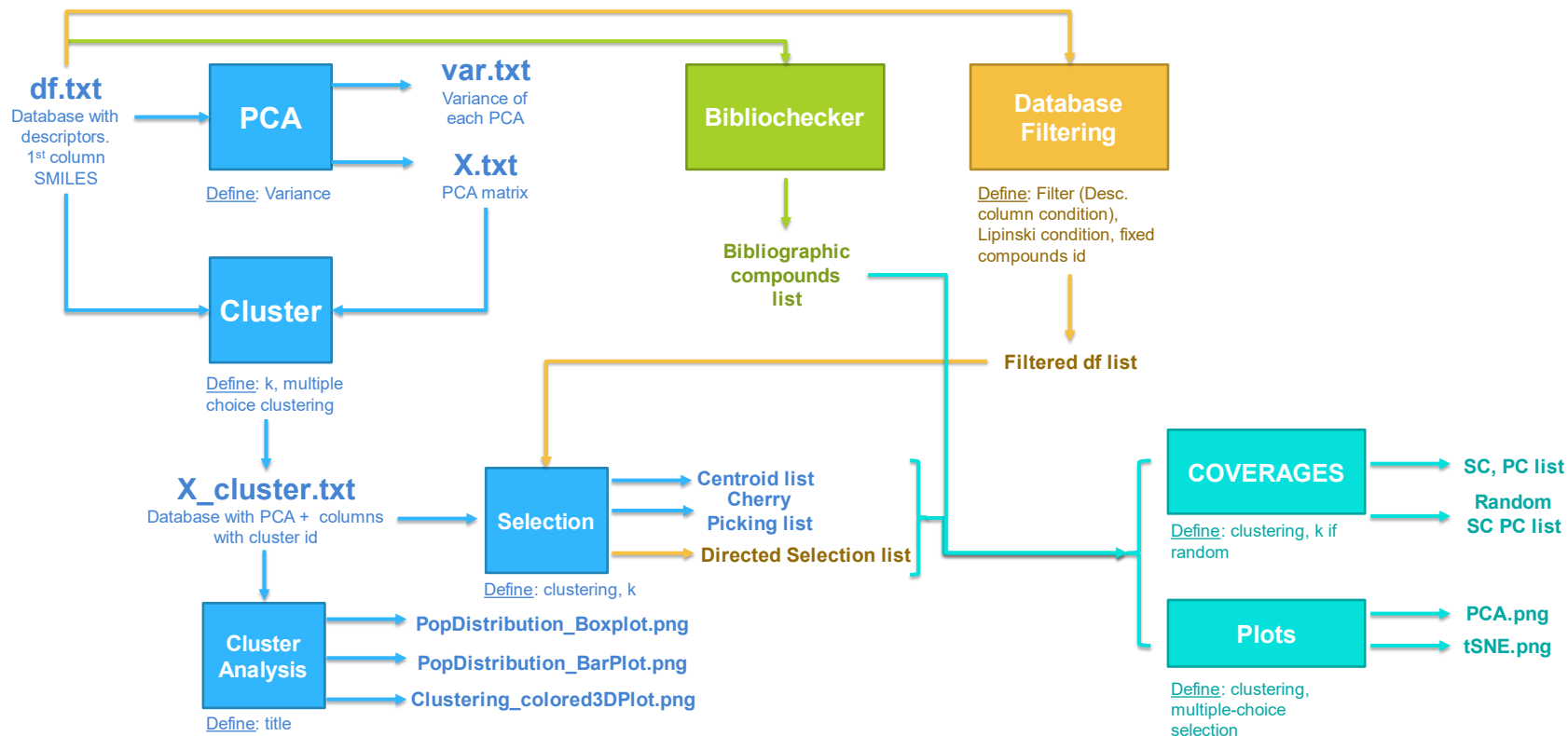


Figure 5.2 Schematic representation of PyLINS workflow

5.1.1. Dataset assembly and curation

5.1.1.1. Combinatorial library enumeration

Markush combinatorial libraries (MCL) were fully enumerated using MarvinSketch for drawing, displaying and enumerating the Markush analogs of the chemical libraries (Marvin 20.21, 2020, ChemAxon Ltd., obtained with an academic license)²³³

5.1.1.2. Data curation and filtering

All datasets including MCL, the small molecule approved by FDA in the period from 2008 to 2020 and ChEMBL database²³⁴ were desalted, protonated at pH 7, and their partial charges were calculated using MMFF94x forcefield using MOE2020.09 software²³⁵. A total of 206 1D and 2D molecular descriptors were calculated using MOE2020.09 (the complete list of molecular descriptors is available in Annex II) to describe the chemical space mathematically.

ChEMBL database was curated selecting only those entries containing organic molecules and a molecular weight lower than 1,000.

Tafenoquine and Dacomitinib were also filtered to describe the chemically accessible space attending to the introduced three criteria stated in Chapter 3 and Chapter 4. RDkit 2019.03.4.0 (rdkit.Chem.Recap() module) package was used to filter those analogs containing undesired structural motifs due to their synthetic unfeasibility. The commercial availability of the necessary reagents for product obtention was checked programmatically using ChemSpiPy 2.0.0. The Lipinski violation filter was implemented through data mining of the datasets as there is a column containing the ['lip_violation'] descriptor.

Additional basic Python packages were involved also in the code definition of the abovementioned filters: matplotlib 3.2.2, numpy 1.21.5 and pandas 1.3.4.

- All these datasets englobe the **df.txt** necessary to start the computational calculations. These must contain all the molecules represented as SMILES in the first column of the matrix.
- The function retrieving the list of commonly consulted chemical suppliers in PyLINS for a given list of reagents is called **ReagentSearcher**

5.1.2. Principal Component Analysis reduction

The key idea of PCA is to find a new system of coordinates in which the input data with dimension \mathbb{M} can be explained by many less variables in a reduced dimension \mathbb{W} with no significant

error. The calculation of these principal components of a certain training can be calculated by the standard matrix factorization named singular value decomposition. This can decompose an original matrix database \mathbb{M} into the dot product of three matrices. This algorithm performs an iterative search of axis directions with the maximum amount of variance in a multidimensional data space leading to an overall transformation from correlated variables to a smaller number of uncorrelated ones for an optimal data presentation. In this new system, the new coordinates are defined by a lineal combination of the original attributes.

Mathematically, the dimensionality reduction of the matrix \mathbb{M} (with n instances and j features) departs from data standardization to lead to matrix \mathbb{Z} as explained in [Eq. 6]. The standardization involves the calculus of the subtraction between every value (p_{nn}) and the mean value of the variable considered ($\overline{p_{jn}}$) divided by the standard deviation ($\sigma_{i,j}$).

$$\mathbb{Z} = \begin{pmatrix} \frac{p_{11} - \overline{p_{j1}}}{\sigma_{p_{j1}}} & \dots & \frac{p_{1n} - \overline{p_{jn}}}{\sigma_{p_{jn}}} \\ \vdots & \ddots & \vdots \\ \frac{p_{n1} - \overline{p_{j1}}}{\sigma_{p_{j1}}} & \dots & \frac{p_{jn} - \overline{p_{jn}}}{\sigma_{p_{jn}}} \end{pmatrix} \quad [\text{Eq. 6}]$$

The new matrix \mathbb{Z} is centralized around zero (mean-centered) and presents a standard deviation of value 1 for all the data in the matrix. This is afterwards used to compute the covariance matrix calculation of the dataset via $\mathbb{X} = \mathbb{Z}^T \cdot \mathbb{Z}$. Finally, through the decomposition of \mathbb{X} the eigenvectors and corresponding eigenvalues can be obtained, these later will explain the percentage of explained variance. The eigenvalues vector (λ) can be obtained by solving $\det(\mathbb{X} - \lambda I) = 0$ and the i th eigenvector (e_i) by solving $\mathbb{X}e_i = \lambda_i e_i$, being. Eigenvectors are finally sorted attending to a decreasing sorting of their corresponding eigenvalues (from higher to lower explained variance) and the desired k eigenvectors are chosen with the largest eigenvalues to form a $n \times k$ dimensional matrix \mathbb{W} with the overall variance of interest. This matrix is obtained as the result of $\mathbb{W} = \mathbb{Z} \cdot e$. The new coordinates will be defined as the projection of the ancient system into the new by solving the linear combination definition described in [Eq. 7] for each principal component.

$$x_1' = p_{11} \cdot e_1 + p_{12} \cdot e_2 + \dots + p_{1k} \cdot e_k \quad [\text{Eq. 7}]$$

The unit vector that defines the j th axis is called the j th principal component (PC). The plot normally is represented by the plane of the first two PCs or the space of the first three PCs (that typically involves above the 60-80% of the overall variance).

For all combinatorial libraries derived from each Markush structure (MCL libraries), the space dimensionality was reduced using Principal Component Analysis (PCA), keeping 95% of the original variance. The PCA was performed through Python scripting using Scikit-Learn 0.23.2²³⁶ and its variance can be interactively visualized thanks to plotly 5.7.0. The visualization plots were favored by both matplotlib 3.2.2 and seaborn 0.11.2 packages.

- Through the **PCA** function it can be obtained the PCs resulting matrix for a desired number of components or explained accumulative variance, the list of variances per PC or an interactive graphic showing the accumulated explained variance per component.

5.1.3. Clustering methodologies

As introduced, clustering is a technique for exploratory data analysis and is used more and more in preliminary analyses of large datasets of small and high dimensionality. It is used as a method of diversity analysis, selection and data reduction. At present, many clustering methodologies have been developed and are used in many studies. These can be generally categorized as *hierarchical* and *non-hierarchical* clustering. Following, a brief review of the main clustering methods used in the chemical databases of the project is introduced.

- **Hierarchic agglomerative clustering (HRC):**

The hierarchic agglomerative classifications are produced in a bottom-up manner by the fusion of individual compounds (also called singletons) into clusters and then the fusion of these into larger clusters with a parent-child relationship resulting in a single cluster that represents the entire data set. Each level of the hierarchy represents a partitioning of the data set into a set of clusters and can be represented through dendrogram plots. These techniques are also known as sequential agglomerative hierarchical non-overlapping (SAHN) methods.

The performance of these algorithms departs from generating a proximity matrix containing all the pairwise proximities between all the pair of clusters or singletons of the dataset. Secondly, the matrix is scanned in order to find the closest pair of clusters/singletons and the merge them into a new cluster by replacing the original term. This step implies the update of the proximity matrix by replacement of the original pair of values by the new distances representing the new cluster proximities with the other remaining clusters. This process is iterated by the continuous modification of the stored matrix until one cluster remains.

Depending on the aggregation criteria chosen to merge the pair of clusters into a final single cluster, different linkage methods have been described^{86,89,237}:

- a) The **single linkage** (HRC single) produces binary tree-like classifications and is characteristic of a range of hierarchic clustering methods in which small clusters of very similar molecules are nested within larger and larger clusters of less closely related molecules. In this case the two clusters are merged by minimizing the distance (D_{ij}) between the closest molecules (x, y) of pair of clusters (C_i and C_j) [Eq. 8].

$$D_{ij} = \min_{x \in C_i, y \in C_j} d(x, y) \quad [\text{Eq. 8}]$$

- b) The **complete linkage** (HRC complete) minimizes the maximum distance between observations of pairs of clusters [Eq. 9].

$$D_{ij} = \max_{x \in C_i, y \in C_j} d(x, y) \quad [\text{Eq. 9}]$$

- c) The **average linkage** (HRC average) minimizes the average of the distances between all observations of pairs of clusters (being n_i, n_j the number of elements in clusters C_i and C_j) [Eq. 10].

$$D_{ij} = \sum_{x \in C_i, y \in C_j} \frac{d(x, y)}{n_i \cdot n_j} \quad [\text{Eq. 10}]$$

- d) The **centroid linkage** (HRC centroid) finds the mean vector location for each of the clusters (or gravity centers) and measures the squared euclidean distance between the two centroids [Eq. 11].

$$D_{ij} = \|\bar{x}_i - \bar{x}_j\|^2 \quad [\text{Eq. 11}]$$

- e) The **median linkage** (HRC median) assigns D_{ij} like the centroid method. When two clusters C_i and C_j are combined into a new cluster, the average of centroids establishes the new centroid. This is also known as the Weighted Pair Group Method with Arithmetic Mean (WPGMC) algorithm.
- f) The **Ward linkage** (HRC Ward) minimizes the total within-cluster variance or the sum of squared differences within all clusters. As a result, it is a variance-minimizing approach and, in this sense, it is similar to the k-means algorithm (later explained) but tackled with a hierarchical agglomerative approach.

A dendrogram is a useful tree diagram to represent the hierarchical clustering arrangement. Each node represents the clusters to which the data belong, and the height represent the distance or

dissimilarity. Figure 5.3 illustrates exemplifying dendrograms evidencing their different distributions and in-cluster connectivity performance depending on the linkage method chosen for the same dataset (first 20 samples of Fischer's Iris Dataset) classification in five clusters.

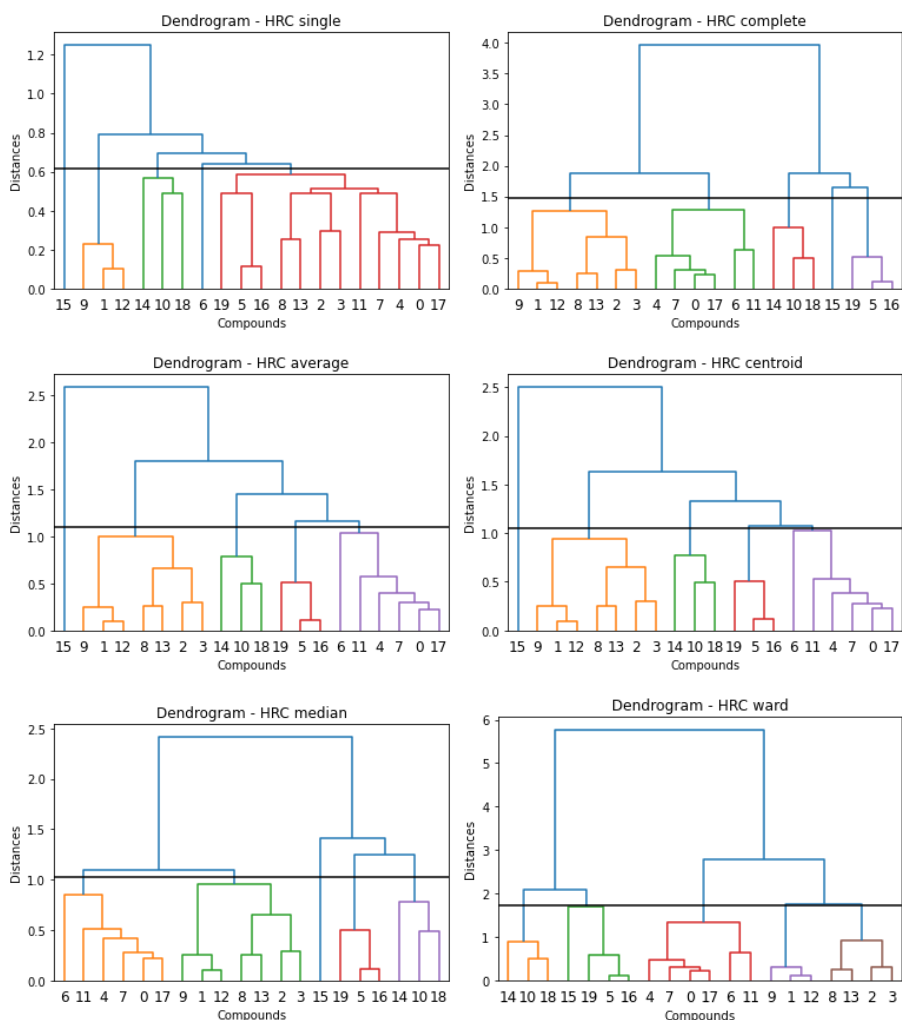


Figure 5.3 Different dendrograms obtained for the hierarchical clustering of the first 20 samples of Fischer's Iris Dataset in $k=5$ clusters with the linkages; single, complete, average, centroid, median and Ward.

For a data set of N compounds, the stored-matrix algorithm for SAHN methods requires $O(N^2)$ time and $O(N^2)$ space for the creation and storage of the proximity matrix while requiring $O(N^3)$ time for the clustering. Thus, this algorithm is very time and resources demanding for large datasets⁸⁶.

HRC algorithms and the calculation of the cophenetic coefficients and dendrograms have been implemented using scipy 1.7.3 package²³⁸.

- **K-means (KMN):**

This relocation algorithm departs from a random selected set of compounds as starting point which serve as cluster centroids. Compounds in their surrounds are assigned to the nearest cluster and the centroids are recalculated continuously as data is merged through an iterative assessment until convergence.

This iterative algorithm tries to partition the dataset into a predefined number of different non-overlapping clusters (k) trying to make the intra-cluster data points to the greatest similarity while keeping the cluster centroids as far (or, in other terms, dissimilar) as possible. Mathematically, data is assigned to a cluster such that the sum of the squared distance between the data points and the centroid of the cluster is at the minimum. This leads to clusters with small variation in its data, which can be translated to more homogeneous or similar data points are within the same cluster.

However, given the iterative nature of KMN and the initialization of a set of centroids at the start of the algorithm, different centroid departing points may lead to different clusters since KMN algorithm may stuck in a local optimum and may not converge to global optimum.

For a data set of N compounds, the complexity of the algorithm is $O(Nmk)$, where k is the number of clusters and m the number of iterations needed until convergence, which is normally negligible²³⁹.

KMN algorithm has been implemented using scipy 1.7.3 packages²³⁸.

- **K-medoid (KMED):**

KMED algorithm is a clustering methodology similar to KMN and was named after Kaufman *et al.*²⁴⁰ work with their PAM algorithm. While KMN tries to minimize the within cluster sum-of-squares via centroid recalculation, KMED tries to minimize the sum of distances between each point and the medoid of its cluster. The medoid consists in a data point of the set which has the least total distance to the other members of its cluster and is used to represent the center of the cluster. Hence, in this case, different distance metrics or similarity coefficients can be used.

The complexity of the algorithm is $O(N^2mk)$, where k is the number of clusters and m the number of iterations needed until convergence. This makes it more suitable for smaller datasets in comparison to KMN.

KMED algorithm has been implemented using pyclust 0.2.0 package²⁴¹.

- **Spectral Clustering:**

Spectral clustering is a methodology with origins in graph theory. This approach is used to identify groups of nodes in a graph based on the edges connecting them²⁴². Hence, the main tools for spectral clustering are graph Laplacian matrices²⁴³. It is a flexible method and enables to cluster non graph data as well. This algorithm uses information from the eigenvalues of special matrices built from the graph or the data set. It is very useful when the shape of the individual clusters is highly non-convex, or in other terms, when the measurement of the center of the cluster is not a suitable description of the complete cluster, for spread values such as when clusters are nested circles on a 2D plane. Hence, it is highly recommended for non-flat geometry data points distribution.

The basic methodology of the algorithm can be explained in four steps:

1. The Laplacian (or normalized Laplacian) L matrix is calculated.
2. The first k eigenvectors are calculated (corresponding to the k smallest eigenvalues of L).
3. A matrix formed by the first k eigenvectors is built-up. In this, the i th row defines the features of i graph node.
4. The graph nodes are clustered based on these features.

Spectral clustering algorithm has been implemented using Scikit-Learn 0.23.2 package²³⁶.

Overall, as it is known in the clustering field, depending on the shape, dimensionality and distribution of the dataset, one algorithm may be more suitable than another, there is no optimal approach which is applicable to all general sets. Each dataset of study should be uniquely considered and assessed although some common applications tend to use explode certain methodologies (e.g., KMED for face recognition).

5.1.4. Binning and Optimum Variance binning

Cell-based or binning is a method partitioning method to discretize the values of a continuous variable into bins or partitions seeking for a maximal occupancy. Through binning methodology, the partition of the descriptor or reduced chemical space is performed via unsupervised techniques which divide it in equal-width and equal-size or equal-frequency interval binning leading to the generation of hypercubic cells in which the compounds are placed according to their property values. The binning process might require satisfying certain constraints which might range from requiring a minimum number of records per bin to monotonicity constraints²⁴⁴. Normally,

the process divides the property ranges in such manner that the number of occupied cells is always equal or less than to the number of desired molecules to select.

Programmatically, the binning methodology sorts the Principal Components axes (PCAs) by descending order of its explained variance. When a certain number of bins is requested (k), the algorithm will divide each axis subsequently until the number of occupied bins are equal or less than k . Hence, each partition results in the formation of 2^n bins, being n the number of divided PCs. However, although the definition of a binning size (k) will divide the dataset in the closest $2^n < k$ value, this will not imply that all of them are occupied. For all cases, it will be confirmed $k_{oc} \leq 2^n \leq k$.

An alternative to simple binning partition (Figure 5.4) has been developed and implemented in this project adapting the optimum binning approach described by Pascual *et al.*^{41,90}. This new approach, called Optimum Variance binning (OV binning) bins the properties with a bias towards the ones that exhibit the largest variation. Hence, those properties with large ranges have a tendency to have more divisions than the others, aiming to form bins as much equal faced as possible while maintaining the desired number of occupied cells.

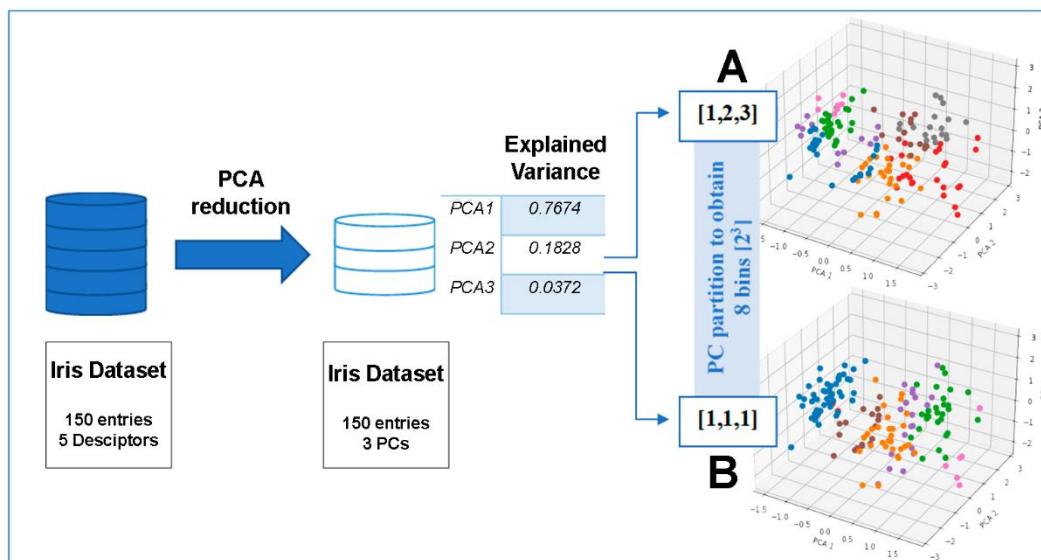


Figure 5.4 Description of the Binning and OV Binning partitioning methods implemented in this study. Fisher's Iris dataset (Fisher, R. A. Ann. Eugen. 1936, 7 (2), 179–188) is used in the example to show how (a) binning partitioned the space in the three firsts PCs

The inner mechanism of OV Binning algorithm also sorts the PCAs by descending order of its explained variance but, assuming that the division of an axis divides its variance in two subaxes, the following partition is reimplemented in the same PCA only if its variance is greater than the explained variance of the following axis. With this procedure, multiple partitions can be observed in the first

PCAs as they contain higher explained variance values proceeding from the Principal Component Analysis. Hence, molecules are binned with a bias toward the PCs that exhibit the largest variation

An example of the algorithms difference between binning and optimum variance binning is shown in Figure 5.5. This shows the representation of the 8 bins formation of the same dataset displayed in its 3D (A, B) and 2D (C, D) chemical space. For 2D examples, partition subaxes are represented when the showed axes are divided: once for PCA1 and PCA2 in binning (C, with a subsequent not-visible third partition in PCA3) and three times for PCA1 in optimum binning as cells have been subdivide attending to balance their grade of variance (D).

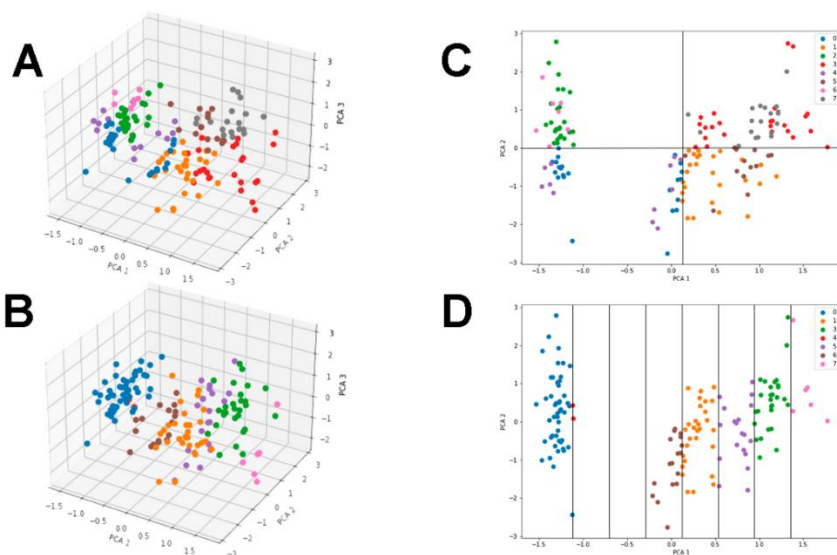


Figure 5.5. Application of binning and optimum binning on an illustrative dataset to divide the chemical space into 8 bins.

The complexity of cell-based techniques does not exceed $O(N)$ making these procedures suitable for large datasets. However, they suffer from edge effects, contrary to cluster-based methods²⁴⁵.

- The desired calculation of all or individual clustering/partitioning methodologies can be achieved through the **Cluster** module which will retrieve a matrix (**X_cluster**) containing the PCA system followed by columns containing the cluster index of each analog.
- Other interesting functions implemented for cluster analysis can be found such as:
- **PopDistribution_Boxplots**: Given an X_Cluster matrix and the desired list of clustering/partitioning names to be analyzed, it retrieves a plot representing the

individual boxplots of the different population distribution for each clustering/partitioning method.

- **Cluster3d_distribution:** Given a X_Cluster matrix, a better exploratory analysis through the clusters/bins formed by a desired algorithm can be performed. This function retrieves: (1) the colored database in the 3D PCA plot depending on their clustering indexes, (2) a bar plot with the population in each cluster and (3) information of the most and less populated clusters. Afterwards, it asks the user to choose any the cluster of interest for the study and displays the chemical space containing only the molecules within the specified cluster.
- **Clusters_3dplot:** Given a X_Cluster matrix and an array of the desired clustering methodologies (clustering_names). Different 3D plots are returned showing the chemical space with the different analogs colored depending on its cluster index.

5.1.5. Library sizes and selections

5.1.5.1. Bibliographic data search

The bibliographic data search has been performed programmatically by using Pubchem's PUG-REST²⁴⁶ Application Programming Interface (API). The whole procedure was integrated into a Python script, which checked the presence of a given compound (entered as a SMILES) on PubChem.

- Through the **PubChemBibliographicDataSearcher** function, the user can introduce the desired string of smiles to be checked and will receive back a dataframe containing the library index and related Pubchem CIDs of the reported molecules.

5.1.5.2. Library sizes

Different numbers of clusters have been studied in the present work to assess the chemical space in terms of population and space coverage. Actually, the number of partitions (whether they are clusters or bins) ultimately determines the number of compounds to pick via rational selection, since they will correspond to the representative compound from each partition.

Preliminary studies were carried out to assess the behavior of a k-optimal, $\frac{\sqrt{N}}{2}$, \sqrt{N} and $10 \cdot \sqrt{N}$ space fragmentation, being N the total number of compounds in each dataset. Finally, \sqrt{N} was taken as a reference in clustering to optimally represent large datasets. Moreover, BD and BCL sizes (N_{BD} and N_{BCL} , correspondingly) were also used to discuss the representativeness of the analogs known until the date, in contrast with a random selection of the same size. Random selections were calculated as the mean value of the coverages represented by N_{BD} and N_{BCL} random samples (with

5000 repetitions) as a contrasting result with data found in the bibliography or other rational selections.

5.1.5.3. Rational selection

Centroid coordinates are calculated for each cluster or bin as the mean of each principal component for compounds in it. The Euclidean distance is therefore calculated between all compounds belonging to a particular cluster or bin and its centroid. The nearest compound to the centroid is selected or, if there are inaccessibility-related issues, the closest chemically accessible compound will be chosen.

An alternative methodology is the cherry-picking selection. In this case, a random compound per cluster is chosen to represent each group of molecules.

- The class **Selection** allows both selection options for a given dataset and clustering/partitioning methodology.
- Also, this same class contains the function **directed_selection** which returns a matrix with compounds sorted in nearest position to cluster's centroid in case synthetic unfeasibility reasons hinders the obtention of the first centroid candidate.

5.1.6. Space and Population coverage calculation

Space and Population coverages are two parameters highly explored in this work to assess the representativeness of given subsets of interest (e.g. BD, rational selections). The equations introduced in Chapter 1 are reproduced here to favor the reader's comprehension.

Space Coverage (SC) represents the percentage of selected (occupied) clusters or bins (k_{oc}) by a given number of selected compounds over the total number of partitions, k [Eq. 4].

$$SC = \frac{k_{oc}}{K} \cdot 100 \quad [\text{Eq. 4}]$$

Given a database of N compounds, Population Coverage (PC) measures the SC weighted by the occupancy in each cluster or bin, by dividing the population of the occupied clusters (n_{oc}) among N [Eq. 5].

$$PC = \frac{n_{oc}}{N} \cdot 100 \quad [\text{Eq. 5}]$$

Complementary, an interesting application of their calculation has been exploited in this research through the representation of heatmaps representing the different algorithm distribution overlap. This departs from setting all the centroids of the studied methodologies and calculates their coverage coefficient (SC or PC) in the other clustering systems.

- These coefficients can be calculated through the **SpaceCoverage** and **PopCoverage** functions.
- The class **coverage_heatmap** enables the heatmap calculation and representation for SC (**spacecov**) and PC (**popcov**) of a given `X_cluster` and the desired clustering/partitioning methodologies.

5.1.7. t-SNE plots

Another approach for high-dimensional data processing is the t-SNE (t-Distributed Stochastic Neighbor Embedding), developed by Van der Maaten⁵⁶ which has gained large popularity for data visualization. Briefly, it is a nonlinear dimensionality reduction which departs from the calculation of the distance matrix calculated through a given metric, then it proceeds establishing a Gaussian probability distribution which is expected to represent the similarities between neighbor samples. Data points are embedded in a low-dimensional space (2D or 3D) attending to joint probabilities (local densities) and divergence is minimized (Figure 5.6). One of the major weaknesses of t-SNE is that the cost function is not convex, as a result of which several optimization parameters need to be chosen.

Two key features are fundamental for the embedding distribution in order to balance the resulting data; the establishment of a perplexity value (which targets the desired amount of neighbors to the central point, it is suggested to use values between 5-50) and the number of iterations to train the model in order to converge²⁴⁷. However, this approach present two notorious drawbacks²⁴⁸; in one hand, it cannot be applied to new data as its sensitive to the curse of the intrinsic dimensionality of the data, with different initializations we can get different result so the modification of the dataset obliges to reevaluate the system, in the other hand, the use of this methodology implies the complexity of distance calculation and is computational consuming in terms of the quadratic calculation of its datapoints. Overall, although it is a good tool for visualization maps for large datasets, it is not recommended or still unclear the use for further reduction purposes as it loses high resolution and a subsequent predictive power. The behavior of t-SNE when reducing data to two or

three coordinate system cannot readily be extrapolated for more dimensions because of the heavy tails of its t-student distribution. Hence, when using large datasets, much part of the local structure of the dataset may not be preserved as a relatively large portion of the probability mass will be placed in the tails of the distribution.

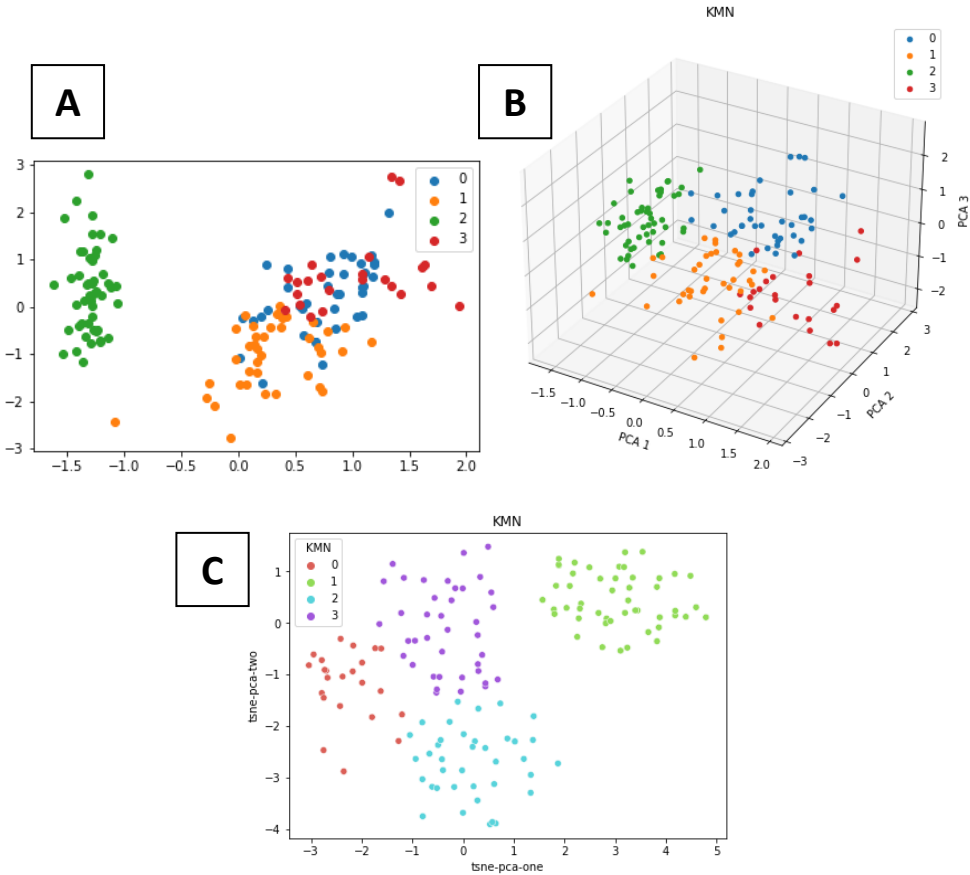


Figure 5.6 Graphical representation of Fisher's Iris Dataset clustered in 4 groups via KMN and visualized by: its two firsts PCs (A), its three firsts PCs (B) and t-SNE plot with perplexity = 80 and 10,000 iterations (C).

- The function `tsne_pylins` simplifies the process for looped optimization studies by only demanding the PCA coordinates, perplexity and iterations values and results in a plot showing the t-SNE studied system.

5.2. Synthesis and compound characterization

5.2.1. Instrumentation

Nuclear Magnetic Resonance spectra ($^1\text{H-NMR}$ and $^{13}\text{C-NMR}$) were recorded in a Varian 400-MR spectrometer ($^1\text{H-NMR}$ 400 MHz and $^{13}\text{C-NMR}$ 100.6 MHz), in the Organic Chemistry Department at IQS under the leadership of Dr. X. Batllori. Chemical shifts are reported in parts per million (ppm) on the δ scale and referenced to the residual signal of the solvent (DMSO- d_6 : 2.5 ppm in $^1\text{H-NMR}$ and 39.52 ppm in $^{13}\text{C-NMR}$, CDCl_3 : 7.26 ppm in $^1\text{H-NMR}$ and 77.16 ppm in $^{13}\text{C-NMR}$, TFA- d : 11.50 ppm in $^1\text{H-NMR}$ and 164.2 ppm and 116 ppm in $^{13}\text{C-NMR}$). Coupling constants (J) are reported in Hertz (Hz). Standard and peak multiplicities are designated as follows: s (singlet), d (doublet), dd (doublet of doublets), ddd (doublet of doublet of doublets), t (triplet), q (quartet), qn (quintet), m (multiplet), br (broad signal). Other spectral splitting pattern designated is * which accounts for exchangeable signal assignments.

Infrared Spectra (IR) were recorded in a Nicolet iS10 FTIR spectrometer with Smart iTr (Thermo Scientific) by Mrs. N. Ruiz, Dr. Albert Gibert and the present author at the Organic and Pharmaceutical Chemistry department at IQS under the leadership of Dr. X. Batllori, using a potassium bromide (KBr) disc or ATR. Wave number values are reported in cm^{-1} . The notation used is: st (stretching), b (bending), oop (out of plane), sc (scissoring), w (wagging).

Organic elemental analysis (OEA) were obtained on a EuroEA 3000 Elemental Analyzer (EuroVector) by Mrs. N. Ruiz at the Organic and Pharmaceutical Chemistry department at IQS under the leadership of Dr. X. Batllori at the Organic and Pharmaceutical Chemistry Department at IQS.

Mass Spectrometry (MS) was conducted on an Agilent Technologies 5975 mass spectrometer operating in electron ionization (EI) mode at 70 eV and at 4kV accelerating potential.

High Resolution Mass Spectrometry (HRMS) was conducted on a SCIEX X500B QTOF high resolution mass spectrometer operating in ESI mode at the Organic and Pharmaceutical Chemistry department at IQS under the leadership of Dr. Ana Belén Cuenca.

The melting point (MP) was determined with a SMP3 melting point apparatus (Stuart Scientific).

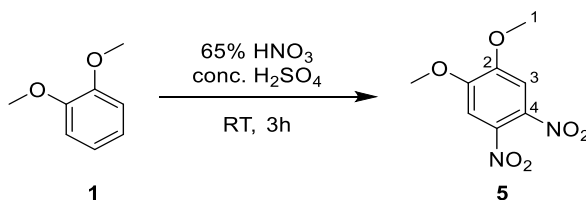
Automatic flash chromatography was performed in an Combiflash[®]Rf (Teledyne Isco) with RediStep[®] silica gel columns or basic alumina columns.

Microwave irradiation experiments were carried out in an InitiatorTM (Biotage) microwave apparatus, operating at a frequency of 2.45 GHz with continuous irradiation power from 0 to 400 W.

Reactions were performed in 5 mL glass tubes, sealed with aluminum/Teflon crimp tops, which can be exposed up to 250 °C and 20 bar internal pressure. Temperature was measured with an IR sensor on the outer surface of the process vial. After the irradiation period, the reaction vessel was cooled rapidly to 50 °C by air jet cooling.

5.2.2. Synthesis of 8-aminoquinoline derivatives

5.2.2.1. Synthesis of 1,2-dimethoxy-4,5-dinitrobenzene



5.00g (36.19 mmol, 1 eq) of veratrole (**1**) were added dropwise to 22.63 ml (9 eq.) of cooled stirred 65% nitric acid keeping the temperature below 50 °C during the addition. After 30 minutes the mixture resulted in a red solution with a dark-red precipitate (the mononitro derivative). 7.14 ml (3.7 eq.) of 96% sulfuric acid were added dropwise and the mixture was stirred at room temperature for 2 hours leading to a dark suspension. Ice and water were added to the suspension to precipitate the product. The solid was filtered, washed with water, and dried *in vacuo* over P₂O₅ to afford 7.72 g of 1,2-dimethoxy-4,5-dinitrobenzene (**5**, **LMF201**) as pale-yellow solid (98% yield).

Spectroscopic data:

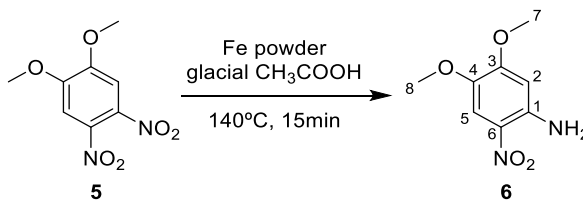
¹H NMR (400 MHz, CDCl₃) δ 7.34 (d, *J* = 0.4 Hz, 2H, C3-H), 4.01 (d, *J* = 0.4 Hz, 6H, C1-H₃).

¹³C NMR (100 MHz, CDCl₃) δ 152.0 (C2), 136.9 (C4), 107.1 (C3), 57.2 (C1).

IR (KBr), ν_{max} (cm⁻¹) 3074 (*st* Csp²-H), 1528 (*asym st* NO₂), 1373 (*sym st* NO₂), 1285 (*st* Ar-O-CH₃), 1234, 1050, 880, 789 (*oop b* C_{ar}-H).

The recorded spectroscopic data are consistent with the previously described (bibliographic yield: 96%)¹⁶⁵.

5.2.2.2. Synthesis of 4,5-dimethoxy-2-nitroaniline



6.00g (26.30 mmol, 1 eq) of intermediate **5** were added to 50 ml of glacial acetic acid in an argon-flushed 250 ml three-neck round-bottom flask equipped with a reflux condenser. The three-neck system was closed, and the mixture was heated to 165 °C until boiling was observed. Rapidly, after removing the heating plate, 4.43g of iron powder (79.33 mmol, 3.0 eq.) were added with vigorous stirring and a quick spontaneous boiling was observed turning the crude mixture in a dark brown color with a black solid precipitate. After the exothermic reaction had subsided (~5 minutes) the mixture was refluxed for 10 minutes and the poured into 300ml ice/water bath observing the formation of an orange-red precipitate which is furtherly filtrated. After filtration, the product is isolated by its dilution using acetone as solvent and separating it from the solid iron traces of the crude. The solvent is afterwards removed by rotary evaporation and the slightly wet product is dried *in vacuo* over P₂O₅ overnight to afford 4.51 g of 4,5-dimethoxy-2-nitroaniline (**6**, **LMF241**) as an orange-red solid (87% yield).

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 7.46 (s, 2H, NH₂), 7.33 (s, 1H, C5-H), 6.51 (s, 1H, C2-H), 3.80 (s, 3H, C8-H₃), 3.72 (s, 3H, C7-H₃).

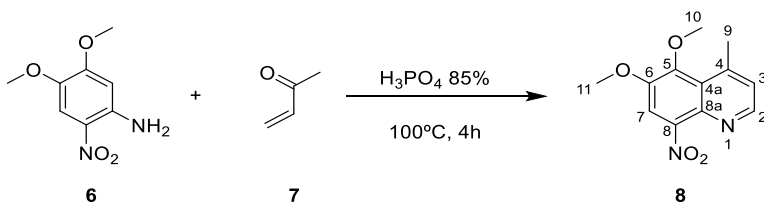
¹³C NMR (100 MHz, DMSO-*d*₆) δ 156.8 (C3), 144.5 (C4), 140.6 (C1), 122.3 (C6), 105.2 (C5), 99.1 (C2), 55.8 (C6), 55.7 (C7).

IR (KBr), ν_{max} (cm⁻¹) 3458 (*asym st* NH₂), 3329 (*sym st* NH₂), 3155 (*st* Csp²-H), 1513 (*asym st* NO₂), 1397 (*sym st* NO₂), 1253 (*st* C-N), 1097, 997, 851, 607.

Elemental analysis: Calculated for C₈H₁₀N₂O₄: C: 48.49%, H: 5.09%, N: 14.14%. Found: C: 48.29%, H: 5.00%, N: 13.79%.

The recorded spectroscopic data are consistent with those previously described (bibliographic yield: 76%)^{249,250}.

5.2.2.3. Synthesis of 5,6-dimethoxy-4-methyl-8-nitroquinoline



2.40g (1.21 mmol, 1.0 eq) of intermediate **6** and 18ml of 85% phosphoric acid was placed in a 50ml round-bottomed flask. The reaction mixture was stirred and heated to 100 °C when 0.22ml of methyl vinyl ketone (2.70 mmol, 2.2 eq.) was added dropwise. The reaction mixture was heated and stirred for 4h and the resulting dark solution was poured onto ice water (50ml). The aqueous crude was alkalinized with 30% ammonia, extracted with AcOEt (2x100ml) and washed with water (3x25ml). The extract was dried over anhydrous Na₂SO₄ and concentrated to dryness. The black residue was finally chromatographed over silica gel using 30:70 Cy-DCM as eluant to afford 1.10g of 5,6-dimethoxy-4-methyl-8-nitroquinoline (**8**, **LMF245**) as a yellow crystalline product (37% yield).

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 8.65 (d, *J* = 4.3 Hz, C2-1H), 8.32 (s, 1H, C7-H), 7.40 (dd, *J* = 4.3, 0.9 Hz, C3-H), 4.00 (d, *J* = 0.6 Hz, C11-H₃), 3.92 (d, *J* = 0.6 Hz, C10-H₃), 2.85 (t, *J* = 0.8 Hz, C9-H₃).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 150.1 (C2), 148.3 (C6), 146.2 (C5), 144.8 (C8a*), 144.0 (C4), 134.9 (C8*), 125.2 (C3), 123.5 (C4a), 112.1 (C7), 61.0 (C10), 57.2 (C11), 22.7 (C9).

IR (KBr), ν_{max} (cm⁻¹) 3431, 2951, 1938, 1592, 1530, (*asym st* NO₂), 1350 (*sym st* NO₂), 1308, 1260 (*st* C-N), 1116, 1044, 971, 855, 768 (*oop b* C_{ar}-H).

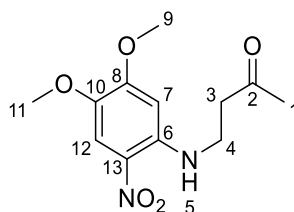
Elemental analysis: Calculated for C₁₂H₁₂N₂O₄: C: 58.06%, H: 4.87%, N: 11.29%. Found: C: 58.06%, H: 4.75%, N: 11.04%.

MS (70 eV, EI) m/z (%): 249.1 (16%) [M+1]⁺, 248.1 (100%) [M]⁺, 218.1 (29%) [M-CH₂O]⁺, 188.1 (32%).

mp: >250 °C

The melting point measured is consistent with previously described data (bibliographic yield: 30%)¹²⁶.

4-((4,5-dimethoxy-2-nitrophenyl)amino)butan-2-one



19

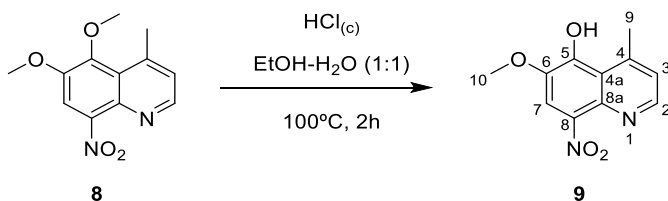
Intermediate (**19**) obtained as a red crystalline solid after column purification of the crude in the isolation of 5,6-dimethoxy-4-methyl-8-nitroquinoline (**8**).

Spectroscopic data:

$^1\text{H NMR}$ (400 MHz, $\text{DMSO-}d_6$) δ 8.52 (t, $J = 5.9$ Hz, 1H, N5-H), 7.44 (d, $J = 2.6$ Hz, 1H, C12-H), 6.41 (d, $J = 2.7$ Hz, 1H, C7-H), 3.92 (d, $J = 2.8$ Hz, 3H, C9*-H₃), 3.73 (d, $J = 2.7$ Hz, 3H, C11*-H₃), 3.58 (dt, $J = 5.9, 6.2$ Hz, 2H, C4-H₂), 2.88 (t, $J = 6.4$ Hz, 2H, C3-H₂), 2.14 (s, 3H, C1-H₃).

$^{13}\text{C NMR}$ (100 MHz, $\text{DMSO-}d_6$) δ 207.4 (C2), 157.7 (C8**), 143.7 (C6), 140.3 (C10**), 122.7 (C13), 106.4 (C12), 95.0 (C7), 56.3 (C9*), 55.7 (C11*), 41.9 (C3), 37.3 (C4), 30.0 (C1).

5.2.2.4. Synthesis of 6-methoxy-4-methyl-8-nitroquinolin-5-ol



1.43g (5.79 mmol, 1.0 eq) of intermediate **8** was refluxed with 5 ml of concentrated hydrochloric acid and 40 ml of ethanol-water (1:1) solution in a 100 ml round-bottomed flask at 100 °C upon 3h. Upon cooling in an ice-water bath, the formation of red crystals was observed. The crude was left to crystallize overnight and was afterwards filtrated and dried *in vacuo* over P_2O_5 to afford 995.8mg of 6-methoxy-4-methyl-8-nitroquinolin-5-ol (**9**, **LMF204**) as a red-brown crystalline product (73% yield).

Spectroscopic data:

$^1\text{H NMR}$ (400 MHz, $\text{TFA-}d$) δ 8.98 (d, $J = 5.8$ Hz, 1H, C2-H), 8.89 (s, 1H, C7-H), 7.86 (d, $J = 5.8$ Hz, 1H, C3-H), 4.24 (s, 3H, C10-H₃), 3.35 (s, 3H, C9-H₃).

^{13}C NMR (100 MHz, TFA-*d*) δ 167.5 (C4*), 155.5 (C5**), 146.6 (C6), 146.4 (C2), 130.3 (C8**), 129.7 (C8a**), 126.2 (C3), 121.3 (C4a*), 119.9 (C7), 59.1 (C10), 27.2 (C9).

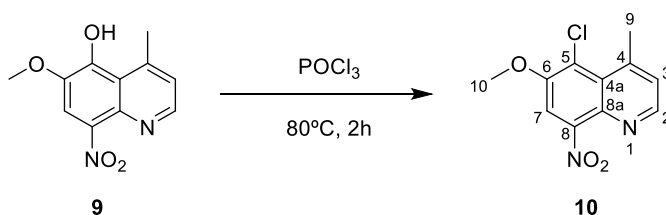
IR (KBr), ν_{max} (cm^{-1}) 3431 (*st* OH), 3080 (*st* C_{ar}-H), 1624, 1591, 1557 (*asym st* NO₂), 1382 (*st* C-N), 1315 (*sym st* NO₂), 1266, 1202, 1155 (*st* C-O), 1096, 1048, 998, 929, 834, 797 (*oop b* C_{ar}-H).

MS (70 eV, EI) m/z (%): 235.1 (19%) [M+1]⁺, 234.1 (100%) [M]⁺, 204.1 (53%) [M-CH₂O]⁺, 189.1 (31%), 174.1 (29%), 146.1 (23%), 120.0 (21%), 117.1 (36%)

mp: >250 °C

Melting point measured is consistent with previously described data (bibliographic yield: 92%)¹²⁶.

5.2.2.5. Synthesis of 5-chloro-6-methoxy-4-methyl-8-nitroquinoline



750.7mg (3.20 mmol, 1.0 eq) of intermediate **9** was placed with 10.8 ml of phosphorous (V) oxychloride in a 5 ml round-bottomed flask and refluxed at 80° for 2h. The mixture was later cooled with an ice-water batch and quenched with aqueous ammonia which was poured dropwise carefully to smoothly hydrolyze the unreacted remaining POCl₃. Base was slowly poured until neutralization avoiding a vigorous exothermic reaction of the crude which may present the formation of hydrogen chloride fumes during neutralization. The residue was extracted with three parts of DCM that were later unified in a single orange organic phase which was dried over anhydrous MgSO₄. The solvent was finally removed under reduced pressure to afford 797.8mg of 5-chloro-6-methoxy-4-methyl-8-nitroquinoline (**10**, LMF205) as a light brown powdery product with quantitative yields (95%).

Spectroscopic data:

^1H NMR (400 MHz, DMSO-*d*₆) δ 8.71 (d, J = 4.3 Hz, 1H, C2-H), 8.43 (s, 1H, C7-H), 7.55 (dd, J = 4.4, 1.0 Hz, 1H, C3-H), 4.05 (s, 3H, C10-H₃), 3.04 (d, J = 1.0 Hz, 3H, C9-H₃).

^{13}C NMR (100 MHz, DMSO-*d*₆) δ 152.5 (C6), 150.0 (C2), 148.5 (C8a*), 144.4 (C4), 134.8 (C8*), 127.1 (C3), 127.0 (C4a), 118.1 (C5*), 110.2 (C7), 57.7 (C10), 24.8 (C9).

IR (KBr), ν_{max} (cm^{-1}) 3426, 3092 (*st* C_{ar}-H), 2944, 2682, 1533 (*asym st* NO₂), 1383, 1340 (*sym st* NO₂), 1253 (*st* C-N), 1106, 1055, 899, 769 (*oop b* C_{ar}-H).

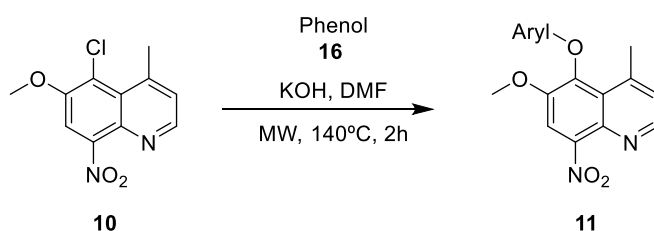
Elemental analysis: Calculated for $C_{11}H_9ClN_2O_3$: C: 52.29%, H: 3.59%, N: 11.09%. Found: C: 51.76%, H: 3.47 %, N: 10.33%.

MS (70 eV, EI) m/z (%): 254.1 (35%) $[M+2]^+$, 253.1 (16%) $[M+1]^+$, 252.1 (100%) $[M]^+$, 222.1 (34%) $[M-CH_2O]^+$, 194.1 (29%), 128.1 (28%).

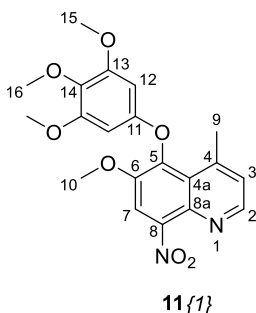
mp: 165-170 °C

Melting point measured is consistent with previously described data (bibliographic yield: 69%)¹²⁶.

5.2.2.6. Synthesis of 5-phenoxy-6-methoxy-4-methyl-8-nitroquinoline derivatives



0.43 mmol (1.1 eq.) of the corresponding phenol (**16**), 0.43 mmol (1.1 eq.) of powder KOH were stirred with 1.5 ml of anhydrous DMF under argon atmosphere at room temperature in a previously dried 5ml microwave glass tube until total dissolution was observed. Then, 0.39 mmol (1.0 eq.) of compound **10** were added with 2ml of anhydrous DMF. The vial was sealed and heated under microwave irradiation at 140° for 2 hours. The solvent was eliminated under reduced pressure and the residue was diluted in AcOEt and washed twice with a LiCl concentrated solution to ensure the total DMF removal. The organic layer was washed with an alkaline KOH solution (pH=14) to remove any excess or unreacted phenol substrate. Finally, the organic solution was washed with water, dried over anhydrous $MgSO_4$ and concentrated under reduce pressure to afford the corresponding 5-phenoxy-6-methoxy-4-methyl-8-nitroquinoline (**11**).

6-methoxy-4-methyl-8-nitro-5-(3,4,5-trimethoxyphenoxy)quinoline (**11{1}**)

Starting from 5-chloro-6-methoxy-4-methyl-8-nitroquinoline (**10**) and 3,4,5-trimethoxyphenol (**16{1}**), the product **11{1}** (**LMF228**) was obtained as a powdery brown solid with 84% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 8.72 (d, *J* = 4.3 Hz, 1H, C2-H), 8.46 (s, 1H, C7-H), 7.44 (dd, *J* = 4.4, 0.9 Hz, 1H, C3-H), 6.12 (s, 2H, C14-H, C14'-H), 3.87 (s, 3H, C10-H₃), 3.65 (d, *J* = 1.0 Hz, 6H, C15-H₃, C15'-H₃), 3.61 (d, *J* = 0.8 Hz, 3H, C16-H₃), 2.69 (d, *J* = 1.0 Hz, 3H, C9-H₃).

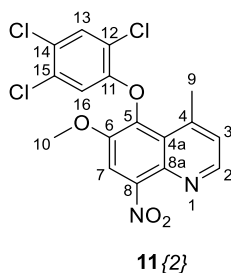
¹³C NMR (100 MHz, DMSO-*d*₆) δ 154.0 (C11), 153.6 (C13, C13'), 150.3 (C2), 148.6 (C6), 146.4 (C5*), 143.2 (C4), 138.4 (C8a*), 135.0 (C8*), 132.6 (C14), 125.7 (C3), 124.2 (C4a), 112.1 (C7), 92.7 (C12, C12'), 60.1 (C16), 57.3 (C10), 56.0 (C15, C15'), 22.5 (C9).

IR (ATR), ν_{max} (cm⁻¹) 2939 (*st* C-H), 2849, 1594, 1525 (*asym st* NO₂), 1524, 1499, 1442, 1323 (*sym st* NO₂), 1224 (*st* C_{ar}-N), 1176, 1123 (*st* C-O), 1011, 974, 763 (*oop b* C_{ar}-H).

MS (70 eV, EI) m/z (%): 401.2 (33%) [M+1]⁺, 400.2 (100%) [M]⁺, 385.2 (85%) [M-CH₂O]⁺, 218.1 (76%) [M+H-C₉H₁₁O₄(*phenolate*)]⁺, 184.1 (61%) [C₉H₁₂O₄(*phenol*)]⁺, 183.1 (63%) [C₉H₁₁O₄(*phenolate*)]⁺, 169.1 (74%), 141.1 (50%) [M+H-C₉H₁₁O₄(*phenolate*)-NO₂-OCH₃]⁺, 117.1 (48%), 69.1 (65%), 57.1 (49%).

HRMS (Q-TOF): m/z calculated for C₂₀H₂₀N₂O₇ [M]⁺: 400.1271. Found: 401.1331 [M+H]⁺

mp: 154-158 °C

6-methoxy-4-methyl-8-nitro-5-(2,4,5-trichlorophenoxy)quinoline (11{2})

Starting from 5-chloro-6-methoxy-4-methyl-8-nitroquinoline (**10**) and 2,4,5-trichlorophenol (**16{2}**), the product **11{2}** (**LMF234**) was obtained as a pale yellow pulverulent solid with 96% yield.

Spectroscopic data:

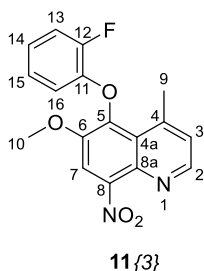
¹H NMR (400 MHz, DMSO-*d*₆) δ 8.74 (d, *J* = 4.1 Hz, 1H, C2-H), 8.50 (s, 1H, C7-H), 8.02 (s, 1H, C13-H), 7.48 (d, *J* = 4.2 Hz, 1H, C3-H), 7.05 (s, 1H, C16-H), 3.88 (s, 3H, C10-H₃), 2.64 (s, 3H, C9-H₃).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 152.1 (C11), 150.3 (C2), 147.9 (C6), 147.2 (C5*), 142.5 (C4), 137.0 (C8a*), 135.0 (C8*), 131.3 (C13), 131.0 (C15), 126.0 (C3), 125.0 (C14), 123.6 (C4a), 120.8 (C12), 116.4 (C16), 112.0 (C7), 57.5 (C10), 21.8 (C9).

IR (ATR), ν_{max} (cm⁻¹) 3101(*st* C_{ar}-H), 2996, 2944 (*st* C-H), 1529 (*asym st* NO₂), 1511, 1338 (*sym st* NO₂), 1254 (*st* C-N), 1055, 846, 777 (*oop b* C_{ar}-H).

HRMS (Q-TOF): *m/z* calculated for C₁₇H₁₁Cl₃N₂O₄ [M]⁺: 411.9784. Found: 412.9853 [M+H]⁺, 414.9826 [M+H+2]⁺, 416.9807 [M+H+4]⁺, 418.9789 [M+H+6]⁺,

mp: 147-160 °C

5-(2-fluorophenoxy)-6-methoxy-4-methyl-8-nitroquinoline (11{3})

Starting from 5-chloro-6-methoxy-4-methyl-8-nitroquinoline (**10**) and 2-fluorophenol (**16{3}**), the product **11{3}** (**LMF236**) was obtained as a brown pulverulent solid with 85% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 8.73 (d, *J* = 4.2 Hz, 1H, C2-H), 8.50 (s, 1H, C7-H), 7.47 (d, *J* = 4.3 Hz, 1H, C3-H), 7.38 (ddd, *J* = 11.6, 7.9, 1.9 Hz, 1H, C13-H), 7.15 – 6.89 (m, 2H, C14-H, C16-H), 6.63 (td, *J* = 8.6, 1.8 Hz, 1H, C15-H), 3.85 (s, 3H, C10-H₃), 2.67 (s, 3H, C9-H₃).

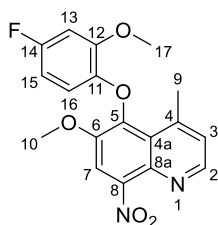
¹³C NMR (100 MHz, DMSO-*d*₆) δ 151.1 (d, *J* = 244.4 Hz, C12), 150.4 (C2), 148.3 (C6), 146.7 (C5*), 145.1 (d, *J* = 10.4 Hz, C11), 142.8 (C4), 138.0 (C8a*), 134.9 (C8), 125.9 (C3), 125.1 (d, *J* = 3.7 Hz, C16), 123.9 (C4a), 123.1 (d, *J* = 6.9 Hz, C14), 116.8 (d, *J* = 17.3 Hz, C13), 115.8 (C15), 111.9 (C7), 57.4 (C10), 21.9 (C9).

IR (ATR), ν_{max} (cm⁻¹) 3067(*st* C_{ar}-H), 2922 (*st* C-H), 2852, 1528 (*asym st* NO₂), 1500, 1352 (*sym st* NO₂), 1312, 1265 (*st* C_{ar}-N), 1246, 1194 (*st* C-O), 1170, 1105, 1078, 1030, 919, 848, 741 (*oop b* C_{ar}-H).

MS (70 eV, EI) m/z (%): 329.1 (34%) [M+1]⁺, 328.1 (100%) [M]⁺, 298.1 (34%) [M-CH₂O]⁺, 172.1 (16%), 144.1 (22%), 116.1 (21%), 109.1 (30%), 89.1 (22%), 75.1 (19%)

HRMS (Q-TOF): m/z calculated for C₁₇H₁₃FN₂O₄ [M]⁺: 328.0859. Found: 329.0921 [M+H]⁺

mp: 132-137 °C

5-(4-fluoro-2-methoxyphenoxy)-6-methoxy-4-methyl-8-nitroquinoline (11{4})

11{4}

Starting from 5-chloro-6-methoxy-4-methyl-8-nitroquinoline (**10**) and 4-fluoro-2-methoxyphenol (**16{4}**), the product **11{4}** (**LMF238**) was obtained as a brown greyish pulverulent solid with 91% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 8.71 (d, *J* = 4.3 Hz, 1H, C2-H), 8.44 (s, 1H, C7-H), 7.43 (d, *J* = 4.4 Hz, 1H, C3-H), 7.07 (dd, *J* = 10.5, 2.9 Hz, 1H, C13-H), 6.54 (td, *J* = 8.6, 2.9 Hz, 1H, C15-H), 6.41 (dd, *J* = 9.0, 5.5 Hz, 1H, C16-H), 3.88 (s, 3H, C10-H₃), 3.82 (s, 3H, C17-H₃), 2.65 (s, C9-H₃).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 157.5 (d, *J* = 237.7 Hz, C14), 150.3 (C2), 149.2 (d, *J* = 10.3 Hz, C12), 148.2 (C6), 146.3 (C5*), 143.2 (C11), 143.1 (C4), 139.0 (C8a*), 135.0 (C8*), 125.6 (C3), 123.9 (C4a),

114.1 (d, $J = 10.2$ Hz, C16), 112.1 (C7), 105.9 (d, $J = 23.1$ Hz, C15), 101.3 (d, $J = 27.6$ Hz, C13), 57.3 (C17), 56.3 (C10), 21.7 (C9).

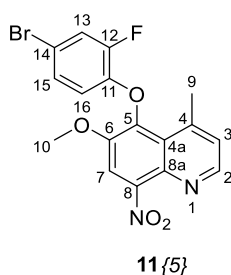
IR (ATR), ν_{\max} (cm^{-1}) 2918 (st C-H), 2849, 1622, 1529 (asym st NO_2), 1505, 1447, 1308 (sym st NO_2), 1256 (st $\text{C}_{\text{ar}}-\text{N}$), 1180 (st C-O), 1154, 1106, 1078, 1027, 950, 833 (oop b $\text{C}_{\text{ar}}-\text{H}$), 768.

MS (70 eV, EI) m/z (%): 359.1 (35%) $[\text{M}+1]^+$, 358.1 (100%) $[\text{M}]^+$, 328.1 (24%) $[\text{M}-\text{CH}_2\text{O}]^+$, 142.1 (44%), 127.1 (57%), 109.1 (29%), 99.1 (40%).

HRMS (Q-TOF): m/z calculated for $\text{C}_{18}\text{H}_{15}\text{FN}_2\text{O}_5$ $[\text{M}]^+$: 358.0965. Found: 359.1026 $[\text{M}+\text{H}]^+$

mp: 153-161 °C

5-(4-bromo-2-fluorophenoxy)-6-methoxy-4-methyl-8-nitroquinoline (**11{5}**)



Starting from 5-chloro-6-methoxy-4-methyl-8-nitroquinoline (**10**) and 4-bromo-2-fluorophenol (**16{5}**), the product **11{5}** (**LMF240**) was obtained as a dark brown pulverulent solid with 75% yield.

Spectroscopic data:

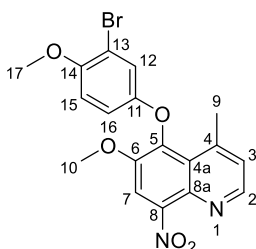
^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ 8.74 (d, $J = 4.3$ Hz, 1H, C2-H), 8.50 (s, 1H, C7-H), 7.74 (dd, $J = 10.7, 2.2$ Hz, 1H, C13-H), 7.48 (dd, $J = 4.2, 0.8$ Hz, 1H, C3-H), 7.20 (dt, $J = 8.9, 2.0$ Hz, 1H, C16-H), 6.66 (dd, $J = 9.0, 1.8$ Hz, 1H, C15-H), 3.86 (s, 3H, C10- H_3), 2.66 (d, $J = 0.8$ Hz, 3H, C9- H_3).

^{13}C NMR (100 MHz, $\text{DMSO}-d_6$) δ 152.3 (d, $J = 228.3$ Hz, C12), 150.4 (C2), 148.1 (C6), 146.9 (C5*), 142.6 (C4), 137.5 (C8a*), 134.8 (C8*), 127.9 (d, $J = 3.7$ Hz, C16), 125.9 (C3), 123.7 (C4a), 123.7 (d, $J = 4.2$ Hz, C14), 120.1 (d, $J = 21.0$ Hz, C13), 117.6 (C15), 113.2 (d, $J = 14.8$ Hz, C11), 111.9 (C7), 57.4 (C10), 21.8 (C9).

IR (ATR), ν_{\max} (cm^{-1}) 3451, 3325, 3094 (st $\text{C}_{\text{ar}}-\text{H}$), 2953 (st C-H), 2830, 1646, 1592, 1510 (asym st NO_2), 1495, 1447, 1423, 1395, 1316 (sym st NO_2), 1231 (st $\text{C}_{\text{ar}}-\text{N}$), 1093, 995, 848 (oop b $\text{C}_{\text{ar}}-\text{H}$), 780, 756.

HRMS (Q-TOF): m/z calculated for $\text{C}_{17}\text{H}_{12}\text{BrFN}_2\text{O}_4$ $[\text{M}]^+$: 405.9965. Found: 407.0032 $[\text{M}+\text{H}]^+$, 409.0013 $[\text{M}+\text{H}+2]^+$

mp: 141-145 °C

5-(3-bromo-4-methoxyphenoxy)-6-methoxy-4-methyl-8-nitroquinoline (11{6})**11{6}**

Starting from 5-chloro-6-methoxy-4-methyl-8-nitroquinoline (**10**) and 3-bromo-4-methoxyphenol (**16{6}**), the product **11{6}** (**LMF301**) was obtained as a brown pulverulent solid with 82% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 8.72 (d, *J* = 4.3 Hz, 1H, C2-H), 8.46 (s, 1H, C7-H), 7.45 (dd, *J* = 4.4, 1.0 Hz, 1H, C3-H), 7.15 (d, *J* = 3.0 Hz, 1H, C12-H), 7.02 (d, *J* = 9.1 Hz, 1H, C15-H), 6.78 (dd, *J* = 9.0, 3.0 Hz, 1H, C16-H), 3.85 (s, 3H, C10-H₃), 3.79 (s, 3H, C17-H₃), 2.66 (d, *J* = 0.8 Hz, 3H, C9-H₃).

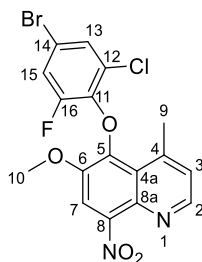
¹³C NMR (100 MHz, DMSO-*d*₆) δ 151.6 (C11), 150.8 (C14), 150.4 (C2), 148.5 (C6), 146.5 (C5*), 142.9 (C4), 138.5 (C8*), 134.9 (C8a*), 125.8 (C3), 124.0 (C4a), 119.4 (C12), 114.6 (C16), 113.4 (C15), 112.0 (C7), 111.1 (C13), 57.3 (C10), 56.5 (C17), 22.4 (C9).

IR (ATR), ν_{max} (cm⁻¹) 3077 (*st* C_{ar}-H), 2935 (*st* C-H), 1738, 1526 (*asym st* NO₂), 1486, 1339 (*sym st* NO₂), 1253 (*st* C-N), 1200 (*st* C-O), 1050, 1034, 869, 749 (*oop b* C_{ar}-H).

MS (70 eV, EI) m/z (%): 420.0 (96%) [M+2]⁺, 419.1 (34%) [M+1]⁺, 418.0 (100%) [M]⁺, 254.0 (19%), 252.0 (56%), 202.0 (30%), 116.0 (36%), 63.0 (34%).

HRMS (Q-TOF): m/z calculated for C₁₈H₁₅BrN₂O₅ [M]⁺: 418.0164. Found: 419.0231 [M+H]⁺, 421.0212 [M+H+2]⁺

mp: 202-211 °C

5-(4-bromo-2-chloro-6-fluorophenoxy)-6-methoxy-4-methyl-8-nitroquinoline (11{7})**11{7}**

Starting from 5-chloro-6-methoxy-4-methyl-8-nitroquinoline (**10**) and 4-bromo-2-chloro-6-fluorophenol (**16{7}**), the residue was purified by automated flash chromatography (silica column, Cy:AcOEt 1:1) to obtain product **11{7}** (**LMF302**) as an orange-colored thick slurry with 28% yield.

Spectroscopic data:

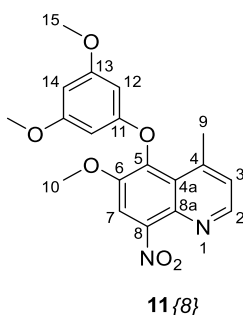
¹H NMR (400 MHz, DMSO-*d*₆) δ 8.75 (d, *J* = 4.3 Hz, 1H, C2-H), 8.40 (s, 1H, C7-H), 7.75 (dd, *J* = 2.3, 1.9 Hz, 1H, C13-H), 7.68 (dd, *J* = 11.1, 2.3 Hz, 1H, C15-H), 7.54 (dd, *J* = 4.4, 1.0 Hz, 1H, C3-H), 3.75 (s, 3H, C10-H₃), 2.81 (d, *J* = 0.8 Hz, 3H, C9-H₃).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 153.1 (d, *J* = 252.2 Hz, C16), 150.6 (C2), 146.0 (C6), 145.7 (C5*), 143.4 (C4), 141.4 (C8a*), 140.6 (d, *J* = 11.6 Hz, C11), 134.7 (C8*), 128.5 (C13), 127.6 (d, *J* = 3.1 Hz, C12), 126.1 (d, *J* = 3.6 Hz, C14), 125.9 (C3), 122.3 (C4a), 119.5 (d, *J* = 22.5 Hz, C15), 112.1 (C7), 57.8 (C10), 22.9 (C9).

IR (ATR), ν_{max} (cm⁻¹) 3087 (*st* C_{ar}-H), 2930 (*st* C-H), 2852, 1530 (*asym st* NO₂), 1465, 1440, 1309 (*sym st* NO₂), 1260 (*st* C-N), 932, 918, 844, 746 (*oop b* C_{ar}-H).

MS (70 eV, EI) m/z (%): 444.0 (26%) [M+4]⁺, 442.0 (100%) [M+2]⁺, 440.0 (80%) [M]⁺, 410.0 (29%) [M-CH₂O]⁺, 261.0 (42%), 226.0 (39%), 220.0 (24%), 116 (34%).

HRMS (Q-TOF): m/z calculated for C₁₇H₁₁BrClFN₂O₄ [M]⁺: 439.9575. Found: 440.9638 [M+H]⁺, 442.9615 [M+H+2]⁺, 444.9602 [M+H+4]⁺

5-(3,5-dimethoxyphenoxy)-6-methoxy-4-methyl-8-nitroquinoline (**11{8}**)

Starting from 5-chloro-6-methoxy-4-methyl-8-nitroquinoline (**10**) and 3,5-dimethoxyphenol (**16{8}**), the product **11{8}** (**LMF230**) was obtained as a metallic black solid with 89% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 8.71 (d, *J* = 4.3 Hz, 1H, C2-H), 8.46 (s, 1H, C7-H), 7.43 (d, *J* = 4.4 Hz, 1H, C3-H), 6.21 (s, 1H, C14-H), 5.93 (s, 2H, C12-H, C12'-H), 3.87 (s, 3H, C10-H₃), 3.68 (s, 6H, C15-H₃, C15'-H₃), 2.66 (s, 3H, C9-H₃).

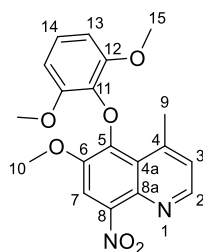
¹³C NMR (100 MHz, DMSO-*d*₆) δ 161.4 (C13, C13'), 159.4 (C11), 150.3 (C2), 148.6 (C6), 146.5 (C5*), 142.9 (C4), 138.1 (C8a*), 134.8 (C8*), 125.7 (C3), 124.0 (C4a), 111.9 (C7), 94.0 (C14), 93.8 (C12, C12'), 57.3 (C10), 55.3 (C15, C15'), 22.3 (C9).

IR (ATR), ν_{max} (cm⁻¹) 3383, 2929 (*st* C-H), 2849, 1589 (*asym st* NO₂), 1531 (*st* C_{ar}-C_{ar}), 1438, 1307 (*sym st* NO₂), 1256 (*st* C_{ar}-N), 1203, 1150, 1128 (*st* C-O), 1051, 1033, 821 (*oop b* C_{ar}-H).

MS (70 eV, EI) m/z (%): 371.1 (50%) [M+1]⁺, 370.1 (100%) [M]⁺, 339.1 (60%) [M-OCH₃]⁺, 293.1 (30%), 151.1 (29%).

HRMS (Q-TOF): m/z calculated for C₁₉H₁₈N₂O₆ [M]⁺: 370.1165. Found: 371.1227 [M+H]⁺

mp: 96-105 °C

5-(2,6-dimethoxyphenoxy)-6-methoxy-4-methyl-8-nitroquinoline (11{9})**11{9}**

Starting from 5-chloro-6-methoxy-4-methyl-8-nitroquinoline (**10**) and 2,6-dimethoxyphenol (**16{9}**), the product **11{9}** (**LMF231**) was obtained as an other crystalline solid with 83% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 8.70 (d, *J* = 4.3 Hz, 1H, C2-H), 8.22 (s, 1H, C7-H), 7.44 (d, *J* = 4.4 Hz, 1H, C3-H), 7.04 (t, *J* = 8.4 Hz, 1H, C14-H), 6.72 (d, *J* = 8.4 Hz, 2H, C13-H, C13'-H), 3.58 (s, 6H, C15-H₃, C15'-H₃), 3.56 (s, 3H, C10-H₃), 2.82 (s, 3H, C9-H₃).

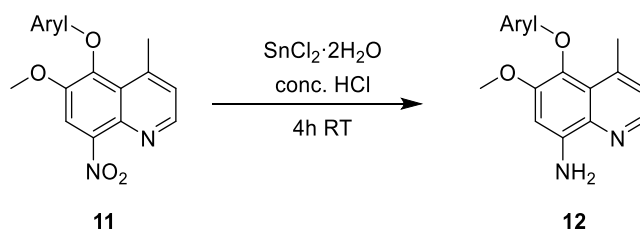
¹³C NMR (100 MHz, DMSO-*d*₆) δ 151.0 (C12, C12'), 150.2 (C2), 145.4 (C6), 145.4 (C5*), 144.8 (C4), 144.0 (C8a*), 135.3 (C11), 135.3 (C8*), 125.0 (C3), 123.6 (C14), 122.7 (C4a), 113.4 (C7), 105.9 (C13, C13'), 57.9 (C10), 56.1 (C15, C15'), 23.3 (C9).

IR (ATR), ν_{max} (cm⁻¹) 3057 (*st* C_{ar}-H), 2942, 2840 (*st* C-H), 1530 (*asym st* NO₂), 1478, 1308 (*sym st* NO₂), 1254 (*st* C-N), 1108 (*st* C-O), 1033, 765 (*oop b* C_{ar}-H), 736.

HRMS (Q-TOF): *m/z* calculated for C₁₉H₁₈N₂O₆ [M]⁺: 370.1165. Found: 371.1225 [M+H]⁺,

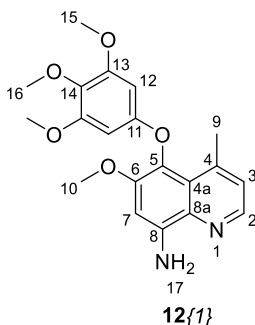
mp: 156-164 °C

5.2.2.7. Synthesis of 5-phenoxy-6-methoxy-4-methylquinoline-8-amine derivatives



1.00 mmol (1.0 eq.) of the corresponding 5-phenoxy-6-methoxy-4-methyl-8-nitroquinoline derivative (**11**) were placed in a 50 ml round-bottomed flask with 5.00 mmol (5.0 eq.) of $\text{SnCl}_2 \cdot 2\text{H}_2\text{O}$ along with a stirring bar. 10 ml of concentrated hydrochloric acid (37%) were added dropwise until total solution was observed (sonication was facilitated in some of the cases) and then the reaction was stirred for 4 hours at room temperature. Later, the mixture was slowly basified with a 2M NaOH solution in an iced-water bath to mitigate the exothermic reaction until pH = 14. A clear color variation was observed during alkalization. The crude was later extracted with three fractions of chloroform. The organic layers were unified, dried with MgSO_4 and concentrated under reduce pressure to afford the corresponding 5-phenoxy-6-methoxy-4-methylquinoline-8-amine derivatives (**12**).

6-methoxy-4-methyl-5-(3,4,5-trimethoxyphenoxy)quinolin-8-amine (**12{1}**)



Starting from 6-methoxy-4-methyl-8-nitro-5-(3,4,5-trimethoxyphenoxy)quinoline **11{1}**, the product **12{1}** (**LMF257**) was obtained as a dark bronze solid with 98% yield.

Spectroscopic data:

^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ 8.41 (d, $J = 4.2$ Hz, 1H, C2-H), 7.17 (d, $J = 4.3$ Hz, 1H, C3-H), 6.87 (s, 1H, C7-H), 6.03 (s, 2H, N17-H₂), 5.99 (s, 2H, C12-H, C12'-H), 3.75 (s, 3H, C10-H₃), 3.60 (s, 6H, C15-H₃, C15'-H₃), 3.58 (s, 3H, C16-H₃), 2.56 (s, 3H, C9-H₃).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 155.8 (C11), 153.4 (C13, C13'), 150.8 (C6), 144.6 (C2), 144.5 (C5*), 141.5 (C4), 132.6 (C8*), 131.8 (C14), 124.8 (C8a*), 124.6 (C3), 123.9 (C4a*), 96.3 (C7), 92.1 (C12, C12'), 60.0 (C16), 56.0 (C10), 55.8 (C15, C15'), 22.6 (C9).

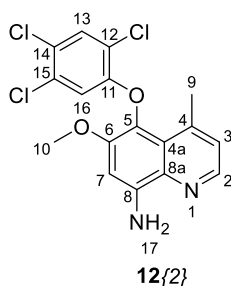
IR (ATR), ν_{max} (cm⁻¹) 3467 (*st* N-H), 3354 (*st* N-H), 2928 (*st* C-H), 1726 (*summation bands*), 1613, 1594 (*b* N-H), 1518 (*sc* N-H), 1499, 1461, 1444, 1440, 1219 (*st* C_{ar}-N), 1126 (*st* C-O), 1044, 1009, 987, 805 (*oop b* C_{ar}-H).

MS (70 eV, EI) m/z (%): 371.1 (6%) [M+1]⁺, 370.1 (27%) [M]⁺, 340.1 (9%) [M-OCH₃]⁺, 203.1 (22%) [M-C₉H₁₂O₃(*aryl*)]⁺, 188.1 (49%) [M-C₉H₁₁O₄(*phenolate*)]⁺, 184.1 (71%) [C₉H₁₁O₄(*phenolate*)]⁺, 169.1 (100%) [C₉H₁₂O₃(*aryl*)]⁺, 159.1 (29%), 141.1 (47%), 111.0 (22%).

HRMS (Q-TOF): m/z calculated for C₂₀H₂₂N₂O₅ [M]⁺: 370.1529. Found: 371.1589 [M+H]⁺

mp: 162-164 °C

6-methoxy-4-methyl-5-(2,4,5-trichlorophenoxy)quinolin-8-amine (**12{2}**)



Starting from 6-methoxy-4-methyl-8-nitro-5-(2,4,5-trichlorophenoxy)quinoline **11{2}**, the product **12{2}** (**LMF263**) was obtained as a yellow solid with 79% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 8.43 (d, *J* = 4.2 Hz, 1H, C2-H), 7.94 (s, 1H, C13-H), 7.21 (dd, *J* = 4.3, 1.0 Hz, 1H, C3-H), 6.88 (s, 1H, C7-H), 6.54 (s, 1H, C16-H), 6.18 (s, 2H, N17-H₂), 3.76 (s, 3H, C10-H₃), 2.51 (d, *J* = 0.9 Hz, 3H, C9-H₃).

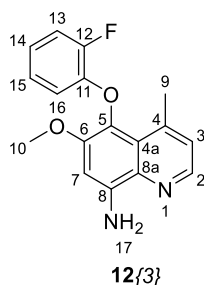
¹³C NMR (100 MHz, DMSO-*d*₆) δ 154.0 (C11), 150.2 (C6), 145.7 (C5*), 144.7 (C2), 140.7 (C4), 132.5 (C8*), 131.2 (C13), 130.6 (C15), 125.0 (C3), 123.5 (C8a*), 123.4 (C14), 123.1 (C4a), 120.4 (C12), 115.4 (C16), 95.7 (C7), 56.0 (C10), 21.7 (C9).

IR (ATR), ν_{max} (cm⁻¹) 3470 (*st* N-H), 3358 (*st* N-H), 3094 (*st* C_{ar}-H), 2928 (*st* C-H), 1615, 1579 (*b* N-H), 1518 (*sc* N-H), 1471, 1443, 1399, 1350, 1224 (*st* C_{ar}-N), 1071 (*st* C-O), 1038, 946, 830 (*w* N-H), 674 (*oop b* C_{ar}-H).

HRMS (Q-TOF): m/z calculated for C₁₇H₁₃Cl₃N₂O₂ [M]⁺: 382.0043. Found: 383.0109 [M+H]⁺, 385.0081 [M+H+2]⁺, 387.0061 [M+H+4]⁺, 389.0042 [M+H+6]⁺,

mp: 160-165 °C

5-(2-fluorophenoxy)-6-methoxy-4-methylquinolin-8-amine (12{3})



Starting from 5-(2-fluorophenoxy)-6-methoxy-4-methyl-8-nitroquinoline **11{3}**, the product **12{3}** (**LMF258**) was obtained as a black solid with quantitative yield (100%).

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 8.42 (d, *J* = 4.2 Hz, 1H, C2-H), 7.35 – 7.25 (m, 1H, C13-H), 7.18 (dd, *J* = 4.3, 1.0 Hz, 1H, C3-H), 6.95 (ddd, *J* = 6.8, 3.6, 1.6 Hz, 2H, C14-H, 16-H), 6.88 (s, 1H, C7-H), 6.45 – 6.36 (m, 1H, C15-H), 6.08 (s, 2H, N17-H₂), 3.74 (s, 3H, C10-H₃), 2.53 (d, *J* = 0.9 Hz, 3H, C9-H₃).

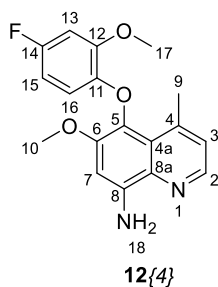
¹³C NMR (100 MHz, DMSO-*d*₆) δ 151.0 (d, *J* = 243.5 Hz, C12), 150.6 (C6), 146.9 (d, *J* = 10.4 Hz, C11), 145.1 (C5*), 144.6 (C2), 141.1 (C4), 132.6 (C8*), 124.8 (d, *J* = 3.7 Hz, C16), 124.7 (C3), 124.1 (C8a*), 123.6 (C4a), 121.5 (d, *J* = 6.7 Hz, C14), 116.4 (d, *J* = 17.6 Hz, C13), 115.1 (C15), 96.1 (C7), 56.0 (C10), 21.8 (C9).

IR (ATR), ν_{max} (cm⁻¹) 3445 (*st* N-H), 3329 (*st* N-H), 2928 (*st* C-H), 1721 (*summation bands*), 1614 (*b* N-H), 1586, 1521 (*sc* N-H), 1497, 1449, 1398, 1254 (*st* C_{ar}-N), 1228 (*st* C_{ar}-F), 1199, 1141, 1102, 1037 (*st* C-O), 941, 829, 750 (*oop b* C_{ar}-H).

MS (70 eV, EI) m/z (%): 299.1 (17%) [M+1]⁺, 298.1 (70%) [M]⁺, 203.1 (45%) [M-C₆H₄F(*aryl*)]⁺, 188.1 (49%) [M-C₆H₄FO(*phenolate*)]⁺, 159.1 (60%), 158.1 (30%), 145.1 (31%), 112.1 (78%) [C₆H₄FO(*phenol*)]⁺, 64.1 (30%).

HRMS (Q-TOF): m/z calculated for C₁₇H₁₅FN₂O₂ [M]⁺: 298.1118. Found: 299.1182 [M+H]⁺

mp: 129-130 °C

5-(4-fluoro-2-methoxyphenoxy)-6-methoxy-4-methylquinolin-8-amine (12{4})

Starting from 5-(4-fluoro-2-methoxyphenoxy)-6-methoxy-4-methyl-8-nitroquinoline **11{4}**, the product **12{4}** (LMF253) was obtained as a black solid with 79% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 8.40 (d, *J* = 4.2 Hz, 1H, C2-H), 7.15 (dd, *J* = 4.3, 0.9 Hz, 1H, C3-H), 6.99 (dd, *J* = 10.6, 2.9 Hz, 1H, C13-H), 6.86 (s, 1H, C7-H), 6.49 (ddd, *J* = 8.9, 8.3, 3.0 Hz, 1H, C15-H), 6.16 (dd, *J* = 8.9, 5.6 Hz, 1H, C16-H), 6.02 (s, 2H, N18-H₂), 3.88 (s, 3H, C17-H₃), 3.72 (s, 3H, C10-H₃), 2.50 (s, C9-H₃).

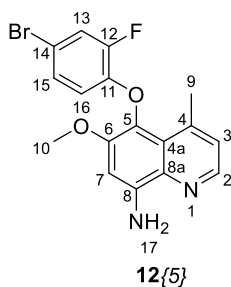
¹³C NMR (100 MHz, DMSO-*d*₆) δ 156.7 (d, *J* = 236.2 Hz, C14), 150.5 (C6), 148.9 (d, *J* = 10.3 Hz, C12), 145.0 (C11), 144.7 (C5*), 144.5 (C2), 141.4 (C4), 132.7 (C8*), 124.8 (C8a*), 124.5 (C3), 123.8 (C4a), 112.9 (d, *J* = 10.2 Hz, C16), 105.6 (d, *J* = 22.7 Hz, C15), 101.0 (d, *J* = 27.4 Hz, C13), 96.4 (C7), 56.1 (C17), 55.9 (C10), 21.6 (C9).

IR (ATR), ν_{max} (cm⁻¹) 3366 (*st* N-H), 2965, 2930 (*st* C-H), 2858, 1720 (*summation bands*), 1613 (*b* N-H), 1583 (*sc* N-H), 1500, 1446, 1202, 1187 (*st* C-O), 1144, 1106, 1029, 936, 831 (*oop b* C_{ar}-H).

MS (70 eV, EI) m/z (%): 329.1 (5%) [M+1]⁺, 328.1 (23%) [M]⁺, 340.1 (9%) [M-OCH₃]⁺, 310.1 (39%) [M-C₉H₁₂O_{3(aryl)}]⁺, 203.1 (20%), 188.1 (100%) [M-C₇H₆FO_{2(phenolate)}]⁺, 174.2 (21%), 159.1 (62%), 158.1 (32%), 145.1 (32%), 142.1 (76%) [C₇H₇FO_{2(phenol)}]⁺, 128.0 (32%), 127.0 (76%) [C₇H₇FO(aryl)]⁺, 99.0 (55%).

HRMS (Q-TOF): m/z calculated for C₁₈H₁₇FN₂O₃ [M]⁺: 328.1223. Found: 329.1286 [M+H]⁺

mp: 100-110 °C

5-(4-bromo-2-fluorophenoxy)-6-methoxy-4-methylquinolin-8-amine (**12{5}**)

Starting from 5-(4-bromo-2-fluorophenoxy)-6-methoxy-4-methyl-8-nitroquinoline **11{5}**, the product **12{5}** (**LMF262**) was obtained as a black solid with 94% yield.

Spectroscopic data:

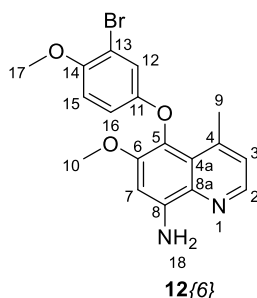
¹H NMR (400 MHz, DMSO-*d*₆) δ 8.42 (d, *J* = 4.2 Hz, 1H, C2-H), 7.64 (dd, *J* = 10.8, 2.4 Hz, 1H, C13-H), 7.20 (dd, *J* = 4.3, 1.0 Hz, 1H, C3-H), 7.16 (ddd, *J* = 8.8, 2.4, 1.5 Hz, 1H, C16-H), 6.87 (s, 1H, C7-H), 6.40 (t, *J* = 9.0 Hz, 1H, C15-H), 6.12 (s, 2H, N17-H₂), 3.75 (s, 3H, C10-H₃), 2.52 (d, *J* = 1.1 Hz, 3H, C9-H₃).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 151.0 (d, *J* = 249.1 Hz, C12), 150.4 (C6), 146.7 (d, *J* = 22.9 Hz, C11), 145.3 (C5), 144.6 (C2), 140.9 (C4), 132.5 (C8), 127.8 (d, *J* = 3.9 Hz, C16), 124.8 (C3), 123.8 (C8a), 123.4 (C4a), 119.7 (d, *J* = 21.0 Hz, C13), 116.9 (C15), 111.6 (d, *J* = 8.4 Hz, C14), 95.9 (C7), 56.0 (C10), 21.8 (C9).

IR (ATR), ν_{max} (cm⁻¹) 3329 (*st* N-H), 2940 (*st* C-H), 1614, 1587 (*b* N-H), 1519 (*sc* N-H), 1490, 1447, 1395, 1262 (*st* C_{ar}-N), 1194 (*st* C-O), 1142, 1037, 941, 831 (*w* N-H), 805 (*oop b* C_{ar}-H).

HRMS (Q-TOF): *m/z* calculated for C₁₇H₁₄BrFN₂O₂ [M]⁺: 376.0223. Found: 377.0291 [M+H]⁺, 379.0272 [M+H+2]⁺,

mp: 88-92 °C

5-(3-bromo-4-methoxyphenoxy)-6-methoxy-4-methylquinolin-8-amine (12{6})

Starting from 5-(3-bromo-4-methoxyphenoxy)-6-methoxy-4-methyl-8-nitroquinoline **11{6}**, the product **12{6}** (**LMF303**) was obtained as a black-greenish solid with 95% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 8.41 (d, *J* = 4.2 Hz, 1H, C2-H), 7.17 (dd, *J* = 4.3, 1.0 Hz, 1H, C3-H), 7.00 (d, *J* = 9.1 Hz, 1H, C15-H), 6.90 (d, *J* = 3.0 Hz, 1H, C12-H), 6.87 (s, 1H, C7-H), 6.67 (dd, *J* = 9.0, 3.0 Hz, 1H, C16-H), 6.05 (s, 2H, N18-H₂), 3.76 (s, 3H, C17-H₃), 3.73 (s, 3H, C10-H₃), 2.53 (d, *J* = 0.7 Hz, 3H, C9-H₃).

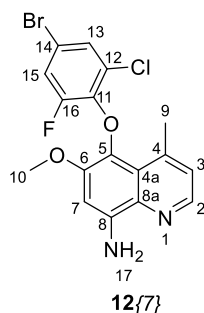
¹³C NMR (100 MHz, DMSO-*d*₆) δ 153.5 (C11), 150.6 (C6), 150.0 (C14), 144.8 (C5), 144.6 (C2), 141.2 (C4), 132.6 (C8*), 124.9 (C8a*), 124.7 (C3), 123.7 (C4a), 118.7 (C12), 114.3 (C16), 113.5 (C15), 110.8 (C13), 96.2 (C7), 56.5 (C17), 56.0 (C10), 22.5 (C9).

IR (ATR), ν_{max} (cm⁻¹) 3489 (*st* N-H), 3379 (*st* N-H), 2935 (*st* C-H), 1614, 1580 (*b* N-H), 1517 (*sc* N-H), 1487, 1446, 1397, 1265 (*st* C_{ar}-N), 1194 (*st* C-O), 1043, 1031, 939, 831 (*w* N-H), 752 (*oop b* C_{ar}-H).

MS (70 eV, EI) m/z (%): 390.1 (50%) [M+2]⁺, 389.1 (9%) (50%) [M+1]⁺, 388.1 (52%) [M]⁺, 204.0 (41%) [C₇H₇BrO₂(*phenol*)+2]⁺, 203.0 (100%) [C₇H₇BrO₂(*phenol*)+1]⁺, 202.0 (35%) [C₇H₇BrO₂(*phenol*)]⁺, 188.9 (33%), 188.1 (82%) [M-C₇H₆BrO₂(*phenolate*)]⁺, 187.1 (39%), 177.1 (42%), 15591.1 (56%), 83.0 (33%).

HRMS (Q-TOF): m/z calculated for C₁₈H₁₇BrN₂O₃ [M]⁺: 388.0423. Found: 389.0486 [M+H]⁺, 391.0467 [M+H+2]⁺

mp: 145-149 °C

5-(4-bromo-2-chloro-6-fluorophenoxy)-6-methoxy-4-methylquinolin-8-amine (**12{7}**)

Starting from 5-(4-bromo-2-chloro-6-fluorophenoxy)-6-methoxy-4-methyl-8-nitroquinoline **11{7}**, the product **12{7}** (**LMF304**) was obtained as a black-greenish solid with 74% yield.

Spectroscopic data:

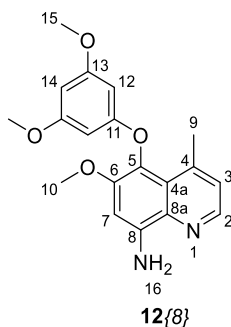
¹H NMR (400 MHz, DMSO-*d*₆) δ 8.44 (d, *J* = 4.2 Hz, 1H, C2-H), 7.65 (dd, *J* = 2.3, 1.9 Hz, 1H, C13-H), 7.53 (dd, *J* = 11.6, 2.4 Hz, 1H, C15-H), 7.26 (dd, *J* = 4.3, 1.0 Hz, 1H, C3-H), 6.74 (s, 1H, C7-H), 5.96 (s, 2H, N17-H₂), 3.63 (s, 3H, C10-H₃), 2.68 (d, *J* = 0.7 Hz, 3H, C9-H₃).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 153.11 (d, *J* = 252.5 Hz, C16), 148.8 (C6), 144.8 (C2), 144.2 (C5), 142.2 (d, *J* = 9.8 Hz, C11), 141.7 (C4), 132.3 (C8), 129.3 (C8a), 128.3 (d, *J* = 3.4 Hz, C13), 126.0 (d, *J* = 5.0 Hz, C12), 124.7 (C3), 122.6 (C4a), 119.5 (d, *J* = 22.7 Hz, C15), 112.8 (d, *J* = 10.7 Hz, C14), 96.0 (C7), 56.4 (C10), 22.7 (C9).

IR (ATR), ν_{max} (cm⁻¹) 3489 (*st* N-H), 3376 (*st* N-H), 2922 (*st* C-H), 1616, 1580 (*b* N-H), 1518 (*sc* N-H), 1487, 1468, 1443, 1398, 1264 (*st* C_{ar}-N), 1223, 1195, 1180 (*st* C-O), 1135, 1031, 940, 844 (*w* N-H), 830 (*oop b* C_{ar}-H).

HRMS (Q-TOF): *m/z* calculated for C₁₇H₁₃BrClFNO₂ [M]⁺: 409.9833. Found: 410.9896 [M+H]⁺, 412.9876 [M+H+2]⁺, 414.9858 [M+H+4]⁺,

mp: 187-193 °C

5-(3,5-dimethoxyphenoxy)-6-methoxy-4-methylquinolin-8-amine (12{8})

Starting from 5-(3,5-dimethoxyphenoxy)-6-methoxy-4-methyl-8-nitroquinoline **11{8}**, the product **12{8}** (**LMF265**) was obtained as a bronzed-colored solid film with 57% yield.

Spectroscopic data:

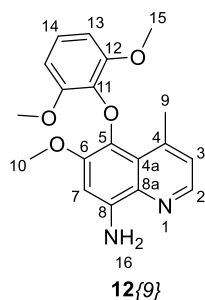
¹H NMR (400 MHz, DMSO-*d*₆) δ 8.40 (d, *J* = 4.2 Hz, 1H, C2-H), 7.16 (dd, *J* = 4.3, 1.0 Hz, 1H, C3-H), 6.86 (s, 1H, C7-H), 6.12 (t, *J* = 2.2 Hz, 1H, C14-H), 6.03 (s, 2H, N16-H₂), 5.82 (d, *J* = 2.2 Hz, 2H, C12-H, C12'-H), 3.74 (s, 3H, C10-H₃), 3.64 (s, 6H, C15-H₃, C15'-H₃), 2.53 (d, *J* = 0.7 Hz, 3H, C9-H₃).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 161.4 (C11), 161.2 (C13, C13'), 150.6 (C6), 144.7 (C5), 144.5 (C2), 141.3 (C4), 132.5 (C8), 124.8 (C8a), 124.6 (C3), 123.8 (C4a), 96.2 (C7), 93.5 (C12, C12'), 92.9 (C14), 56.0 (C15, C15'), 55.1 (C10), 22.4 (C9).

IR (ATR), ν_{max} (cm⁻¹) 3469 (*st* N-H), 3363 (*st* N-H), 2931 (*st* C-H), 2835, 1591 (*b* N-H), 1516 (*sc* N-H), 1441, 1399, 1208, 1146 (*st* C-O), 1123, 1053, 816 (*oop b* C_{ar}-H).

HRMS (Q-TOF): *m/z* calculated for C₁₉H₂₀N₂O₄ [M]⁺: 340.1423. Found: 341.1476 [M+H]⁺,

mp: 131-140 °C

5-(2,6-dimethoxyphenoxy)-6-methoxy-4-methylquinolin-8-amine (**12{9}**)

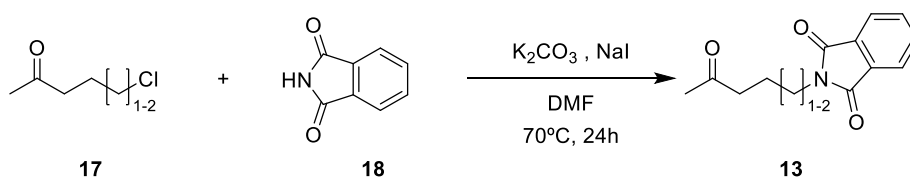
Starting from 5-(2,6-dimethoxyphenoxy)-6-methoxy-4-methyl-8-nitroquinoline **11{9}**, the product **12{9}** (**LMF264**) was obtained as an ochre solid with 13% yield.

Spectroscopic data:

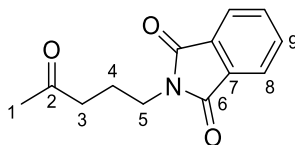
¹H NMR (400 MHz, DMSO-*d*₆) δ 8.41 (d, *J* = 4.2 Hz, 1H, C2-H), 7.18 (dd, *J* = 4.3, 0.9 Hz, 1H, C3-H), 6.92 (dd, *J* = 8.6, 8.0 Hz, 1H, C14-H), 6.65 (d, *J* = 8.3 Hz, 2H, C13-H, C13-H'), 6.65 (s, 1H, C7), 5.67 (s, 2H, N16-H₂), 3.50 (s, 6H, C15-H₃, C15'-H₃), 3.46 (s, 3H, C10-H₃), 2.72 (d, *J* = 0.6 Hz, 3H, C9-H₃).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 151.4 (C12, C12'), 148.1 (C6), 144.6 (C2), 143.4 (C4), 142.1 (C5), 137.4 (C8a), 133.0 (C8), 132.8 (C11), 123.9 (C3), 123.4 (C4a), 122.1 (C14), 106.6 (C13, C13'), 98.2 (C7), 56.8 (C10), 56.2 (C15, C15'), 23.2 (C9).

5.2.2.8. Synthesis of 2-(oxoalkyl)isoindoline-1,3-dione derivatives



3.71 mmol (1.0 eq.) of the corresponding oxoalkyl chloride derivative (**17**), 4.09 mmol (1.1 eq.) of phthalimide, 5.57 mmol (1.5 eq.) of K₂CO₃ and 3.71 mmol (1.0 eq.) of NaI were placed in a previously dried 10 ml round-bottomed flask with 5 ml of anhydrous DMF along with a stirring bar. The mixture was stirred at 70 °C for 24 hours under argon atmosphere. The solvent was later eliminated under reduced pressure and the residue was diluted in AcOEt and washed twice with a LiCl concentrated solution to ensure the total DMF removal. The organic phase was washed with a 1M NaOH solution to remove any excess or unreacted **18** substrate. Finally, the organic solution was washed with water, dried over anhydrous MgSO₄ and concentrated under reduce pressure to afford the 2-(oxoalkyl)isoindoline-1,3-dione (**13**).

2-(4-oxopentyl)isoindoline-1,3-dione (13{1})**13{1}**

Starting from 5-chloropentan-2-one **17{1}**, the product **13{1}** (**LMF255**) was obtained as dark yellow solid with 39% yield.

Spectroscopic data:

¹H NMR (400 MHz, CDCl₃) δ 7.83 (dd, *J* = 5.3, 3.1 Hz, 2H, C8-H, C8'-H), 7.71 (dd, *J* = 5.4, 3.1 Hz, 2H, C9-H, C9'-H), 3.70 (t, *J* = 6.7 Hz, 2H, C5-H₂), 2.49 (t, *J* = 7.2 Hz, 2H, C3-H₂), 2.13 (s, 3H, C1-H₃), 1.95 (p, *J* = 6.9 Hz, 2H, C4-H₂).

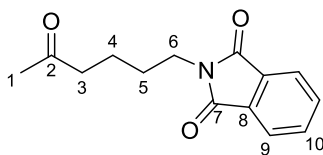
¹³C NMR (100 MHz, DMSO-*d*₆) δ 207.6 (C2), 168.6 (C6, C6'), 134.1 (C9, C9'), 132.2 (C7, C7'), 123.4 (C8, C8'), 40.7 (C3), 37.3 (C5), 30.1 (C1), 22.8 (C4).

IR (ATR), ν_{max} (cm⁻¹) 2928 (*st* C-H), 1770, 1705 (*st* C=O), 1396 (*rock* C-H), 1362, 1333, 1163, 1047, 873, 719 (*oop b* C_{ar}-H).

MS (70 eV, EI) m/z (%): 232.2 (2%) [M+1]⁺, 231.1 (9%) [M]⁺, 188.1 (72%) [C₁₀H₁₁NO₂]⁺, 174.1 (100%) [C₉H₉NO₂]⁺, 161.1 (26%), 160.1 (67%) [CH₂Phth]⁺, 130.1 (17%), 104.0 (24%), 77.0 (24%), 76.0 (32%).

mp: 70-74 °C

Spectroscopic data and melting point are consistent with those previously described (bibliographic yield: 74%)²⁵¹.

2-(5-oxohexyl)isoindoline-1,3-dione (13{2})**13{2}**

Starting from 6-chlorohexan-2-one **17{2}**, the product **13{2}** (**LMF256**) was obtained as a pale-yellow solid with 89% yield.

Spectroscopic data:

¹H NMR (400 MHz, CDCl₃) δ 7.83 (dd, *J* = 5.5, 3.0 Hz, 2H, C9-H, C9'-H), 7.71 (dd, *J* = 5.5, 3.0 Hz, 2H, C10-H, C10'-H), 3.69 (t, *J* = 6.9 Hz, 2H, C6-H₂), 2.49 (t, *J* = 7.2 Hz, 2H, C3-H₂), 2.13 (s, 3H, C1-H₃), 1.73 – 1.57 (m, 4H, C4-H₂, 15-H₂).

¹³C NMR (100 MHz, CDCl₃) δ 208.5 (C2), 168.6 (C7, C7'), 134.1 (C10, C10'), 132.2 (C8, C8'), 123.4 (C9, C9'), 43.0 (C3), 37.6 (C6), 30.1 (C1), 28.1 (C5), 20.9 (C4).

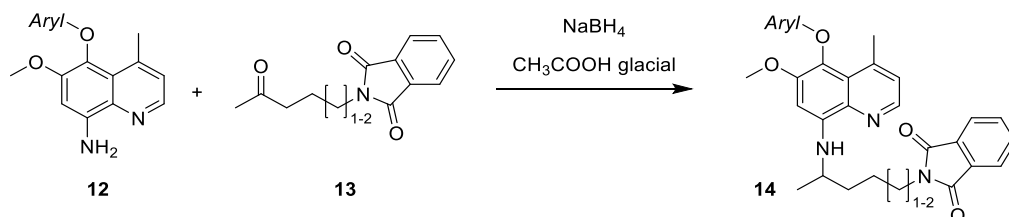
IR (ATR), ν_{max} (cm⁻¹) 2941 (*st* C-H), 1770, 1702 (*st* C=O), 1395 (*rock* C-H), 1358, 1045, 716 (*oop* *b* C_{ar}-H).

MS (70 eV, EI) m/z (%): 246.2 (2%) [M+1]⁺, 245.2 (13%) [M]⁺, 202.1 (13%) [M-C₂H₄O]⁺, 189.1 (10%), 188.1 (72%) [C₁₀H₁₁NO₂]⁺, 161.1 (33%), 160.1 (100%) [CH₂Phth]⁺, 130.1 (20%).

mp: 62-63 °C

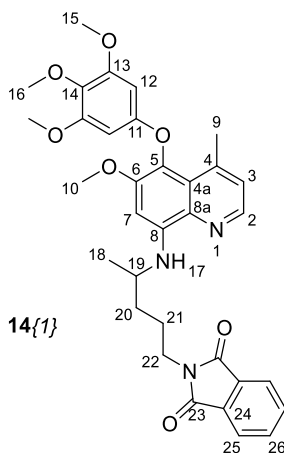
Spectroscopic data and melting point are consistent with those previously described (bibliographic yield: 89%)²⁵².

5.2.2.9. Synthesis of 2-(5-((4-(phenoxy)-3-methoxy-5-methylnaphthalen-1-yl)amino)alkyl)isoindoline-1,3-dione derivatives



A mixture of 0.91 mmol (1.0 eq.) of the corresponding **12** analog and 1.3 mmol (1.4 eq.) of the desired compound **13** was placed in a previously dried 50 ml round-bottomed flask along with 10 ml of glacial acetic acid. The mixture was stirred under argon atmosphere at room temperature for 30 minutes. Later, 1.76 mmol (1.9 eq.) of sodium borohydride were added portionwise (the formation of bubbles could be seen) maintaining the temperature at 25-30 °C making use of a water bath. After the addition, the reaction mixture was stirred for 30 min and the poured carefully into an ice-cold solution of 2M sodium hydroxide. The clear formation of a precipitate was seen. This was filtered and chromatographed over silica gel using an automated flash chromatography (silica column, Cy: AcOEt 1:1) to finally afford the 2-(5-((4-(phenoxy)-3-methoxy-5-methylnaphthalen-1-yl)amino)alkyl)isoindoline-1,3-dione (**14**).

2-(4-((6-methoxy-4-methyl-5-(3,4,5-trimethoxyphenoxy)quinolin-8-yl)amino)pentyl)isoindoline-1,3-dione (14{1})



Starting from 6-methoxy-4-methyl-5-(3,4,5-trimethoxyphenoxy)quinolin-8-amine **12{1}** and 2-(4-oxopentyl)isoindoline-1,3-dione **13{1}**, the product **14{1}** (**LMF267**) was obtained as a yellow solid with 52% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 8.36 (d, *J* = 4.2 Hz, 1H, C2-H), 7.83 – 7.78 (m, 4H, C25-H, C25'-H, C26-H, C26'-H), 7.17 (dd, *J* = 4.3, 0.9 Hz, 1H, C3-H), 6.60 (s, 1H, C7-H), 6.28 (d, *J* = 9.1 Hz, 1H, N17-H), 5.98 (s, 2H, C12-H, C12'-H), 3.83 – 3.74 (m, 1H, C19-H), 3.75 (s, 3H, C10-H₃), 3.62 – 3.60 (m, 2H, C22-H₂), 3.58 (s, 6H, C15-H₃, C15'-H₃), 3.57 (s, 3H, C16-H₃), 2.54 (d, *J* = 0.7 Hz, 3H, C9-H₃), 1.81 – 1.53 (m, 4H, C20-H₂, C21-H₂), 1.22 (d, *J* = 6.3 Hz, 3H, C18-H₃).

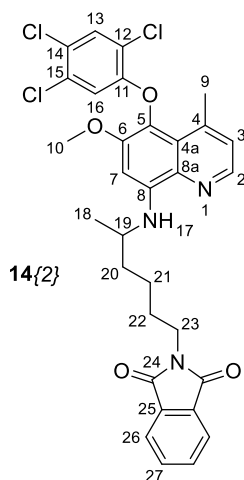
¹³C NMR (100 MHz, DMSO-*d*₆) δ 168.2 (C23, C23'), 156.0 (C11), 153.6 (C13, C13'), 151.2 (C6), 144.5 (C2), 143.2 (C5*), 141.9 (C4), 134.5 (C26, C26'), 132.9 (C8*), 132.0 (C14), 131.6 (C25, C25'), 125.0 (C3), 124.5 (C8a*), 124.1 (C4a), 123.1 (C24, C24'), 93.3 (C7), 92.4 (C12, C12'), 60.3 (C16), 56.4 (C10), 55.9 (C15, C15'), 47.1 (C19), 37.7 (C22), 33.3 (C20), 24.9 (C21), 22.7 (C9), 20.3 (C18).

IR (ATR), ν_{max} (cm⁻¹) 3378 (*st* N-H), 2931 (*st* C-H), 2853 (*st* C-H), 1709 (*st* C=O), 1581 (*b* N-H), 1526 (*st* C_{ar}-C_{ar}), 1469 (*st* C_{ar}-C_{ar}), 1449, 1396, 1210 (*st* C-O), 930, 874, 716 (*w* N-H).

MS (70 eV, EI) m/z (%): 586.4 (28%) [M+1]⁺, 585.4 (70%) [M]⁺, 418.3 (15%), 398.3 (26%), 397.3 (100%), [M-C₁₁H₁₀NO₂]⁺, 367.3 (10%), 309.2 (18%), 215.2 (21%) [C₁₃H₁₄N₂O]⁺, 160.1 (75%) [C₉H₆NO₂]⁺, 130.1 (20%).

HRMS (Q-TOF): m/z calculated for C₃₃H₃₅N₃O₇ [M]⁺: 585.2475. Found: 586.2515 [M+H]⁺,

mp: 70-74 °C

2-((6-methoxy-4-methyl-5-(2,4,5-trichlorophenoxy)quinolin-8-yl)amino)hexyl)isoindoline-1,3-dione (14{2})

Starting from 6-methoxy-4-methyl-5-(2,4,5-trichlorophenoxy)quinolin-8-amine **12{2}** and 2-(5-oxohexyl)isoindoline-1,3-dione **13{2}**, the product **14{2}** (**LMF272**) was obtained as a yellow solid with 28% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 8.37 (d, *J* = 4.3 Hz, 1H, C2-H), 7.97 – 7.92 (s, 1H, C13-H), 7.80 (s, 4H, C26-H, C26'-H, C27-H, C27'-H), 7.21 (dd, *J* = 4.3, 0.9 Hz, 1H, C3-H), 6.62 (m, 2H, C7-H, C16-H), 6.39 (d, *J* = 8.9 Hz, 1H, N17-H), 3.81 (s, 3H, C10-H₃), 3.80 – 3.72 (m, 1H, C19-H), 3.57 (t, *J* = 7.0 Hz, 2H, C23-H₂), 2.50 (s, 3H, C9-H₃), 1.67 (m, 4H, C20-H₂, C21-H₂), 1.39 (m, 2H, C22-H₂), 1.24 (d, *J* = 6.2 Hz, 3H, C18-H₃).

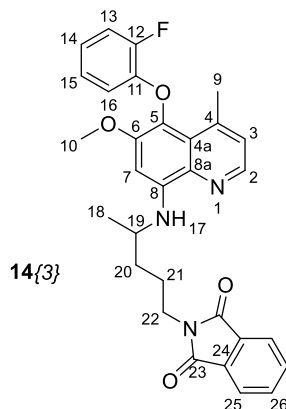
¹³C NMR (100 MHz, DMSO-*d*₆) δ 167.9 (C24, C24'), 154.0 (C11), 150.6 (C6), 144.4 (C2), 144.0 (C8a*), 140.9 (C4), 134.3 (C27, C27'), 132.6 (C5*), 131.5 (C25, C25'), 131.2 (C13), 130.6 (C15), 125.2 (C3), 123.5 (C14), 123.1 (C4a), 122.9 (C26, C26'), 122.8 (C8*), 120.4 (C12), 115.5 (C16), 92.4 (C7), 56.3 (C10), 47.0 (C19), 37.3 (C23), 35.7 (C20), 27.9 (C22), 22.9 (C21*), 21.7 (C9), 20.4 (C18).

IR (ATR), ν_{max} (cm⁻¹) 3382 (*st* N-H), 2936 (*st* C-H), 2851 (*st* C-H), 1709 (*st* C=O), 1611 (*b* N-H), 1526 (*st* C_{ar}-C_{ar}), 1449 (*st* C_{ar}-C_{ar}), 1449, 1395, 1350, 1224 (*st* C-O), 1074 (*st* C-N), 878, 717 (*w* N-H).

HRMS (Q-TOF): *m/z* calculated for C₃₁H₂₈Cl₃N₃O₄ [M]⁺: 611.1145 Found: 612.1197 [M+H]⁺, 614.1174 [M+H+2]⁺, 616.1163 [M+H+4]⁺, 618.1173 [M+H+6]⁺

mp: 76-80 °C

2-(4-((5-(2-fluorophenoxy)-6-methoxy-4-methylquinolin-8-yl)amino)pentyl)isoindoline-1,3-dione (14{3})



Starting from 5-(2-fluorophenoxy)-6-methoxy-4-methylquinolin-8-amine **12{3}** and 2-(4-oxopentyl)isoindoline-1,3-dione **13{1}**, the product **14{3}** (**LMF266**) was obtained as a thick yellow slurry with 33% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 8.39 (d, *J* = 4.2 Hz, 1H, C2-H), 7.86 – 7.79 (m, 4H, C25-H, C25'-H, C26-H, C26'-H), 7.34 – 7.26 (m, 1H, C13-H), 7.20 (dd, *J* = 4.3, 0.9 Hz, 1H, C3-H), 7.02 – 6.89 (m, 2H, C14-H, C16-H), 6.61 (s, 1H, C7-H), 6.43 – 6.37 (m, 1H, C15-H), 6.35 (d, *J* = 9.1 Hz, 1H, N17-H), 3.82 (m, 1H, C19-H), 3.76 (s, 3H, C10-H₃), 3.63 (t, *J* = 6.5 Hz, 2H, C22-H₃), 2.52 (d, *J* = 0.8 Hz, 3H, C9-H₃), 1.85 – 1.56 (m, 4H, C20-H₂, C21-H₂), 1.23 (d, *J* = 6.3 Hz, 3H, C18-H₃).

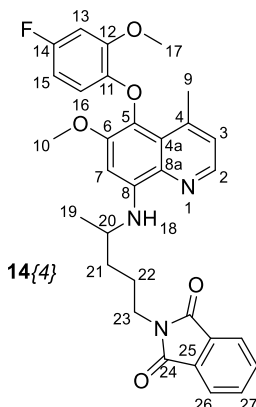
¹³C NMR (100 MHz, DMSO-*d*₆) δ 168.0 (C23, C23'), 151.0 (d, *J* = 243.7 Hz, C11), 150.9 (C6), 146.9 (d, *J* = 10.0 Hz, C11), 144.3 (C2), 143.5 (C8a*), 141.3 (C4), 134.3 (C26, C26'), 132.8 (C5*), 131.6 (C24, C24'), 125.0 (C3), 124.8 (d, *J* = 3.7 Hz, C16), 123.6 (C8*), 123.6 (C4a), 122.9 (C25, C25'), 121.5 (d, *J* = 6.8 Hz, C14), 116.4 (d, *J* = 17.5 Hz, C13), 115.1 (C15), 92.8 (C7), 56.2 (C10), 47.0 (C19), 37.5 (C22), 33.2 (C20), 24.8 (C21), 21.8 (C9), 20.2 (C18).

IR (ATR), ν_{max} (cm⁻¹) 3400 (*st* N-H), 2931 (*st* C-H), 2855 (*st* C-H), 1706 (*st* C=O), 1614 (*b* N-H), 1525 (*st* C_{ar}-C_{ar}), 1498 (*st* C_{ar}-C_{ar}), 1454, 1397, 1365, 1257 (*st* C-O), 1193 (*st* C-N), 1044, 792, 744, 721 (*w* N-H).

MS (70 eV, EI) m/z (%): 514.4 (17%) [M+1]⁺, 513.4 (42%) [M]⁺, 326.2 (23%), 325.2 (100%) [M-C₁₁H₁₀NO₂]⁺, 215.2 (28%) [C₁₃H₁₄N₂O]⁺, 160.1 (42%) [C₉H₆NO₂]⁺.

HRMS (Q-TOF): m/z calculated for C₃₀H₂₈FN₃O₄ [M]⁺: 513.2064 Found: 514.2106 [M+H]⁺

2-(4-((5-(4-fluoro-2-methoxyphenoxy)-6-methoxy-4-methylquinolin-8-yl)amino)pentyl)isoindoline-1,3-dione (14{4})



Starting from 5-(4-fluoro-2-methoxyphenoxy)-6-methoxy-4-methylquinolin-8-amine **12{4}** and 2-(4-oxopentyl)isoindoline-1,3-dione **13{1}**, the product **14{4}** (**LMF260**) was obtained as a yellow crystalline solid with 34% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 8.37 (d, *J* = 4.2 Hz, 1H, C2-H), 7.89 – 7.77 (m, 4H, C25-H, C25'-H, C26-H, C26'-H), 7.17 (dd, *J* = 4.3, 0.9 Hz, 1H, C3-H), 6.99 (dd, *J* = 10.6, 2.9 Hz, 1H, C13-H), 6.60 (s, 1H, C7-H), 6.54 – 6.44 (m, 1H, C15-H), 6.30 (d, *J* = 9.0 Hz, 1H, N18-H), 6.15 (dd, *J* = 9.0, 5.6 Hz, 1H, C16-H), 3.88 (s, 3H, C17-H₃), 3.80 (s, 1H, C20-H), 3.74 (s, 3H, C10-H₃), 3.63 (t, *J* = 6.5 Hz, 2H, C23-H₂), 2.49 (s, 3H, C9-H₃), 1.85 – 1.54 (m, 4H, C21-H₂, C22-H₂), 1.23 (d, *J* = 6.3 Hz, 3H, C19-H₃).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 168.0 (C24, C24'), 156.7 (d, *J* = 236.1 Hz, C14), 150.9 (C6), 148.8 (d, *J* = 10.3 Hz, C12), 145.0 (C11), 144.2 (C2), 143.1 (C8a*), 141.7 (C4), 134.3 (C27, C27'), 132.9 (C5*), 131.6 (C25, C25'), 124.8 (C3), 124.2 (C8*), 123.8 (C4a), 123.0 (C26, C26'), 112.9 (d, *J* = 10.3 Hz, C16), 105.6 (d, *J* = 22.5 Hz, C15), 101.0 (d, *J* = 27.5 Hz, C13), 93.1 (C7), 56.2 (C10), 56.1 (C17), 47.0 (C20), 37.5 (C23), 33.3 (C21), 24.9 (C22), 21.6 (C9), 20.3 (C19).

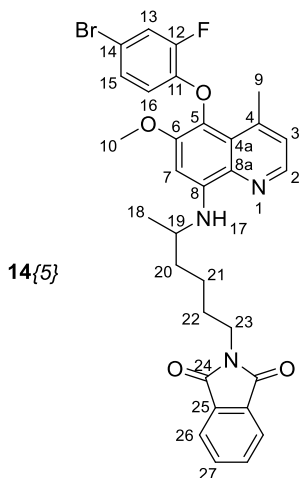
IR (KBr), ν_{max} (cm⁻¹) 3367 (*st* N-H), 2937 (*st* C-H), 1709 (*st* C=O), 1610 (*b* N-H), 1525, 1502 (*st* C_{ar}-C_{ar}), 1450 (*st* C_{ar}-C_{ar}), 1395, 1187 (*st* C-O), 1106 (*st* C_{ar}-N), 1028, 951, 831, 718 (*w* N-H).

MS (70 eV, EI) m/z (%): 544.3 (15%) [M+1]⁺, 543.3 (33%) [M]⁺, 356.2 (15%), 355.2 (61%) [M-C₁₁H₁₀NO₂]⁺, 216.1 (17%), 215.2 (28%) [C₁₃H₁₄N₂O]⁺, 188.1 (58%) [C₁₁H₁₀NO₂]⁺, 161.2 (18%), 160.2 (100%) [C₉H₆NO₂]⁺, 130.1 (33%), 84.0 (33%).

HRMS (Q-TOF): m/z calculated for C₃₁H₃₀FN₃O₅ [M]⁺: 543.2169 Found: 544.2206 [M+H]⁺

mp: 117-121 °C

2-(5-((5-(4-bromo-2-fluorophenoxy)-6-methoxy-4-methylquinolin-8-yl)amino)hexyl)isoindoline-1,3-dione (14{5})



Starting from 5-(4-bromo-2-fluorophenoxy)-6-methoxy-4-methylquinolin-8-amine **12{5}** and 2-(5-oxohexyl)isoindoline-1,3-dione **13{2}**, the product **14{5}** (**LMF271**) was obtained as a yellow slurry with 31% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 8.36 (d, *J* = 4.2 Hz, 1H, C2-H), 7.81 (s, 4H, C26-H, C26'-H, C27-H, C27-H'), 7.64 (dd, *J* = 10.8, 2.4 Hz, 1H, C13-H), 7.20 (dd, *J* = 4.3, 0.9 Hz, 1H, C3-H), 7.17 (ddd, *J* = 8.8, 2.4, 1.5 Hz, 1H, C16-H), 6.61 (s, 1H, C7-H), 6.43 (t, *J* = 9.0 Hz, 1H, C15-H), 6.34 (d, *J* = 8.9 Hz, 1H, N17-H), 3.79 (s, 3H, C10-H₃), 3.77 – 3.72 (m, 1H, C19-H), 3.58 (t, *J* = 7.0 Hz, 2H, C23-H₂), 2.52 (d, *J* = 0.6 Hz, 3H, C9-H₃), 1.77 – 1.29 (m, 6H, C20-H₂, C21-H₂, C22-H₂), 1.23 (d, *J* = 6.3 Hz, 3H, C18-H₃).

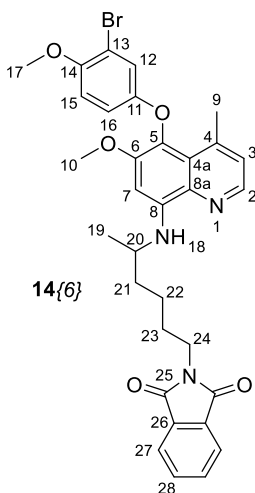
¹³C NMR (100 MHz, DMSO-*d*₆) δ 168.0 (C24, C24'), 151.0 (d, *J* = 248.7 Hz, C12), 150.8 (C6), 146.5 (d, *J* = 10.4 Hz, C11), 144.4 (C2), 143.7 (C8a*), 141.1 (C4), 134.3 (C27, C27'), 132.6 (C5*), 131.5 (C25, C25'), 127.8 (d, *J* = 3.5 Hz, C16), 125.1 (C3), 123.4 (C4a), 123.2 (C8*), 122.9 (C26, C26'), 119.8 (d, *J* = 20.6 Hz, C13), 117.0 (C15), 111.5 (d, *J* = 8.5 Hz, C14), 92.5 (C7), 56.3 (C10), 47.1 (C19), 37.3 (C23), 35.8 (C20), 27.9 (C22*), 23.0 (C21*), 21.8 (C9), 20.4 (C18).

IR (ATR), ν_{max} (cm⁻¹) 3379 (st N-H), 2931 (st C-H), 2855 (st C-H), 1707 (st C=O), 1616 (b N-H), 1526 (st C_{ar}-C_{ar}), 1500 (st C_{ar}-C_{ar}), 1395, 1353, 1261 (st C-O), 1194 (st C-N), 717 (w N-H).

MS (70 eV, EI) m/z (%): 607.3 (35%) [M+2]⁺, 606.3 (6%) [M+1]⁺, 605.3 (36%) [M]⁺, 418.3 (16%), 417.3 (54%) [M-C₁₁H₁₀NO₂]⁺, 406.1 (21%), 405.3 (51%), 403.3 (59%) [M-C₁₂H₁₂NO₂]⁺, 216.1 (30%), 215.3 (83%), 215.1 (100%) [C₁₃H₁₄N₂O]⁺, 203.1 (44%) [C₁₂H₁₂NO₂]⁺, 188.1 (23%) [C₆H₄BrF]⁺, 175.1 (29%), 173.1 (21%) [C₁₀H₇NO₂]⁺, 172.1 (24%), 160.1 (69%) [C₉H₆NO₂]⁺, 130.1 (25%), 56.1 (22%).

HRMS (Q-TOF): m/z calculated for C₃₁H₂₉BrFN₃O₄ [M]⁺: 605.1325 Found: 606.1360 [M+H]⁺, 608.1361 [M+H+2]⁺

2-(5-((5-(3-bromo-4-methoxyphenoxy)-6-methoxy-4-methylquinolin-8-yl)amino)hexyl)isoindoline-1,3-dione (14{6})



Starting from 5-(3-bromo-4-methoxyphenoxy)-6-methoxy-4-methylquinolin-8-amine **12{6}** and 2-(5-oxohexyl)isoindoline-1,3-dione **13{2}**, the product **14{6}** (**LMF305**) was obtained as a thick yellow slurry with 43% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 8.35 (d, *J* = 4.2 Hz, 1H, C2-H), 7.82 – 7.80 (m, 4H, C27-H, C27'-H, C28-H, C28'-H), 7.17 (dd, *J* = 4.3, 0.9 Hz, 1H, C3-H), 7.00 (d, *J* = 9.2 Hz, 1H, C15-H), 6.94 (d, *J* = 3.0 Hz, 1H, C12-H), 6.68 (dd, *J* = 9.0, 3.0 Hz, 1H, C16-H), 6.60 (s, 1H, C7-H), 6.29 (d, *J* = 8.9 Hz, 1H, N18-H), 3.78 (s, 3H, C10-H₃), 3.76 (s, 3H, C17-H₃), 3.73 (m, 1H, C20-H), 3.57 (dt, *J* = 9.7, 6.9 Hz, 2H, C24-H₂), 2.53 (d, *J* = 0.6 Hz, 3H, C9-H₃), 1.79 – 1.40 (m, 6H, C21-H₂, C22-H₂, C23-H₂), 1.23 (d, *J* = 6.3 Hz, 3H, C19-H₃).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 167.9 (C25, C25'), 153.5 (C11), 151.0 (C6), 150.0 (C14), 144.3 (C2), 143.3 (C8a*), 141.4 (C4), 134.3 (C28, C28'), 132.7 (C5*), 131.5 (C26, C26'), 124.9 (C3), 124.4 (C8*),

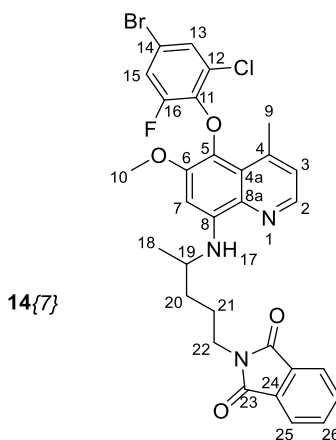
123.8 (C4a), 123.0 (C27, C27'), 118.8 (C12), 114.3 (C16), 113.5 (C15), 110.9 (C13), 92.9 (C7), 56.5 (C17), 56.3 (C10), 47.1 (C20), 37.2 (C24), 35.8 (C21), 27.9 (C23), 22.9 (C22), 22.5 (C9), 20.4 (C19).

IR (ATR), ν_{\max} (cm⁻¹) 3393 (st N-H), 2929 (st C-H), 2856 (st C-H), 1705 (st C=O), 1608 (b N-H), 1525 (st C_{ar}-C_{ar}), 1488, 1395 (st C_{ar}-C_{ar}), 1353, 1193 (st C-O), 1042 (st C-N), 717 (w N-H).

MS (70 eV, EI) m/z (%): 619.3 (67%) [M+2]⁺, 618.3 (11%) [M+1]⁺, 617.3 (69%) [M]⁺, 418.3 (34%), 417.3 (95%), 415.2 (74%) [M-C₁₂H₁₂NO₂]⁺, 215.1 (62%) [C₁₃H₁₄N₂O]⁺, 203.1 (18%) [C₁₂H₁₂NO₂]⁺, 188.1 (81%) [C₁₁H₁₀NO₂]⁺, 175.1 (28%), 174.1 (21%) [C₁₀H₈NO₂]⁺, 160.2 (100%) [C₉H₆NO₂]⁺, 77.1 (26%), 76.1 (32%), 56.1 (25%).

HRMS (Q-TOF): m/z calculated for C₃₂H₃₂BrN₃O₅ [M]⁺: 617.1525 Found: 618.1576 [M+H]⁺, 620.1561 [M+H+2]⁺

2-(4-((5-(4-bromo-2-chloro-6-fluorophenoxy)-6-methoxy-4-methylquinolin-8-yl)amino)pentyl)isoindoline-1,3-dione (14{7})



Starting from 5-(4-bromo-2-chloro-6-fluorophenoxy)-6-methoxy-4-methylquinolin-8-amine **12{7}** and 2-(4-oxopentyl)isoindoline-1,3-dione **13{1}**, the product **14{7}** (**LMF307**) was obtained as a thick yellow slurry with 43% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 8.38 (d, *J* = 4.2 Hz, 1H, C2-H), 7.80 (m, 4H, C25-H, C25'-H, C26-H, C26'-H), 7.61 (t, *J* = 1.9 Hz, 1H, C13-H), 7.49 (dd, *J* = 11.5, 2.3 Hz, 1H, C15-H), 7.25 (*J* = 4.3, 0.9 Hz, 1H, C3-H), 6.44 (s, C7-H), 6.21 (d, *J* = 8.9 Hz, 1H, C7-H), 3.72 (m, 1H, C19-H), 3.60 (s, 3H, C10-H₃), 2.64 (d, *J* = 0.7 Hz, 3H, C9-H₃), 1.78 – 1.47 (m, 4H, C20-H₂, C21-H₂), 1.17 (d, *J* = 6.2 Hz, 3H, C18H₃).

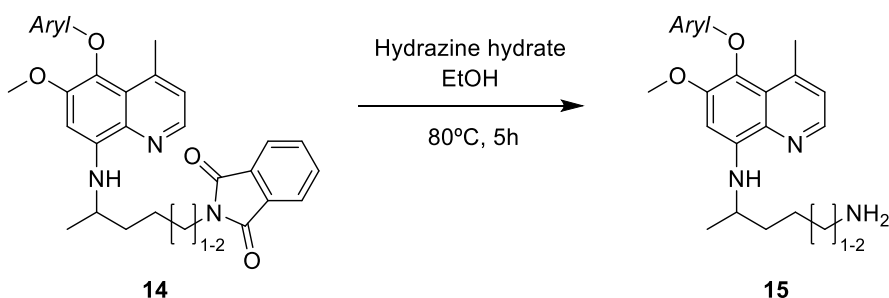
^{13}C NMR (100 MHz, DMSO- d_6) δ 168.0 (C23, C23'), 154.4 (d, J = 249.5 Hz, C16), 149.1 (C6), 144.5 (C2), 143.1 (C8a*), 142.6 (d, J = 8.4 Hz, C11), 141.9 (C4), 134.3 (C26, C26'), 132.5 (C5*), 131.5 (C23, C23'), 128.9 (C8*), 128.3 (d, J = 3.9 Hz, C13), 126.1 (d, J = 3.4 Hz, C12**), 124.9 (C3), 122.9 (C25, C25'), 122.6 (C4a), 119.5 (d, J = 22.9 Hz, C15), 112.9 (d, J = 9.7 Hz, C14), 93.0 (C7), 56.7 (C10), 46.9 (C19), 37.5 (C22), 32.9 (C20), 24.6 (C21), 22.7 (C9), 20.1 (C18).

IR (ATR), ν_{max} (cm^{-1}) 3383 (st N-H), 2929 (st C-H), 2851 (st C-H), 1708 (st C=O), 1613 (b N-H), 1572, 1525 (st $\text{C}_{\text{ar}}-\text{C}_{\text{ar}}$), 1487 (st $\text{C}_{\text{ar}}-\text{C}_{\text{ar}}$), 1395, 1252 (st C-O), 929, 718 (w N-H).

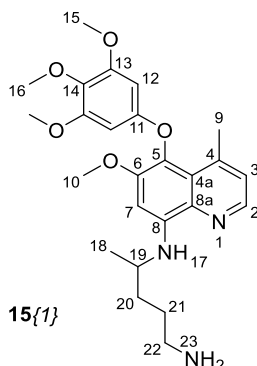
MS (70 eV, EI) m/z (%): 629.3 (9%) $[\text{M}+4]^+$, 627.3 (22%) $[\text{M}+2]^+$, 626.3 (5%) $[\text{M}+1]^+$, 625.3 (18%) $[\text{M}]^+$, 593.3 (30%), 591.3 (24%) $[\text{M}-\text{Cl}]^+$, 439.2 (54%), 437.2 (43%) $[\text{M}-\text{C}_{11}\text{H}_{10}\text{NO}_2]^+$, 405.2 (991%), 403.2 (76%) $[\text{M}-\text{C}_6\text{H}_2\text{BrClFO}_{(\text{phenolate})}]^+$, 389.2 (17%), 387.2 (15%), 215.1 (37%) $[\text{C}_{13}\text{H}_{14}\text{N}_2\text{O}]^+$, 203.1 (17%) $[\text{C}_{12}\text{H}_{12}\text{NO}_2]^+$, 187.1 (81%) $[\text{C}_{11}\text{H}_9\text{NO}_2]^+$, 175.1 (15%), 161.1 (16%), 160.1 (100%) $[\text{C}_9\text{H}_6\text{NO}_2]^+$, 104.1 (21%), 77.1 (15%), 76.1 (14%).

HRMS (Q-TOF): m/z calculated for $\text{C}_{30}\text{H}_{26}\text{BrClFN}_3\text{O}_4$ $[\text{M}]^+$: 625.0779 Found: 626.0823 $[\text{M}+\text{H}]^+$, 628.0802 $[\text{M}+\text{H}+2]^+$, 630.0800 $[\text{M}+\text{H}+4]^+$

5.2.2.10. Synthesis of *N*-(6-methoxy-4-methyl-5-(phenoxy)quinolin-8-yl)alkyldiamine derivatives



A mixture of 0.21 mmol (1.0 eq.) of the corresponding **14** analog and 8 ml of ethanol were placed in a 10 ml round-bottomed flask and refluxed at 80 °C for 30 minutes. Next, 0.07 ml (5.7 eq.) of hydrazine hydrate were added and stirred at the same temperature for 5 hours. The reaction mixture was evaporated under vacuum and the residue was partitioned twice with 3 ml of 10% KOH and 3 ml of AcOEt. The resulting organic layers were washed four times with 3 ml of water, dried with anhydrous sodium sulfate and the solvent was finally removed by rotary evaporation to yield the pure desired *N*-(6-methoxy-4-methyl-5-(phenoxy)quinolin-8-yl)alkyldiamine derivative (**15**).

***N*⁴-(6-methoxy-4-methyl-5-(3,4,5-trimethoxyphenoxy)quinolin-8-yl)pentane-1,4-diamine (**14{1}**)**

Starting from 2-(4-((6-methoxy-4-methyl-5-(3,4,5-trimethoxyphenoxy)quinolin-8-yl)amino)pentyl)isoindoline-1,3-dione **14{1}**, the product **15{1}** (**LMF270**) was obtained as an orange solid with 83% yield.

Spectroscopic data:

¹H NMR (400 MHz, CD₃OD) δ 8.39 (d, *J* = 4.3 Hz, 1H, C2-H), 7.17 (dd, *J* = 4.3, 1.0 Hz, 1H, C3-H), 6.65 (s, 1H, C7-H), 6.06 (s, 2H, C12-H, C12'-H), 3.86 (s, 3H, C10-H₃), 3.85 – 3.76 (m, 1H, C19-H), 3.71 (s, 3H, C16-H₃), 3.68 (s, 6H, C15-H₃, C15-H₃), 2.85 – 2.76 (m, 2H, C22-H₂), 2.66 (d, *J* = 1.0 Hz, 3H, C9-H₃), 1.84 – 1.65 (m, 4H, C20-H₂, C21-H₂), 1.36 (d, *J* = 6.4 Hz, 3H, C18-H₃).

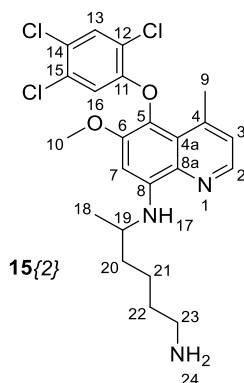
¹³C NMR (100 MHz, CD₃OD) δ 158.0 (C11), 155.1 (C13, C13'), 152.8 (C6), 145.7 (C2), 144.9 (C8a*), 144.2 (C4), 134.9 (C5*), 133.5 (C14), 127.1 (C8*), 125.9 (C4a), 125.9 (C3), 94.9 (C7), 93.6 (C12, C12'), 61.2 (C16), 57.2 (C10), 56.5 (C15, C15'), 49.0 (C19), 35.0 (C20), 28.5 (C21), 23.3 (C9), 20.9 (C18).

IR (ATR), ν_{max} (cm⁻¹) 3378 (*st* N-H), 2934 (*st* C-H), 2832 (*st* C-H), 1595 (*st* C_{ar}-C_{ar}), 1525 (*b* N-H), 1445 (*st* C_{ar}-C_{ar}), 1219 (*st* C_{ar}-N), 1223 (*st* C-O), 1040 (*st* C-N), 988, 809 (*w* N-H).

MS (70 eV, EI) m/z (%): 456.3 (39%) [M+1]⁺, 455.3 (100%) [M]⁺, 397.2 (94%) [M- C₃H₈N]⁺, 371.2 (22%) [M- C₅H₁₂N]⁺, 203.1 (22%), 84.1 (39%), 57.1 (75%) [C₄H₉]⁺.

HRMS (Q-TOF): m/z calculated for C₂₅H₃₃N₃O₅ [M]⁺: 455.2420 Found: 612.1192 [M+H]⁺, 614.1168 [M+H+2]⁺

mp: 87-92 °C

***N*⁵-(6-methoxy-4-methyl-5-(2,4,5-trichlorophenoxy)quinolin-8-yl)hexane-1,5-diamine (15{2})**

Starting from 2-(5-((6-methoxy-4-methyl-5-(2,4,5-trichlorophenoxy)quinolin-8-yl)amino)hexyl)isoindoline-1,3-dione **14{2}**, the product **15{2}** (**LMF274**) was obtained as an orange solid with 93% yield.

Spectroscopic data:

¹H NMR (400 MHz, CD₃OD) δ 8.41 (d, *J* = 4.3 Hz, 1H, C2-H), 7.68 (s, 1H, C13-H), 7.20 (dd, *J* = 4.3, 0.9 Hz, 1H, C3-H), 6.62 (s, 1H, C7-H), 6.52 (s, 1H, C16-H), 3.87 (s, 3H, C10-H₃), 3.80 (m, 1H, C19-H), 2.76 (t, *J* = 6.9 Hz, 2H, C23-H₂), 2.61 (d, *J* = 0.9 Hz, 3H, C9-H₃), 1.84 – 1.66 (m, 2H, C20-H₂), 1.68 – 1.51 (m, 4H, C21-H₂, C22-H₂), 1.36 (d, *J* = 6.3 Hz, 3H, C18-H₃).

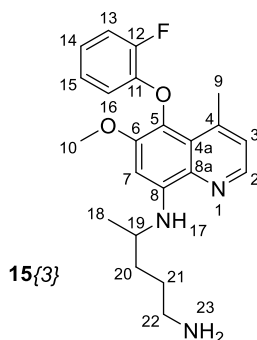
¹³C NMR (100 MHz, CD₃OD) δ 155.8 (C13), 152.2 (C6), 145.8 (C2), 145.7 (C8a*), 143.3 (C4), 134.7 (C5*), 132.4 (C15), 132.2 (C13), 126.3 (C3), 125.5 (C14), 125.5 (C8*), 125.0 (C4a), 122.3 (C12), 116.9 (C16), 93.7 (C7), 57.1 (C10), 49.0, (C19), 41.9 (C23), 37.7 (C20), 32.1 (C22), 24.5 (C21), 22.6 (C9), 20.8 (C18).

IR (ATR), ν_{max} (cm⁻¹) 3381 (*st* N-H), 2926 (*st* C-H), 2851 (*st* C-H), 1609 (*st* C_{ar}-C_{ar}), 1525 (*b* N-H), 1448 (*st* C_{ar}-C_{ar}), 1350 (*st* C_{ar}-N), 1224 (*st* C-O), 1075 (*st* C-N), 926, 753 (*w* N-H).

MS (70 eV, EI) m/z (%): 487.2 (2%) [M+6]⁺, 485.2 (5%) [M+4]⁺, 483.2 (38%) [M+2]⁺, 481.1 (38%) [M]⁺, 413.5 (32%) [M+4-C₄H₁₀N]⁺, 411.1 (93%) [M+2-C₄H₁₀N]⁺, 409.1 (100%) [M-C₄H₁₀N]⁺, 215.1 (51%) [C₁₃H₁₄N₂O]⁺, 203.1 (54%), 98.1 (61%), 57.1 (75%) [C₄H₉]⁺.

HRMS (Q-TOF): m/z calculated for C₂₃H₂₆Cl₃N₃O₂ [M]⁺: 481.1091 Found: 482.1153 [M+H]⁺, 484.1127 [M+H+2]⁺, 486.1113 [M+H+4]⁺, 488.1106 [M+H+6]⁺,

mp: 94-98 °C

***N*⁴-(5-(2-fluorophenoxy)-6-methoxy-4-methylquinolin-8-yl)pentane-1,4-diamine (15{3})**

Starting from 2-(4-((5-(2-fluorophenoxy)-6-methoxy-4-methylquinolin-8-yl)amino)pentyl)isoindoline-1,3-dione **14{3}**, the product **15{3}** (**LMF269**) was obtained as a yellow solid with 89% yield.

Spectroscopic data:

¹H NMR (400 MHz, CD₃OD) δ 8.39 (d, *J* = 4.3 Hz, 1H, C2-H), 7.23 – 7.18 (m, 1H, C13-H), 7.17 (dd, *J* = 4.3, 1.0 Hz, 1H, C3-H), 6.95 – 6.90 (m, 2H, C14-H, C16-H), 6.63 (s, 1H, C7-H), 6.50 – 6.39 (m, 1H, C15-H), 3.84 (s, 3H, C10-H), 3.83 – 3.76 (m, 1H, C19-H), 2.76 (t, *J* = 7.0 Hz, 2H, C22-H₂), 2.64 (d, *J* = 0.7 Hz, 3H, C9-H₃), 1.82 – 1.63 (m, 4H, C20-H₂, C21-H₂), 1.36 (d, *J* = 6.3 Hz, 3H, C18-H₃).

¹³C NMR (100 MHz, CD₃OD) δ 153.3 (d, *J* = 244.4 Hz, C12), 152.5 (C6), 148.77 (d, *J* = 10.3 Hz, C11), 145.7 (C2), 145.1 (C8*), 143.9 (C4), 134.9 (C8a*), 126.5 (C5*), 126.0 (C3), 125.7 (C4a), 125.4 (d, *J* = 3.8 Hz, C16), 122.6 (d, *J* = 6.7 Hz, C14), 117.2 (d, *J* = 17.9 Hz, C13), 116.4 (C15), 94.5 (C7), 57.1 (C10), 49.0 (C19), 42.3 (C22), 35.2 (C20), 29.6 (C21), 22.7 (C9), 20.9 (C18).

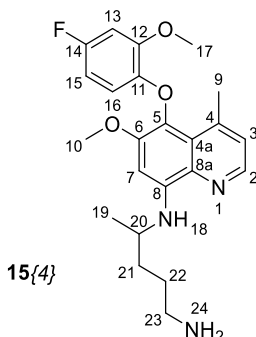
IR (ATR), ν_{max} (cm⁻¹) 3374 (*st* N-H), 2935 (*st* C-H), 2860 (*st* C-H), 1609 (*st* C_{ar}-C_{ar}), 1525 (*bend* N-H), 1497 (*st* C_{ar}-C_{ar}), 1256 (*st* C_{ar}-N), 1192 (*st* C-O), 1103 (*st* C-N), 931, 747 (*w* N-H).

MS (70 eV, EI) m/z (%): 384.3 (15%) [M+1]⁺, 383.3 (42%) [M]⁺, 325.2 (100%) [M-C₃H₈N]⁺, 299.2 (21%), 215.1 (29%) [C₁₃H₁₄N₂O]⁺, 203.1 (21%), 188.1 (27%) [C₁₁H₁₂N₂O]⁺, 175.1 (13%), 159.1 (18%).

HRMS (Q-TOF): m/z calculated for C₂₂H₂₆FN₃O₂ [M]⁺: 383.2009 Found: 384.2064 [M+H]⁺

mp: 88-103 °C

***N*⁴-(5-(4-fluoro-2-methoxyphenoxy)-6-methoxy-4-methylquinolin-8-yl)pentane-1,4-diamine (15{4})**



Starting from 2-(4-((5-(4-fluoro-2-methoxyphenoxy)-6-methoxy-4-methylquinolin-8-yl)amino)pentyl)isoindoline-1,3-dione **14{4}**, the product **15{4}** (**LMF261**) was obtained as an ochre-brownish solid with 94% yield.

Spectroscopic data:

¹H NMR (400 MHz, CD₃OD) δ 8.38 (d, *J* = 4.3 Hz, 1H, C3-H), 7.15 (dd, *J* = 4.3, 0.9 Hz, 1H, C4-H), 6.89 (dd, *J* = 10.3, 2.9 Hz, 1H, C13-H), 6.63 (s, 1H, C7-H), 6.44 (ddd, *J* = 8.9, 8.1, 2.9 Hz, 1H, C15-H), 6.25 (dd, *J* = 8.9, 5.5 Hz, 1H, C16-H), 3.95 (s, 3H, C17-H₃), 3.83 (s, 3H, C10-H₃), 3.84 – 3.74 (m, 1H, C20-H), 2.84 – 2.76 (m, 2H, C23-H₂), 2.62 (d, *J* = 0.9 Hz, 3H, C9-H₃), 1.80 – 1.70 (m, 4H, C21-H₂, C22-H₂), 1.36 (d, *J* = 6.3 Hz, 3H, C19-H₃).

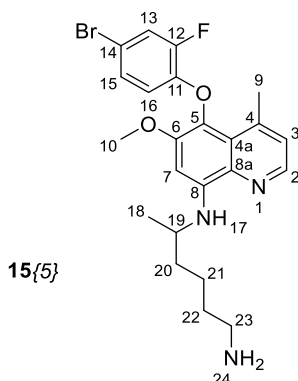
¹³C NMR (100 MHz, CD₃OD) δ 158.9 (d, *J* = 237.6 Hz, C14), 152.5 (C6), 150.6 (d, *J* = 10.0 Hz, C12), 146.7 (C11), 145.7 (C2), 144.9 (C8a*), 144.2 (C4), 135.0 (C5*), 127.1 (C8*), 125.9 (C3), 123.4 (C4a), 114.5 (d, *J* = 9.8 Hz, C16), 106.7 (d, *J* = 23.1 Hz, C15), 101.8 (d, *J* = 27.8 Hz, C13), 94.8 (C7), 57.1 (C10), 56.8 (C17), 49.0 (C20), 42.0 (C23), 35.0 (C21), 28.8 (C22), 22.5 (C9), 20.9 (C19).

IR (ATR), ν_{max} (cm⁻¹) 3383 (*st* N-H), 2926 (*st* C-H), 2859 (*st* C-H), 1608 (*st* C_{ar}=C_{ar}), 1525 (*b* N-H), 1501, 1449 (*st* C_{ar}-C_{ar}), 1402 (*st* C_{ar}-N), 1187 (*st* C-F), 1148 (*st* C-O), 1103 (*st* C-N), 1030, 949, 832 (*w* N-H).

MS (70 eV, EI) m/z (%): 413.2 (24%) [M]⁺, 355.2 (44%) [M-C₃H₈N]⁺, 329.1 (11%) [M-C₅H₁₁N]⁺, 188.1 (100%) [C₁₃H₁₄N₂O]⁺, 159.1 (69%), 84.1 (23%) [C₅H₁₁N]⁺, 55.1 (38%).

HRMS (Q-TOF): m/z calculated for C₂₃H₂₈FN₃O₃ [M]⁺: 413.2115 Found: 414.2170 [M+H]⁺

mp: 102-104 °C

***N*⁵-(5-(4-bromo-2-fluorophenoxy)-6-methoxy-4-methylquinolin-8-yl)hexane-1,5-diamine (15{5})**

Starting from 2-(5-((5-(4-bromo-2-fluorophenoxy)-6-methoxy-4-methylquinolin-8-yl)amino)hexyl)isoindoline-1,3-dione **14{5}**, the product **15{5}** (**LMF273**) was obtained as a yellow-greenish solid with 93% yield. The product showed absorbance and fluorescence at UV at 366 nm.

Spectroscopic data:

¹H NMR (400 MHz, CD₃OD) δ 8.38 (d, *J* = 4.3 Hz, 1H, C2-H), 7.41 (dd, *J* = 10.6, 2.3 Hz, 1H, C13-H), 7.17 (dd, *J* = 4.3, 1.0 Hz, 1H, C3-H), 7.09 (ddd, *J* = 8.9, 2.3, 1.6 Hz, 1H, C16-H), 6.60 (s, 1H, C7-H), 6.41 (t, *J* = 8.9 Hz, 1H, C15-H), 3.84 (s, 3H, C10-H₃), 3.77 (qt, *J* = 6.3 Hz, 1H, C19-H), 2.72 (t, *J* = 6.9 Hz, 2H, C23-H₂), 2.63 (s, 3H), 1.83 – 1.46 (m, 6H, C20-H₂, C21-H₂, C22-H₂), 1.34 (d, *J* = 6.3 Hz, 3H, C18-H₃).

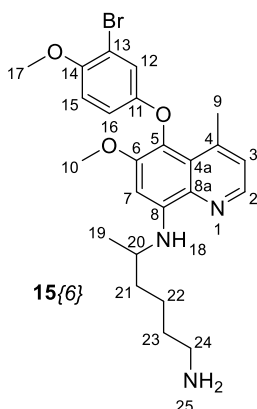
¹³C NMR (100 MHz, CD₃OD) δ 153.1 (d, *J* = 249.8 Hz, C12), 152.4 (C6), 148.3 (d, *J* = 10.3 Hz, C11), 145.7 (C2), 145.4 (C8a*), 143.6 (C4), 134.8 (C5*), 128.5 (d, *J* = 3.9 Hz, C16), 126.1 (C3), 125.8 (C8*), 125.5 (C4a*), 120.7 (d, *J* = 21.3 Hz, C13), 118.0 (d, *J* = 2.0 Hz, C15), 113.2 (d, *J* = 8.0 Hz, C14), 94.1 (C7), 57.0 (C10), 49.0 (C19) 42.2 (C23), 37.7 (C22), 33.0 (C20), 24.6 (C21), 22.7 (C9), 20.8 (C18).

IR (ATR), ν_{max} (cm⁻¹) 3380 (st N-H), 2930 (st C-H), 2857 (st C-H), 1610 (st C_{ar}-C_{ar}), 1525 (b N-H), 1489 (st C_{ar}-C_{ar}), 1260 (st C_{ar}-N), 1194 (st C-O), 1118 (st C-N), 1037, 806 (w N-H).

MS (70 eV, EI) m/z (%): 477.2 (58%) [M+2]⁺, 476.3 (28%) [M+1]⁺, 475.2 (59%) [M]⁺, 443.2 (31%), 441.2 (30%), 405.1 (91%) 403.1 (100%) [M+2-C₄H₁₀N]⁺, 403.1 (100%) [M-C₄H₁₀N]⁺, 371.1 (47%), 369.1 (52%), 215.1 (38%) [C₁₃H₁₄N₂O]⁺, 175.1 (32%), 160.2 (26%), 98.1 (42%), 84.1 (33%) [C₅H₁₁N]⁺, 55.1 (25%).

HRMS (Q-TOF): m/z calculated for C₂₃H₂₇BrFN₃O₂ [M]⁺: 475.1271 Found: 476.1334 [M+H]⁺, 478.1317 [M+H+2]⁺

mp: 82-94 °C

***N*⁵-(5-(3-bromo-4-methoxyphenoxy)-6-methoxy-4-methylquinolin-8-yl)hexane-1,5-diamine (15{6})**

Starting from 2-(5-((5-(3-bromo-4-methoxyphenoxy)-6-methoxy-4-methylquinolin-8-yl)amino)hexyl)isoindoline-1,3-dione **14{6}**, the product **15{6}** (**LMF306**) was obtained as a yellow solid with 90% yield.

Spectroscopic data:

¹H NMR (400 MHz, CD₃OD) δ 8.39 (d, *J* = 4.3 Hz, 1H, C2-H), 7.16 (dd, *J* = 4.3, 1.0 Hz, 1H, C3-H), 6.93 (d, *J* = 9.1 Hz, 1H, C15-H), 6.91 (d, *J* = 3.0 Hz, 1H, C12-H), 6.69 (dd, *J* = 9.0, 3.0 Hz, 1H, C16-H), 6.62 (d, *J* = 9.0 Hz, 1H, C7-H), 3.84 (s, 3H, C17-H₃), 3.81 (s, 3H, C10-H₃), 3.80 – 3.70 (m, 1H, C20-H), 2.74 – 2.68 (t, *J* = 6.9 Hz, 2H, C24-H₂), 2.63 (d, *J* = 0.7 Hz, 3H, C9-H₃), 1.84 – 1.47 (m, 6H, C21-H₂, C22-H₂, C23-H₂), 1.35 (d, *J* = 6.3 Hz, 3H, C19-H₃).

¹³C NMR (100 MHz, CD₃OD) δ 155.4 (C11), 152.6 (C6), 152.1 (C14), 145.7 (C2), 145.0 (C8a*), 143.9 (C4), 134.9 (C4), 127.2 (C8*), 125.9 (C3), 125.8 (C4a), 120.6 (C12), 115.5 (C16), 114.3 (C15), 112.9 (C13), 94.7 (C7), 57.2 (C17), 57.1 (C10), 49.0 (C20), 42.2 (C24), 37.7 (C21), 33.0 (C23), 24.6 (C22), 23.2 (C9), 20.9 (C19).

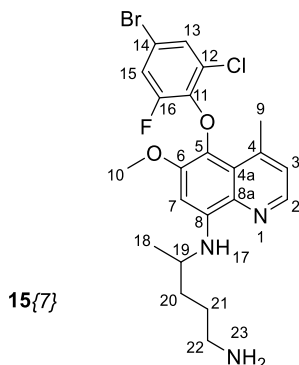
IR (ATR), ν_{max} (cm⁻¹) 3375 (*st* N-H), 2923 (*st* C-H), 2846 (*st* C-H), 1607 (*st* C_{ar}-C_{ar}), 1524 (*b* N-H), 1487 (*st* C_{ar}-C_{ar}), 1450 (*st* C_{ar}-C_{ar}), 1263 (*st* C-O), 1193 (*st* C_{ar}-N), 1135 (*st* C-N), 1040 (*st* C_{ar}-Br), 803, 752 (*w* N-H).

MS (70 eV, EI) m/z (%): 489.2 (25%) [M+2]⁺, 488.2 (9%) [M+1]⁺, 487.2 (25%) [M]⁺, 417.1 (54%) [M+2-C₄H₁₀N]⁺, 415.1 (56%) [M-C₄H₁₀N]⁺, 287.2 (18%), 216.1 (21%), 215.1 (100%) [C₁₃H₁₄N₂O]⁺, 203.1 (31%), 188.1 (34%) [C₁₁H₁₂N₂O]⁺, 175.1 (18%), 159.1 (17%), 98.1 (32%), 97.1 (20%), 85.1 (28%), 84.1 (26%) [C₅H₁₁N]⁺, 71.1 (34%), 69.1 (20%), 57.1 (41%) [C₄H₉]⁺, 55.1 (27%).

HRMS (Q-TOF): m/z calculated for $C_{24}H_{30}BrN_3O_3$ $[M]^+$: 487.1471 Found: 488.1526 $[M+H]^+$, 490.1509 $[M+H+2]^+$

mp: 98-103 °C

***N*⁴-(5-(4-bromo-2-chloro-6-fluorophenoxy)-6-methoxy-4-methylquinolin-8-yl)pentane-1,4-diamine (15{7})**



Starting from 2-(4-((5-(4-bromo-2-chloro-6-fluorophenoxy)-6-methoxy-4-methylquinolin-8-yl)amino)pentyl)isoindoline-1,3-dione **14{7}**, the product **15{7}** (**LMF308**) was obtained as a yellow solid with quantitative yield (100%).

Spectroscopic data:

¹H NMR (400 MHz, CD₃OD) δ 8.40 (d, J = 4.3 Hz, 1H, C2-H), 7.47 (s, 1H, C13-H), 7.28 (d, J = 11.4 Hz, 1H, C15-H), 7.23 (dd, J = 4.4, 1.0 Hz, 1H, C3-H), 6.50 (s, 1H, C7-H), 3.75 (m, 1H, C19-H), 3.72 (s, 3H, C10-H₃), 2.78 (d, J = 0.8 Hz, 3H, C9-H₃), 2.76 (t, J = 7.0 Hz, 2H, C22-H₂), 1.78 – 1.62 (m, 4H, C20-H₂, C21-H₂), 1.33 (d, J = 6.3 Hz, 3H, C18-H₃).

¹³C NMR (100 MHz, CD₃OD) δ 155.2 (d, J = 252.9 Hz, C16), 150.7 (C6), 145.8 (C2), 144.4 (C4), 144.3 (C8a*), 144.1 (d, J = 9.8 Hz, C11), 134.7 (C5*), 129.5 (C13), 128.2 (C8*), 126.0 (C3), 124.6 (C4a), 120.2 (d, J = 23.3 Hz, C15), 119.1 (d, J = 5.3 Hz, C12), 114.1 (d, J = 10.7 Hz, C14), 94.5 (C7), 57.4 (C10), 49.0 (C19), 42.1 (C2), 35.0 (C20), 29.1 (C21), 23.5 (C9), 20.9 (C18).

IR (ATR, ν_{max} (cm⁻¹)) 3380 (st N-H), 2932 (st C-H), 2857 (st C-H), 1572 (st C_{ar}-C_{ar}), 1525 (b N-H), 1467 (st C_{ar}-C_{ar}), 1403 (st C_{ar}-C_{ar}), 1262 (st C_{ar}-O), 1213 (st C-N), 929, 844 (w N-H).

MS (70 eV, EI) m/z (%): 499.1 (10%) $[M+4]^+$, 497.2 (36%) $[M+2]^+$, 496.2 (8%) $[M+1]^+$, 495.2 (29%) $[M]^+$, 441.1 (20%), 439.1 (76%), 437.1 (64%) $[M-C_3H_8N]^+$, 215.1 (28%) $[C_{13}H_{14}N_2O]^+$, 203.1 (51%), 188.1 (20%) $[C_{11}H_{12}N_2O]^+$, 175.1 (29%), 129.1 (28%), 126.1 (58%), 111.1 (42%), 105.1 (55%), 98.1 (51%), 97.1 (56%),

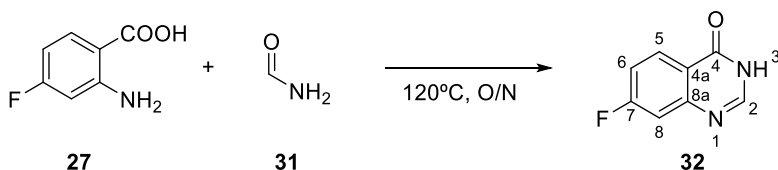
85.1 (52%) [C₅H₁₂N]⁺, 84.1 (56%) [C₅H₁₁N]⁺, 83.1 (59%), 81.1 (40%), 71.1 (68%), 69.1 (76%), 59.1 (87%), 57.1 (100%) [C₄H₉]⁺, 55.1 (95%).

HRMS (Q-TOF): m/z calculated for C₂₂H₂₄BrClFN₃O₂ [M]⁺: 495.0724 Found: 496.0782 [M+H]⁺, 498.0760 [M+H+2]⁺, 500.0750 [M+H+4]⁺

mp: 88-92 °C

5.2.3. Synthesis of quinazoline derivatives

5.2.3.1. Synthesis of 7-fluoroquinazolin-4(3H)-one



10.00g (64.5 mmol, 1.0 eq.) of 2-amino-4-fluorobenzoic acid (**27**) were added to a 50 ml round-bottomed flask along with 22.2 ml (10.0 eq.) of formamide (**31**). The mixture was heated and stirred at 120 °C and left reacting overnight. The reaction was checked by TLC ensuring the completely disappearance of the starting materials (**27**). Later, the mixture was cooled to room temperature and poured into ice-cold water with the subsequent formation of a pale brown precipitate. This was filtered, washed three times with water and dried *in vacuo* over P₂O₅ to afford 8.18 g of 7-fluoroquinazolin-4(3H)-one (**32**, **LMF314**) as pale-brown solid (77% yield).

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 12.34 (s, 1H, N3-H), 8.18 (dd, *J* = 8.8, 6.4 Hz, 1H, C5-H), 8.13 (s, 1H, C2-H), 7.45 (dd, *J* = 10.1, 2.5 Hz, 1H, C8-H), 7.39 (td, *J* = 8.7, 2.6 Hz, 1H, C6-H).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 165.6 (d, *J* = 251.0 Hz, C7), 160.1 (C4), 150.9 (C8a), 146.9 (C2), 129.0 (d, *J* = 10.9 Hz, C5), 119.6 (C4a), 115.3 (d, *J* = 23.5 Hz, C6), 112.3 (d, *J* = 21.4 Hz, C8).

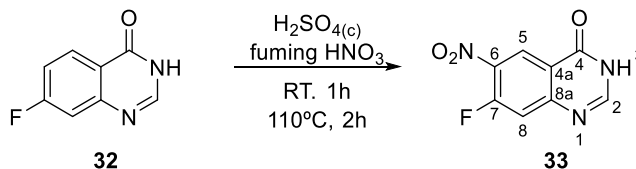
IR (ATR), ATR (cm⁻¹) 3180 (*st* N-H), 3042, 2891 (*st* C-H), 1683 (*st* C=O), 1617, 1450, 1252 (*st* C-N), 1127, 890 (*oop b* C_{ar}-H), 777.

HRMS (Q-TOF): m/z calculated for C₈H₅FN₂O [M]⁺: 164.0386. Found: 165.0460 [M+H]⁺

mp: 247-248 °C

The recorded spectroscopic data are consistent with the previously described (bibliographic yield: 66%)^{206,224}.

5.2.3.2. Synthesis of 7-fluoro-6-nitroquinazolin-4(3H)-one



10 ml of fuming nitric acid were poured dropwise over 10 ml of concentrated sulfuric acid and mixed at 0 °C in an ice-water bath. The mixture was then added dropwise to a 50 ml round-bottomed flask containing 5.00g (20.0 mmol, 1.0 eq.) of 7-fluoroquinazolin-4(3H)-one (**32**) and left stirring at room temperature for 1 hour. Later, the mixture was slowly heated to 110 °C and stirred for 2 hours turning into an orange color. The mixture was poured into 200 ml of ice-water forming a pale-yellow precipitate and the aqueous mixture was allowed to warm to room temperature and totally precipitate overnight. The precipitate was then separated from the suspension by vacuum filtration, washed three times with water and dried *in vacuo* over P₂O₅ to yield 5.84 g of 1,2-7-fluoro-6-nitroquinazolin-4(3H)-one (**33**, LMF324) as a pale-yellow solid (92% yield).

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 12.82 (s, 1H, N3-H), 8.75 (d, *J* = 8.3 Hz, 1H, C5-H), 8.32 (s, 1H, C2-H), 7.80 (d, *J* = 12.3 Hz, 1H, C8-H).

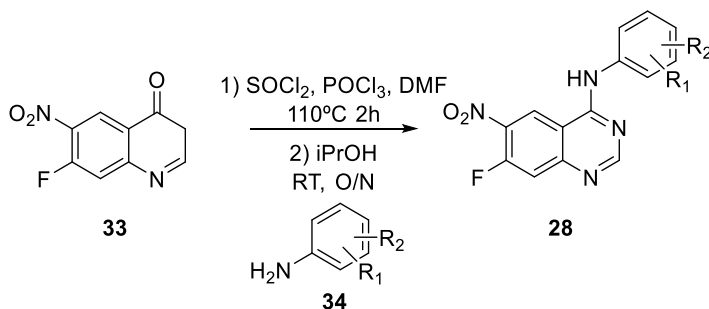
¹³C NMR (100 MHz, DMSO-*d*₆) δ 159.4 (C4), 157.6 (d, *J* = 265.6 Hz, C7), 154.1 (d, *J* = 13.6 Hz, C8a), 150.0 (C2), 135.4 (d, *J* = 9.8 Hz, C6), 125.6 (C4a), 119.3 (d, *J* = 2.3 Hz, C5), 115.6 (d, *J* = 23.5 Hz, C8).

IR (ATR), *v*_{max} (cm⁻¹) 3183 (*st* N-H), 3021, 2892, 1672 (*st* C=O), 1625, 1606 (*st* C_{ar}-C_{ar}), 1571 (*st asym* NO₂), 1521, 1342, 1326 (*st sym* NO₂), 1280, 1262 (*st* C-N), 937, 874 (*oop b* C_{ar}-H).

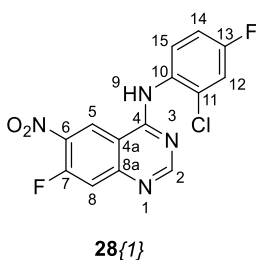
HRMS (Q-TOF): *m/z* calculated for C₈H₄FN₃O₃ [M]⁺: 209.0237. Found: 210.0310 [M+H]⁺

mp: 277-282 °C

The recorded spectroscopic data are consistent with the previously described (bibliographic yield: 86%)^{206,224}.

5.2.3.3. Synthesis of 7-fluoro-*N*-phenyl-6-nitroquinazolin-4-amines

0.95 mmol (1.0 eq.) of 7-fluoro-6-nitroquinazolin-4(3*H*)-one (**33**), 23.30 mmol (24.5 eq.) of thionyl chloride, 3.20 mmol (3.4 eq.) of phosphorus (V) oxychloride and 0.26 mmol (0.27 eq.) of dry DMF were stirred in a 10 ml round-bottomed flask and heated to 100° for 2 hours. Later the transparent yellow mixture was cooled to room temperature and the solvent removed *in vacuo* with three washes of toluene to totally remove the DMF traces. The pale yellow remaining solid corresponding to the chlorinated intermediate derivative was used further on without purification. This was suspended in 20 ml of isopropanol (iPrOH) in a 50 ml round-bottomed flask and 1.36 mmol (1.5 eq.) of the corresponding aniline (**34**) were added. The flask was set to stir at room temperature overnight becoming a suspension. The following day, the total precipitation of the desired product was favored by adding 15 ml of hexane. After 30 minutes, the suspension was filtered with suction, washed twice with hexane and dried *in vacuo* to remove the excess of isopropanol or hexane to afford the desired *N*-phenyl-7-fluoro-6-nitroquinazolin-4-amine (**28**).

***N*-(2-chloro-4-fluorophenyl)-7-fluoro-6-nitroquinazolin-4-amine (28{1})**

Starting from 7-fluoro-6-nitroquinazolin-4(3*H*)-one (**33**) and 2-chloro-4-fluoroaniline (**34{1}**), the product **28{1}** (**LMF332**) was obtained as a powdery yellow solid with 97% yield.

Spectroscopic data:

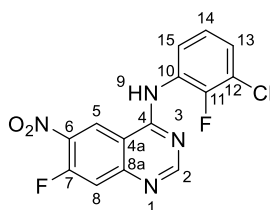
¹H NMR (400 MHz, DMSO-*d*₆) δ 11.14 (s, 1H, N9-H), 9.69 (d, *J* = 7.8 Hz, 1H, C5-H), 8.71 (s, 1H, C2-H), 7.93 (d, *J* = 12.0 Hz, 1H, C8-H), 7.67 (dd, *J* = 8.6, 2.9 Hz, 1H, C12-H), 7.60 (dd, *J* = 8.8, 5.7 Hz, 1H, C15-H), 7.37 (ddd, *J* = 8.8, 8.2, 2.9 Hz, 1H, C14-H).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 160.9 (d, *J* = 248.1 Hz, C13), 160.3 (C4), 157.4 (d, *J* = 268.1 Hz, C7), 155.6 (C2), 147.5 (C8a), 147.4 (C10), 136.4 (d, *J* = 9.9 Hz, C6), 131.9 (d, *J* = 11.7 Hz, C11), 130.8 (d, *J* = 10.2 Hz, C15), 125.6 (C5), 117.3 (d, *J* = 26.2 Hz, C12), 115.3 (d, *J* = 22.4 Hz, C14), 111.0 (d, *J* = 23.1 Hz, C8), 110.2 (C4a).

IR (ATR), ν_{max} (cm⁻¹) 2993, 2648, 2605, 1618 (*b* N-H), 1548 (*asym st* NO₂), 1491, 1434 (*st* C_{ar}-C_{ar}), 1381 (*sym st* NO₂), 1334 (*st* C_{ar}-N), 1203, 901, 877, 798 (*w* N-H), 755 (*oop b* C_{ar}-H).

HRMS (Q-TOF): *m/z* calculated for C₁₄H₇ClF₂N₄O₂ [M]⁺: 336.0226. Found: 337.0296 [M+H]⁺, 339.0274 [M+H+2]⁺

mp: 249-252 °C

***N*-(3-chloro-2-fluorophenyl)-7-fluoro-6-nitroquinazolin-4-amine (28{2})****28{2}**

Starting from 7-fluoro-6-nitroquinazolin-4(3*H*)-one (**33**) and 3-chloro-2-fluoroaniline (**34{2}**), the product **28{2}** (**LMF320**) was obtained as a powdery yellow solid with 99% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 11.05 (s, 1H, N9-H), 9.67 (d, *J* = 7.9 Hz, 1H, C5-H), 8.75 (s, 1H, C2-H), 7.93 (d, *J* = 12.1 Hz, 1H, C8-H), 7.59 (ddd, *J* = 8.3, 6.8, 1.7 Hz, 1H, C15-H), 7.52 (ddd, *J* = 8.3, 6.8, 1.7 Hz, 1H, C13-H), 7.35 (td, *J* = 8.1, 8.1, 1.4 Hz, 1H, C14-H).

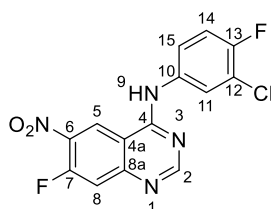
¹³C NMR (100 MHz, DMSO-*d*₆) δ 157.0 (d, *J* = 266.3 Hz, C7), 156.9 (C2), 156.9 (C4), 152.3 (d, *J* = 250.2 Hz, C11), 150.6 (C8a), 136.1 (C6), 136.0 (d, *J* = 10.0 Hz, C10), 128.8 (C13), 127.0 (C15), 125.3 (C14), 125.2 (d, *J* = 3.3 Hz, C5), 120.4 (d, *J* = 16.4 Hz, C12), 117.1 (C4a), 113.0 (d, *J* = 21.9 Hz, C8).

IR (ATR), ν_{max} (cm^{-1}) 3041 (*st* N-H), 2913, 2456, 1624 (*b* N-H), 1575 (*asym st* NO_2), 1480, 1439 (*st* $\text{C}_{\text{ar}}-\text{C}_{\text{ar}}$), 1365 (*sym st* NO_2), 1346 (*st* $\text{C}_{\text{ar}}-\text{N}$), 1050, 901 (*w* N-H), 778 (*oop b* $\text{C}_{\text{ar}}-\text{H}$).

HRMS (Q-TOF): m/z calculated for $\text{C}_{14}\text{H}_7\text{ClF}_2\text{N}_4\text{O}_2$ $[\text{M}]^+$: 336.0226. Found: 337.0298 $[\text{M}+\text{H}]^+$, 339.0274 $[\text{M}+\text{H}+2]^+$

mp: 245-249 °C

***N*-(3-chloro-4-fluorophenyl)-7-fluoro-6-nitroquinazolin-4-amine (28{3})**



28{3}

Starting from 7-fluoro-6-nitroquinazolin-4(3*H*)-one (**33**) and 3-chloro-4-fluoroaniline (**34{3}**), the product **28{3}** (**LMF339**) was obtained as an orange solid with 77% yield.

Spectroscopic data:

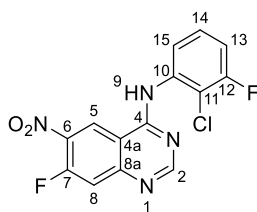
^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ 11.31 (s, 1H, N9-H), 9.76 (d, $J = 7.8$ Hz, 1H, C5-H), 8.86 (s, 1H, C2-H), 8.11 (dd, $J = 6.8, 2.6$ Hz, 1H, C11-H), 7.93 (d, $J = 12.0$ Hz, 1H, C8-H), 7.79 (ddd, $J = 9.0, 4.3, 2.6$ Hz, 1H, C15-H), 7.53 (t, $J = 9.1$ Hz, 1H, C14-H).

^{13}C NMR (100 MHz, $\text{DMSO}-d_6$) δ 158.8 (C4), 157.0 (d, $J = 266.8$ Hz, C7), 156.7 (C2), 154.6 (d, $J = 245.4$ Hz, C13), 150.0 (C8a), 136.0 (d, $J = 9.7$ Hz, C6), 134.7 (C10), 125.4 (C11), 125.1 (C12), 124.9 (C5), 124.1 (d, $J = 7.3$ Hz, C15), 116.8 (d, $J = 21.9$ Hz, C14), 112.7 (d, $J = 21.6$ Hz, C8), 110.9 (C4a).

IR (ATR), ν_{max} (cm^{-1}) 2995, 2547, 1619 (*b* N-H), 1576 (*asym st* NO_2), 1496, 1440 (*st* $\text{C}_{\text{ar}}-\text{C}_{\text{ar}}$), 1379 (*sym st* NO_2), 1330 (*st* $\text{C}_{\text{ar}}-\text{N}$), 1262, 1206, 801 (*w* N-H), 783 (*oop b* $\text{C}_{\text{ar}}-\text{H}$).

HRMS (Q-TOF): m/z calculated for $\text{C}_{14}\text{H}_7\text{ClF}_2\text{N}_4\text{O}_2$ $[\text{M}]^+$: 336.0226. Found: 337.0299 $[\text{M}+\text{H}]^+$, 339.0274 $[\text{M}+\text{H}+2]^+$

mp: 232-238 °C

***N*-(2-chloro-3-fluorophenyl)-7-fluoro-6-nitroquinazolin-4-amine (28{4})****28{4}**

Starting from 7-fluoro-6-nitroquinazolin-4(3*H*)-one (**33**) and 2-chloro-3-fluoroaniline (**34{4}**), the product **28{4}** (**LMF349**) was obtained as a powdery pale-yellow solid with 92% yield.

Spectroscopic data:

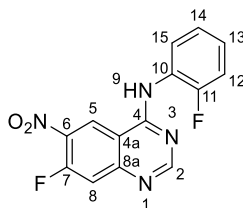
¹H NMR (400 MHz, DMSO-*d*₆) δ 11.23 (s, 1H, N9-H), 9.65 (d, *J* = 7.8 Hz, 1H, C5-H), 8.68 (s, 1H, C2-H), 7.91 (d, *J* = 12.0 Hz, 1H, C8-H), 7.51 (m, C14-H), 7.45 (m, C15-H), 7.39 (d, *J* = 7.8 Hz, 1H, C13-H).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 159.5 (C4), 158.1 (d, *J* = 246.7 Hz, C7), 157.4 (d, *J* = 267.6 Hz, C12), 155.6 (C2), 148.2 (s, C8a), 136.9 (C10), 136.3 (d, *J* = 9.8 Hz, C6), 128.6 (d, *J* = 9.0 Hz, C14), 125.6 (C5), 124.7 (C15), 118.1 (d, *J* = 16.8 Hz, C11), 115.6 (d, *J* = 21.2 Hz, C13), 111.5 (d, *J* = 22.5 Hz, C8), 110.6 (C4a).

IR (ATR), ν_{max} (cm⁻¹) 3030 (*st* N-H), 2456, 1616 (*b* N-H), 1570 (*asym st* NO₂), 1551, 1464, 1427 (*st* C_{ar}-C_{ar}), 1376 (*sym st* NO₂), 1337 (*st* C_{ar}-N), 986, 926 (*w* N-H), 790 (*oop b* C_{ar}-H), 776.

HRMS (Q-TOF): *m/z* calculated for C₁₄H₇ClF₂N₄O₂ [M]⁺: 336.0226. Found: 337.0296 [M+H]⁺, 339.0273 [M+H+2]⁺

mp: 276-280 °C

7-fluoro-*N*-(2-fluorophenyl)-6-nitroquinazolin-4-amine (28{5})**28{5}**

Starting from 7-fluoro-6-nitroquinazolin-4(3*H*)-one (**33**) and 2-fluoroaniline (**34{5}**), the product **28{5}** (**LMF351**) was obtained as a yellow solid with 96% yield.

Spectroscopic data:

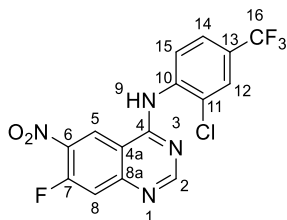
¹H NMR (400 MHz, DMSO-*d*₆) δ 11.80 (s, 1H, N9-H), 9.82 (d, *J* = 7.8 Hz, 1H, C5-H), 8.83 (s, 1H, C2-H), 7.99 (d, *J* = 11.8 Hz, 1H, C8-H), 7.56 (td, *J* = 7.8, 1.6 Hz, 1H, C14-H), 7.47 – 7.39 (m, 2H, C15-H, C12-H), 7.33 (td, *J* = 7.4, 1.8 Hz, 1H, C13-H).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 160.0 (C4), 157.3 (d, *J* = 267.5 Hz, C7), 155.9 (C2), 156.6 (d, *J* = 248.7 Hz, C11), 148.1 (C8a), 136.2 (d, *J* = 9.8 Hz, C6), 129.2 (d, *J* = 7.8 Hz, C15), 128.4 (C14), 125.6 (C5), 124.8 (d, *J* = 3.6 Hz, C13), 124.4 (d, *J* = 12.4 Hz, C10), 116.3 (d, *J* = 19.7 Hz, C12), 111.3 (d, *J* = 22.5 Hz, C8), 110.4 (C4a).

IR (ATR), ν_{max} (cm⁻¹) 3037 (*st* N-H), 29893, 2460, 1621 (*b* N-H), 1572 (*asym st* NO₂), 1498, 1434 (*st* C_{ar}-C_{ar}), 1376 (*sym st* NO₂), 1343 (*st* C_{ar}-N), 1253, 1045, 777 (*w* N-H), 757 (*oop b* C_{ar}-H).

HRMS (Q-TOF): *m/z* calculated for C₁₄H₈F₂N₄O₂ [M]⁺: 302.0615. Found: 303.0680 [M+H]⁺

mp: 233-236 °C

***N*-(2-chloro-4-(trifluoromethyl)phenyl)-7-fluoro-6-nitroquinazolin-4-amine (28{6})****28{6}**

Starting from 7-fluoro-6-nitroquinazolin-4(3*H*)-one (**33**) and 2-chloro-4-(trifluoromethyl)aniline (**34{6}**), the product **28{6}** (**LMF331**) was obtained as a powdery pale-yellow solid with 61% yield.

Spectroscopic data:

¹H NMR (400 MHz, CD₃OD) δ 9.62 (dd, *J* = 7.4, 0.4 Hz, 1H, C5-H), 8.87 (s, 1H, C2-H), 8.02 (dq, *J* = 1.6, 0.8 Hz, 1H, C12-H), 7.88 (dd, *J* = 10.8, 0.4 Hz, 1H, C8), 7.83 (d, *J* = 1.6 Hz, 2H, C14-H, C15-H)

¹³C NMR (100 MHz, CD₃OD) δ 162.8 (C3), 160.4 (d, *J* = 272.2 Hz, C7), 155.3 (C2), 145.3 (C8a), 139.2 (C6), 138.5 (C10), 133.4 (C11), 132.9 (q, *J* = 33.6 Hz, C13), 131.2 (C15), 128.6 (q, *J* = 4.0 Hz, C12), 126.5 (C5), 126.1 (q, *J* = 3.7 Hz, C14), 124.5 (q, *J* = 272.3 Hz, C16), 111.4 (d, *J* = 2.7 Hz, C4a), 110.6 (d, *J* = 25.0 Hz, C8).

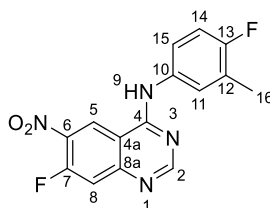
¹⁹F NMR (376 MHz, CD₃OD) δ -65.10 (s, CF₃), -107.49 (dd, *J* = 10.6, 7.3 Hz, F).

IR (ATR), ν_{\max} (cm^{-1}) 3028 (*st* N-H), 2895, 1619 (*b* N-H), 1570, 1550 (*asym st* NO_2), 1431 (*st* $\text{C}_{\text{ar}}\text{-C}_{\text{ar}}$), 1374 (*sym st* NO_2), 1323 (*st* $\text{C}_{\text{ar}}\text{-N}$), 1262, 1175, 1125 (*st* CF_3), 1078, 1038, 811 (*w* N-H), 707 (*oop b* $\text{C}_{\text{ar}}\text{-H}$).

HRMS (Q-TOF): m/z calculated for $\text{C}_{15}\text{H}_7\text{ClF}_4\text{N}_4\text{O}_2$ $[\text{M}]^+$: 386.0194. Found: 387.0261 $[\text{M}+\text{H}]^+$, 389.0240 $[\text{M}+\text{H}+2]^+$

mp: 234-238 °C

7-fluoro-N-(4-fluoro-3-methylphenyl)-6-nitroquinazolin-4-amine (28{7})



28{7}

Starting from 7-fluoro-6-nitroquinazolin-4(3*H*)-one (**33**) and 4-fluoro-3-methylaniline (**34{7}**), the product **28{7}** (**LMF350**) was obtained as a yellow solid with 82% yield.

Spectroscopic data:

^1H NMR (400 MHz, $\text{DMSO-}d_6$) δ 11.49 (s, 1H, N9-H), 9.79 (d, $J = 7.7$ Hz, 1H, C5-H), 8.85 (s, 1H, C2-H), 7.94 (d, $J = 11.9$ Hz, 1H, C8-H), 7.61 (dt, $J = 8.2, 4.0$ Hz, 1H, C15-H), 7.58-7.55 (m, 1H, C11-H), 7.26 (t, $J = 9.1$ Hz, 1H, C14-H), 2.29 (d, $J = 1.8$ Hz, 3H, C16-H3).

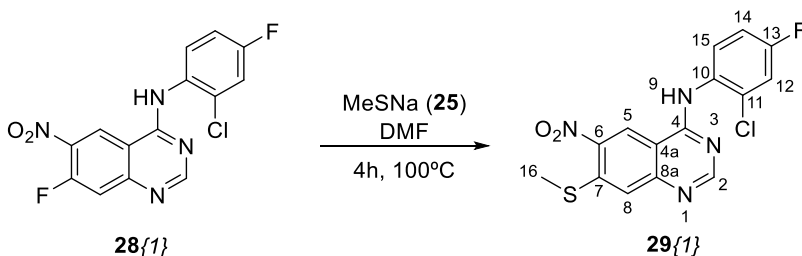
^{13}C NMR (100 MHz, $\text{DMSO-}d_6$) δ 159.0 (C4), 158.5 (d, $J = 242.6$ Hz, C13), 157.1 (d, $J = 267.3$ Hz, C7), 156.0 (C2), 148.4 (C8a), 136.0 (d, $J = 10.0$ Hz, C6), 132.9 (d, $J = 3.1$ Hz, C10), 127.2 (d, $J = 5.1$ Hz, C15), 125.4 (C5), 124.6 (d, $J = 18.4$ Hz, C12), 123.5 (d, $J = 8.3$ Hz, C11), 115.2 (d, $J = 23.6$ Hz, C14), 111.4 (d, $J = 22.5$ Hz, C8), 110.7 (d, $J = 2.1$ Hz, C4a), 14.3 (d, $J = 3.1$ Hz, C1).

IR (ATR), ν_{\max} (cm^{-1}) 2989 (*st* C-H), 2637, 1616 (*b* N-H), 1577 (*asym st* NO_2), 1495, 1382 (*st* $\text{C}_{\text{ar}}\text{-C}_{\text{ar}}$), 1363 (*sym st* NO_2), 1330 (*st* $\text{C}_{\text{ar}}\text{-N}$), 1208, 902, 802 (*w* N-H), 772 (*oop b* $\text{C}_{\text{ar}}\text{-H}$).

HRMS (Q-TOF): m/z calculated for $\text{C}_{15}\text{H}_{10}\text{F}_2\text{N}_4\text{O}_2$ $[\text{M}]^+$: 386.0194. Found: 387.0261 $[\text{M}+\text{H}]^+$

mp: 256-258 °C

5.2.3.4. Synthesis of *N*-(2-chloro-4-fluorophenyl)-7-(methylthio)-6-nitroquinazolin-4-amine (**29{1}**)



516 mg (1.53 mmol, 1.0 eq.) of *N*-(2-chloro-4-fluorophenyl)-7-fluoro-6-nitroquinazolin-4-amine (**28{1}**) were dissolved in 2.5 ml of DMF in a 10 ml round-bottomed flask and stirred at room temperature. After total dissolution of reagent **28{1}**, 434.4 mg (6.13 mmol, 4.0 eq.) of sodium methanethiolate (**25**) were added and the mixture was set to stir at 100 °C for 4 hours. The solvent was removed under reduced pressure and the crude was washed three times with toluene to ensure the total removal of DMF. The crude was later partitioned with ethyl acetate and a saturated solution of sodium bicarbonate. The orange organic phases were finally washed with brine, dried with anhydrous sodium sulphate and the solvent was removed under reduced pressure to afford 467.2 mg of the product **29{1}** (**LMF369**) which was obtained as a smelly yellow solid with 86% yield.

Spectroscopic data:

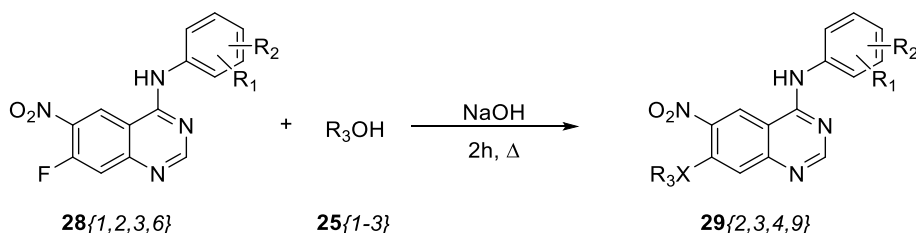
¹H NMR (400 MHz, DMSO-*d*₆) δ 10.55 (s, 1H, N9-H), 9.57 (s, 1H, C5-H), 8.54 (s, 1H, C2-H), 7.68 (s, 1H, C8-H), 7.65 – 7.56 (m, 2H, C12-H, C15-H), 7.34 (td, *J* = 8.4, 2.7 Hz, 1H, C14-H), 2.63 (s, 3H, C16-H₃).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 160.4 (d, *J* = 247.6 Hz, C13), 159.9 (C4), 158.6 (C2), 151.8 (C8a), 142.8 (C6), 142.1 (C7), 132.2 (d, *J* = 10.7 Hz, C11), 131.8 (C10), 131.1 (d, *J* = 10.1 Hz, C15), 124.0 (C8), 123.4 (C5), 117.1 (d, *J* = 26.2 Hz, C12), 115.0 (d, *J* = 22.0 Hz, C14), 110.5 (C4a), 15.6 (C16).

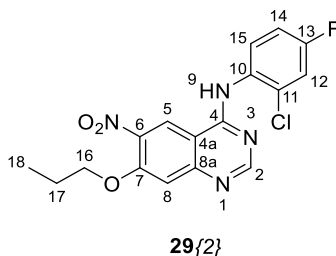
IR (ATR), ν_{max} (cm⁻¹) 3399 (*st* N-H), 3028 (*st* C_{ar}-H), 2918 (*st* C-H), 1609, 1558 (*asym st* NO₂), 1526, 1493, 1406 (*st* C_{ar}-N), 1346 (*sym st* NO₂), 1328, 1305, 1118, 1078, 893, 795 (*oop b* C_{ar}-H).

HRMS (Q-TOF): *m/z* calculated for C₁₅H₁₀ClFN₄O₂S [M]⁺: 364.0197. Found: 365.0260 [M+H]⁺, 367.0240 [M+H+2]⁺

mp: 230-233 °C

5.2.3.5. Synthesis of 7-alkoxy-*N*-phenyl-6-nitroquinazolin-4-amines (method A)

0.30 mmol (1.0 eq.) of the substrate **28**{1,2,3,6} were dissolved in an excess of 10 ml of the desired alcohol **25**{1-2} in a 10 ml round-bottomed flask. 3.00 mmol (10.0 eq.) of sodium hydroxide were later carefully added and the mixture was refluxed at the alcohol's boiling point temperature for 2 hours. The reaction was checked by TLC ensuring the completely disappearance of the starting material (**28**). Later, the crude was precipitated over a solution of saturated sodium bicarbonate, the solid was collected by filtration, washed with water three times and dried *in vacuo* to afford the desired 7-alkoxy-*N*-phenyl-6-nitroquinazolin-4-amine (**29**).

***N*-(2-chloro-4-fluorophenyl)-6-nitro-7-propoxyquinazolin-4-amine (**29**{2})**

Starting from *N*-(2-chloro-4-fluorophenyl)-7-fluoro-6-nitroquinazolin-4-amine (**28**{1}) and 1-propanol (**25**{1}), the product **29**{2} (**LMF401**) was obtained as a dark brown solid with 80% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 10.19 (s, 1H, N9-H), 9.12 (s, 1H, C5-H), 8.41 (s, 1H, C2-H), 7.61 – 7.50 (m, 2H, C12-H, C15-H), 7.40 (s, 1H, C8-H), 7.29 (td, $J = 8.5, 2.9$ Hz, 1H, C14-H), 4.25 (t, $J = 6.3$ Hz, 2H, C16-H₂), 1.80 (sextet, $J = 7.0$ Hz, 2H, C17-H₂), 1.02 (t, $J = 7.4$ Hz, 3H, C18-H₃).

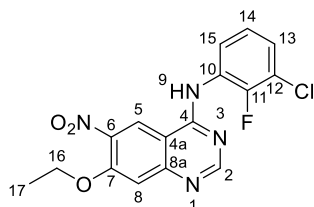
¹³C NMR (100 MHz, DMSO-*d*₆) δ 159.9 (d, $J = 246.0$ Hz, C13), 159.0 (C4), 157.6 (C2), 154.0 (C7), 153.3 (C8a), 138.6 (C6), 131.6 (d, $J = 11.6$ Hz, C11), 130.5 (d, $J = 8.8$ Hz, C15), 126.2 (C10), 122.1 (C5), 116.9 (d, $J = 25.8$ Hz, C12), 114.8 (d, $J = 22.1$ Hz, C14), 109.8 (C8), 108.2 (C4a), 71.0 (C16), 21.7 (C17), 10.3 (C18).

IR (ATR), ν_{max} (cm^{-1}) 3371 (*st* N-H), 3077 (*st* C_{ar} -H), 2970 (*st* C-H), 2877, 1624, 1570, 1525 (*asym st* NO_2), 1495, 1415, 1357 (*sym st* NO_2), 1331, 1232, 1186, 849, 794 (*oop b* C_{ar} -H).

HRMS (Q-TOF): m/z calculated for $\text{C}_{17}\text{H}_{14}\text{ClFN}_4\text{O}_3$ $[\text{M}]^+$: 376.0738. Found: 377.0800 $[\text{M}+\text{H}]^+$, 379.078 $[\text{M}+\text{H}+2]^+$

mp: 154-156 °C

***N*-(3-chloro-2-fluorophenyl)-7-ethoxy-6-nitroquinazolin-4-amine (29{3})**



29{3}

Starting from *N*-(3-chloro-2-fluorophenyl)-7-fluoro-6-nitroquinazolin-4-amine (**28{2}**) and ethanol (**25{2}**), the product **29{3}** (**LMF323**) was obtained as a powdery yellow solid with 98% yield.

Spectroscopic data:

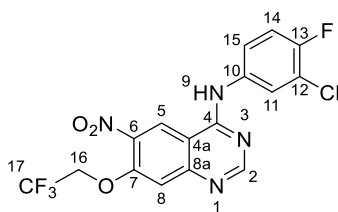
^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ 10.12 (s, 1H, N9-H), 8.93 (s, 1H, C5-H), 8.24 (s, 1H, C2-H), 7.34 (ddd, $J = 8.0, 7.1, 1.7$ Hz, 1H, C15-H), 7.29 – 7.20 (m, 1H, C13-H), 7.19 (s, 1H, C8-H), 7.14 (td, $J = 8.0, 1.3$ Hz, 1H, C14-H), 4.29 (q, $J = 6.9$ Hz, 2H, C16- H_2), 1.39 (t, $J = 7.0$ Hz, 3H, C17- H_3).

^{13}C NMR (100 MHz, $\text{DMSO}-d_6$) δ 158.5 (C4), 158.2 (C2), 154.2 (C8a), 153.7 (C7), 151.6 (d, $J = 246.6$ Hz, C11), 143.8 (C10), 137.2 (C6), 125.7 (d, $J = 2.4$ Hz, C15), 124.5 (C14), 124.5 (C13), 123.0 (C5), 119.8 (d, $J = 16.80$ Hz, C12), 110.7 (C4a), 108.9 (C8), 65.2 (C16), 14.2 (C17).

IR (ATR), ν_{max} (cm^{-1}) 3384 (*asym st* NH), 2989 (*st* C-H), 1629 (*ring st* C_{ar} - C_{ar}), 1571 (*asym st* NO_2), 1457, 1355, 1326 (*sym st* NO_2), 1234 (*st* C-N), 1105, 1023, 937, 848, 765 (*oop b* C_{ar} -H).

HRMS (Q-TOF): m/z calculated for $\text{C}_{16}\text{H}_{12}\text{ClFN}_4\text{O}_3$ $[\text{M}]^+$: 362.0582. Found: 363.0640 $[\text{M}+\text{H}]^+$, 365.062 $[\text{M}+\text{H}+2]^+$

mp: 201-203 °C

***N*-(3-chloro-4-fluorophenyl)-6-nitro-7-(2,2,2-trifluoroethoxy)quinazolin-4-amine (29{4})****29{4}**

Starting from *N*-(3-chloro-4-fluorophenyl)-7-fluoro-6-nitroquinazolin-4-amine (**28{3}**) and 2,2,2-trifluoroethan-1-ol (**25{3}**), the product **29{4}** (**LMF355**) was obtained as a yellow solid after its purification by automated flash chromatography (silica column, Cy:AcOEt-0.4%TEA 1:1) with 69% yield.

Spectroscopic data:

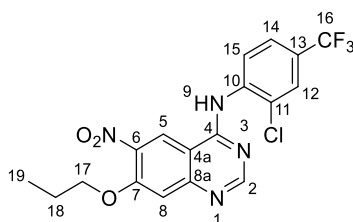
¹H NMR (400 MHz, DMSO-*d*₆) δ 10.24 (s, 1H, N9-H), 9.30 (s, 1H, C5-H), 8.71 (s, 1H, C2-H), 8.15 (dd, *J* = 6.8, 2.6 Hz, 1H, C11-H), 7.79 (ddd, *J* = 9.0, 4.3, 2.7 Hz, 1H, C15-H), 7.67 (s, 1H, C8-H), 7.47 (t, *J* = 9.1 Hz, 1H, C14-H), 5.17 (q, *J* = 8.6 Hz, 2H, C16-H₂).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 157.9 (C4), 157.7 (C2), 153.7 (d, *J* = 243.9 Hz, C13), 153.2 (C8a), 151.9 (C7), 138.4 (C6), 135.8 (d, *J* = 3.1 Hz, C10), 124.0 (C11), 123.5 (q, *J* = 277.4 Hz, C17), 122.8 (d, *J* = 6.9 Hz, C15), 122.4 (C5), 118.9 (d, *J* = 18.6 Hz, C12), 116.7 (d, *J* = 21.7 Hz, C14), 111.4 (C8), 109.0 (C4a), 65.7 (q, *J* = 34.9 Hz, C16).

IR (ATR), ν_{max} (cm⁻¹) 3390 (*st* N-H), 3090, 3052 (*st* C_{ar}-H), 1623 (*b* N-H), 1572, 1520 (*asym st* NO₂), 1495 (*st* C_{ar}-C_{ar}), 1415, 1355 (*sym st* NO₂), 1330 (*st* C_{ar}-N), 1286 (*st* CF₃), 1229, 1157 (*st* C-O), 1106, 978, 845 (*w* N-H), 805, 747 (*oop b* C_{ar}-H).

HRMS (Q-TOF): *m/z* calculated for C₁₆H₉ClF₄N₄O₃ [M]⁺: 416.0299. Found: 417.0370 [M+H]⁺, 419.0350 [M+H+2]⁺

mp: 214-215 °C

N*-(2-chloro-4-(trifluoromethyl)phenyl)-6-nitro-7-propoxyquinazolin-4-amine **29{9}***29{9}**

Starting from *N*-(2-chloro-4-(trifluoromethyl)phenyl)-7-fluoro-6-nitroquinazolin-4-amine (**28{6}**) and 1-propanol (**25{1}**), the product **29{9}** (**LMF340**) was obtained as a yellow slightly greyish powder with 88% yield.

Clear carbon assignment was unsuccessfully achieved as there was not easy spin relaxation even when raising the system temperature until 101 °C. Bidimensional NMR spectra did not show clear cross-signals for a proper assignment.

Spectroscopic data:

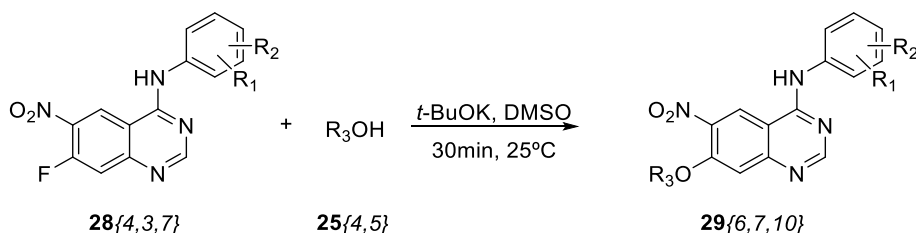
¹H NMR (400 MHz, 101 °C, DMSO-*d*₆) δ 10.38 (s, 1H, N9-H), 8.98 (s, 1H, C5-H), 8.37 (s, 1H, C2-H), 7.83 (d, *J* = 2.0 Hz, 1H, C12-H), 7.77 – 7.65 (m, 2H, C14-H, C15-H), 7.38 (s, 1H, C8-H), 4.26 (t, *J* = 6.3 Hz, 2H, C17-H₂), 1.84 (qt, *J* = 7.3, 6.3 Hz, 2H, C18-H₂), 1.04 (t, *J* = 7.4 Hz, 3H, C19-H₃).

¹³C NMR (100 MHz, 101 °C, DMSO-*d*₆) δ 208.0, 155.8 (C2*), 140.6 (C10*), 128.4 – 128.1 (C12), 126.3 (C15*), 126.0 (C5*), 123.5 (C14), 111.0 (C8*), 72.6 (C17), 23.2 (C18), 11.5 (C19).

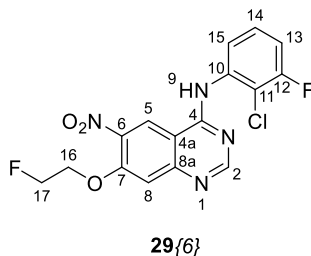
IR (ATR), ν_{max} (cm⁻¹) 3421 (*st* N-H), 2970, 2879 (*st* C-H), 1623 (*b* N-H), 1535 (*asym st* NO₂), 1416 (*st* C_{ar}-C_{ar}), 1322 (*sym st* NO₂), 1309 (*st* C_{ar}-N), 1176, 1112 (*st* CF₃), 1076 (*st* C-O), 1054, 964, 913, 850 (*w* N-H).

HRMS (Q-TOF): *m/z* calculated for C₁₈H₁₄ClF₃N₄O₃ [M]⁺: 426.0707. Found: 427.0770 [M+H]⁺, 429.0750 [M+H+2]⁺

mp: 212-214 °C

5.2.3.6. Synthesis of 7-alkoxy-*N*-phenyl-6-nitroquinazolin-4-amines (method B)

A suspension of 1.1 mmol (1.0 eq.) of the substrate **28**{3,4,7} and 1.65 mmol of the desired alcohol **25**{4,5} in 5 ml of DMSO was prepared in a 10 ml round-bottomed flask and stirred at room temperature. After adding 3.3 mmol (3.0 eq.) of sodium *tert*-butoxide, the reaction was set to stir at 25 °C for 30 minutes. Later, sufficient water was added to quench the reaction and ensure the complete precipitation of the product. This solid was collected by filtration, washed with water three times and dried *in vacuo* to afford the desired 7-alkoxy-*N*-phenyl-6-nitroquinazolin-4-amine (**29**).

***N*-(2-chloro-3-fluorophenyl)-7-(2-fluoroethoxy)-6-nitroquinazolin-4-amine (29{6})**

Starting from *N*-(2-chloro-3-fluorophenyl)-7-fluoro-6-nitroquinazolin-4-amine (**28**{4}) and 2-fluoroethan-1-ol (**25**{4}), the product **29**{6} (**LMF364**) was obtained as a brown solid with 54% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 10.59 (s, 1H, N9-H), 9.18 (s, 1H, C5-H), 8.47 (s, 1H, C2-H), 7.52 (s, 1H, C8-H), 7.49 – 7.41 (m, 1H, C14-H), 7.41 – 7.34 (m, 2H, C13-H, C15-H), 4.82 (dt, *J* = 40.8, 3.5 Hz, 2H, C17-H₂), 4.60 (dt, *J* = 22.9, 3.7 Hz, 2H, C16-H₂).

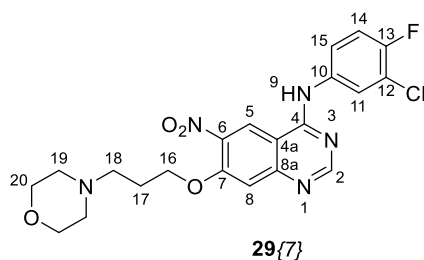
¹³C NMR (100 MHz, DMSO-*d*₆) δ 158.1 (d, *J* = 245.7 Hz, C12), 157.0 (C4), 153.6 (C2), 152.8 (C7), 146.7 (C8a), 138.8 (C6), 134.3 (C10), 128.2 (d, *J* = 9.2 Hz, C14), 124.5 (C15), 122.2 (C5), 117.7 (d, *J* = 21.6 Hz, C11), 114.3 (d, *J* = 22.7 Hz, C13), 110.3 (C8), 108.4 (C4a), 81.6 (d, *J* = 167.4 Hz, C17), 69.2 (d, *J* = 18.9 Hz, C16).

IR (ATR), ν_{\max} (cm^{-1}) 3377 (*st* N-H), 3060, (*st* C_{ar} -H), 2981 (*st* C-H), 1625 (*b* N-H), 1571 (*st* C_{ar} - C_{ar}), 1522 (*asym st* NO_2), 1472 (*b* C-H), 1446, 1360 (*sym st* NO_2), 1330 (*st* C_{ar} -N), 1252 (*w* CH_2F), 1226 (*st* C-O), 933, 773 (*w* N-H).

HRMS (Q-TOF): m/z calculated for $\text{C}_{16}\text{H}_{11}\text{ClF}_2\text{N}_4\text{O}_3$ $[\text{M}]^+$: 380.0488. Found: 381.0550 $[\text{M}+\text{H}]^+$, 383.0530 $[\text{M}+\text{H}+2]^+$

mp: 139-142 °C

***N*-(3-chloro-4-fluorophenyl)-7-(3-morpholinopropoxy)-6-nitroquinazolin-4-amine (29{7})**



Starting from *N*-(3-chloro-4-fluorophenyl)-7-fluoro-6-nitroquinazolin-4-amine (**28{3}**) and 3-morpholinopropan-1-ol (**25{5}**), the product **29{7}** (**LMF402**) was obtained as an other solid with 85% yield.

Spectroscopic data:

^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ 10.13 (s, 1H, N9-H), 9.20 (s, 1H, C5-H), 8.65 (s, 1H, C2-H), 8.15 (dd, $J = 6.9, 2.4$ Hz, 1H, C11-H), 7.79 (m, 1H, C15-H), 7.51 – 7.38 (m, 2H, C7-H, C14-H), 4.33 (t, $J = 6.1$ Hz, 2H, C16-H₂), 3.57 (t, $J = 4.5$ Hz, 4H, C20-H₂, C20'-H₂), 2.45 (t, $J = 7.1$ Hz, 2H, C18-H₂), 2.37 (s br., 4H, C19-H₂, C19'-H₂), 1.93 (p, $J = 6.6$ Hz, 2H, C17-H₂).

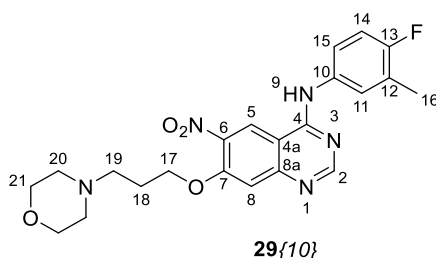
^{13}C NMR (100 MHz, $\text{DMSO}-d_6$) δ 157.8 (C4), 157.3 (C2), 153.9 (C7), 153.60 (d, $J = 243.7$ Hz, C13), 153.3 (C8a), 139.5 (C10), 138.9 (C6), 123.9 (d, $J = 6.8$ Hz, C11), 122.6 (d, $J = 6.9$ Hz, C15), 121.8 (C5), 118.9 (d, $J = 18.5$ Hz, C12), 116.7 (d, $J = 21.7$ Hz, C14), 110.2 (C8), 107.9 (C4a), 68.0 (C16), 66.2 (C20, C20'), 54.5 (C18), 53.4 (C19, C19'), 25.4 (C17).

IR (ATR), ν_{\max} (cm^{-1}) 3296 (*st* N-H), 2954, 2870, 2811 (*st* C-H), 1624 (*b* N-H), 1568 (*st* C_{ar} - C_{ar}), 1495 (*asym st* NO_2), 1420 (*b* C-H), 1343 (*sym st* NO_2), 1310, 1230 (*st* C_{ar} -N), 1115 (*st* C-O), 865, 800 (*w* N-H), 779 (*w* N-H), 757.

HRMS (Q-TOF): m/z calculated for $\text{C}_{21}\text{H}_{21}\text{ClFN}_5\text{O}_4$ $[\text{M}]^+$: 461.1266. Found: 462.1320 $[\text{M}+\text{H}]^+$, 464.1310 $[\text{M}+\text{H}+2]^+$

mp: 123-126 °C

***N*-(4-fluoro-3-methylphenyl)-7-(3-morpholinopropoxy)-6-nitroquinazolin-4-amine (29{10})**



Starting from 7-fluoro-*N*-(4-fluoro-3-methylphenyl)-6-nitroquinazolin-4-amine (**28{7}**) and 3-morpholinopropan-1-ol (**25{5}**), the product **29{10}** (**LMF358**) was obtained as an ochre solid with 80% yield.

Spectroscopic data:

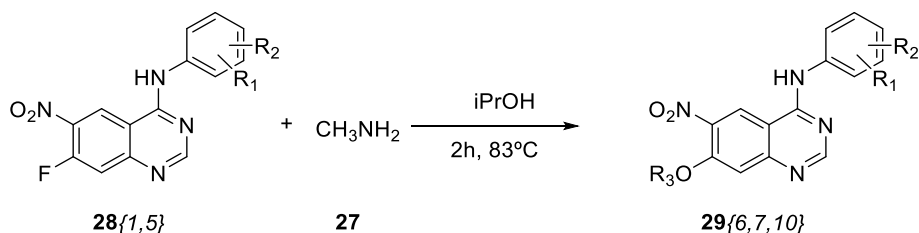
¹H NMR (400 MHz, DMSO-*d*₆) δ 10.05 (s, 1H, N9-H), 9.23 (s, 1H, 1H, C5-H), 8.60 (s, 1H, C2-H), 7.67 – 7.59 (m, 2H, C11-H, C15-H), 7.45 (s, 1H, C8-H), 7.18 (t, *J* = 9.1 Hz, 1H, C14-H), 4.34 (t, *J* = 6.1 Hz, 2H, C17-H₂), 3.57 (t, *J* = 4.6 Hz, 4H, C21-H₂, C21'-H₂), 2.45 (t, *J* = 7.1 Hz, 2H, C19-H₂), 2.37 (s br., 4H, C20-H₂, C20'-H₂), 2.27 (d, *J* = 2.0 Hz, 3H, C16-H₃), 1.94 (p, *J* = 6.5 Hz, 2H, C18-H₂).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 158.1 (C4), 157.6 (C2), 157.4 (d, *J* = 240.4 Hz, C13), 153.8 (C7), 153.3 (C8a), 138.8 (C6), 134.5 (d, *J* = 2.9 Hz, C10), 125.7 (d, *J* = 4.8 Hz, C11), 124.1 (d, *J* = 18.2 Hz, C12), 122.0 (d, *J* = 8.0 Hz, C15), 121.8 (C5), 114.8 (d, *J* = 23.1 Hz, C14), 110.1 (C8), 107.9 (C4a), 67.9 (C17), 66.2 (C21, C21'), 54.5 (C19), 53.4 (C20, C20'), 25.4 (C18), 14.4 (d, *J* = 3.1 Hz, C16).

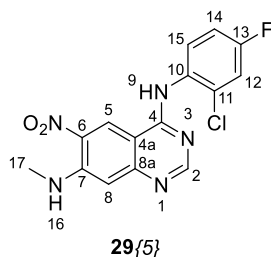
IR (ATR), ν_{max} (cm⁻¹) 3399 (*st* N-H), 2957, 2846, 2813 (*st* C-H), 1622 (*b* N-H), 1574 (*st* C_{ar}-C_{ar}), 1518 (*asym st* NO₂), 1500, 1423 (*b* C-H), 1353 (*sym st* NO₂), 1323 (*st* C_{ar}-N), 1229 (*w* C-H), 1116, 1103 (*st* C-O), 801 (*w* N-H).

HRMS (Q-TOF): *m/z* calculated for C₂₂H₂₄FN₅O₄ [M]⁺: 441.1812. Found: 442.1870 [M+H]⁺

mp: 170-172 °C

5.2.3.7. Synthesis of N^4 -phenyl-6-nitroquinazolin-4,6-diamines

A solution of 1.5 mmol (1.0 eq.) of the substrate **28**{1,5} was prepared in 10 ml of isopropanol in a 50 ml round-bottomed flask and stirred at room temperature. After adding 2.98 mmol (2.0 eq.) of methylamine, the reaction was refluxed at 83°C for 2 hours. Later, sufficient water was added to quench the reaction and ensure the complete precipitation of the product. This solid was collected by filtration, washed with water three times and dried *in vacuo* to afford the desired 7-alkoxy- N -phenyl-6-nitroquinazolin-4-amine (**29**). Later, the crude was precipitated over a solution of saturated sodium bicarbonate, the solid was collected by filtration, washed with water and was purified by automated flash chromatography (silica column, Cy:AcOEt-0,2% TEA, 1:1) to afford the desired N^4 -phenyl-6-nitroquinazolin-4,6-diamine (**29**).

 N^4 -(2-chloro-4-fluorophenyl)- N^7 -methyl-6-nitroquinazoline-4,7-diamine (**29**{5})

Starting from N -(2-chloro-4-fluorophenyl)-7-fluoro-6-nitroquinazolin-4-amine (**28**{1}) and methylamine (**27**), the product **29**{5} (**LMF356**) was obtained as a crystalline needle-shaped red solid with 65% yield.

Spectroscopic data:

^1H NMR (400 MHz, $\text{DMSO}-d_6$) δ 10.26 (s, 1H, N9-H), 9.45 (s, 1H, C5-H), 8.30 (s, 1H, C2-H), 7.98 (d, $J = 5.1$ Hz, 1H, N16-H), 7.57 (m, 2H, C12-H, C15-H), 7.30 (td, $J = 8.3, 2.8$ Hz, 1H, C14-H), 6.90 (s, 1H, C8-H), 2.99 (d, $J = 4.8$ Hz, 3H, C17-H3).

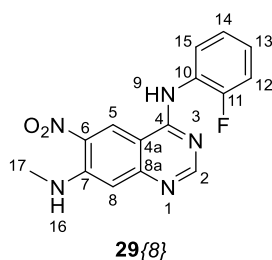
^{13}C NMR (100 MHz, DMSO- d_6) δ 160.2 (d, J = 246.5 Hz, C13), 159.9 (C4), 158.3 (C2), 153.4 (C7), 146.7 (C8a), 133.0 (C6), 132.2 (C10), 131.3 (d, J = 9.3 Hz, C15), 124.8 (C5), 116.9 (d, J = 25.8 Hz, C12), 114.8 (d, J = 22.2 Hz, C14), 106.6 (C8), 104.6 (C4a), 29.9 (C17).

IR (ATR), ν_{max} (cm^{-1}) 3420 (st N-H), 2990 (st C-H), 2164, 1623, 1564, 1528 (asym st NO_2), 1502, 1415, 1359 (sym st NO_2), 1231, 846, 793 (oop b C_{ar} -H).

HRMS (Q-TOF): m/z calculated for $\text{C}_{15}\text{H}_{11}\text{ClFN}_5\text{O}_2$ $[\text{M}]^+$: 347.0585. Found: 348.0650 $[\text{M}+\text{H}]^+$, 350.0620 $[\text{M}+\text{H}+2]^+$

mp: 243-245 °C

N^4 -(2-fluorophenyl)- N^7 -methyl-6-nitroquinazoline-4,7-diamine (29{8})



Starting from 7-fluoro- N -(2-fluorophenyl)-6-nitroquinazolin-4-amine (**28{5}**) and methylamine (**27**), the product **29{8}** (**LMF365**) was obtained as a crystalline orange solid with 66% yield.

Spectroscopic data:

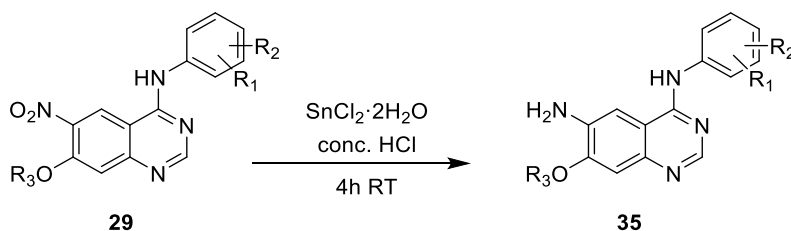
^1H NMR (400 MHz, DMSO- d_6) δ 10.23 (s, 1H, $\text{N}9^{\text{*}}\text{-H}$), 9.45 (s, 1H, $\text{N}16^{\text{*}}\text{-H}$), 8.34 (s, 1H, C5-H), 7.99 (s, 1H, C2-H), 7.52 (t, J = 7.3 Hz, 1H, $\text{C}14^{\text{**}}\text{-H}$), 7.37 – 7.21 (m, 3H, $\text{C}15^{\text{**}}\text{-H}$, C12-H, C13-H), 6.89 (s, 1H, C8-H), 2.99 (d, J = 4.9 Hz, 3H, $\text{C}17\text{-H}_3$).

^{13}C NMR (100 MHz, DMSO- d_6) δ 159.5 (C4), 158.3 (C2), 156.6 (d, J = 240.7 Hz, C11), 153.5 (C8a), 146.7 (C7), 133.0 (C6), 128.5 (C14 *), 127.6 (C15 *), 125.0 (C5), 124.4 (d, J = 3.6 Hz, C13), 116.6 (d, J = 19.8 Hz, C10), 116.0 (d, J = 20.5 Hz, C10), 106.6 (C8), 104.8 (C4a), 29.9 (C17).

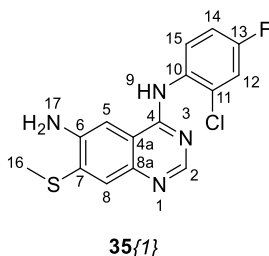
IR (ATR), ν_{max} (cm^{-1}) 3420 (st N-H), 3074 (st C_{ar} -H), 2946 (st C-H), 1621 (b N-H), 1555 (st $\text{C}_{\text{ar}}\text{-C}_{\text{ar}}$), 1536 (asym st NO_2), 1489 (b C-H), 1400 (st $\text{C}_{\text{ar}}\text{-C}_{\text{ar}}$), 1360 (rock C-H), 1333 (sym st NO_2), 1260, 1235 (st $\text{C}_{\text{ar}}\text{-N}$), 749 (w N-H).

HRMS (Q-TOF): m/z calculated for $\text{C}_{15}\text{H}_{12}\text{FN}_5\text{O}_2$ $[\text{M}]^+$: 313.0975. Found: 314.1040 $[\text{M}+\text{H}]^+$

mp: 208-210 °C

5.2.3.8. Synthesis of 7-substituted-*N*⁴-phenyl-6-nitroquinazolin-4,6-diamines

1.00 mmol (1.0 eq.) of the corresponding 7-substituted-*N*(2-fluorophenyl)-6-nitroquinazolin-4-amine (**29**) were placed in a 50 ml round-bottomed flask with 5.00 mmol (5.0 eq.) of $\text{SnCl}_2 \cdot 2\text{H}_2\text{O}$ along with a stirring bar. 10 ml of concentrated hydrochloric acid (37%) were added dropwise until total solution was observed (sonication was facilitated in some of the cases) and then the reaction was stirred for 4 hours at room temperature. Later, the mixture was slowly basified with a 2M NaOH solution in an iced-water bath to mitigate the exothermic reaction until pH = 14. A clear color variation was observed during alkalization. The crude was later extracted with three fractions of chloroform. The organic layers were unified, dried with MgSO_4 and concentrated under reduce pressure to afford the 7-substituted-*N*⁴-phenyl-6-nitroquinazolin-4,6-diamine derivatives (**35**).

***N*⁴-(2-chloro-4-fluorophenyl)- 7-(methylthio)-6-nitroquinazolin-4,6-diamine (**35{1}**)**

Starting from *N*-(2-chloro-4-fluorophenyl)-7-(methylthio)-6-nitroquinazolin-4-amine (**29{1}**), the product **35{1}** (**LMF372**) was obtained as a smelly pale-yellow solid with 92% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 9.29 (s, 1H, N9-H), 8.21 (s, 1H, C2-H), 7.64 (dd, *J* = 8.9, 5.9 Hz, 1H, C15-H), 7.55 (dd, *J* = 8.6, 2.9 Hz, 1H, C12-H), 7.45 (s, 1H, C5-H), 7.37 (s, 1H, C8-H), 7.28 (td, *J* = 8.6, 2.9 Hz, 1H, C14-H), 5.39 (s, 2H, N17-H₂), 2.59 (s, 3H, C16-H₃).

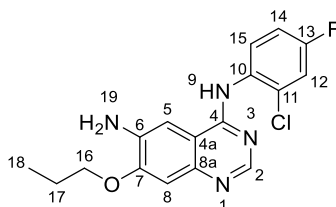
¹³C NMR (100 MHz, DMSO-*d*₆) δ 159.5 (d, *J* = 245.5 Hz, C13), 157.0 (C4), 150.8 (C2), 143.9 (C8a), 143.2 (C7), 133.4 (C6), 133.3 (d, *J* = 3.4 Hz, C10), 131.4 (d, *J* = 11.1 Hz, C11), 130.3 (d, *J* = 9.0 Hz, C15), 123.2 (C5), 116.7 (d, *J* = 25.9 Hz, C12), 114.6 (d, *J* = 22.0 Hz, C14), 113.3 (C4a), 101.6 (C8), 14.4 (C16).

IR (ATR), ν_{\max} (cm^{-1}) 3466 (*asym st* NH), 3369, 3301, 3141(*sym st* NH), 2917 (*st* C-H), 1630 (*sc* N-H), 1523 (*st* C_{ar}-C_{ar}), 1482, 1414, 1183 (*st* C-N), 889 (*oop b* C_{ar}-H), 825 (*w* N-H).

HRMS (Q-TOF): m/z calculated for C₁₅H₁₂ClFN₄S [M]⁺: 334.0455. Found: 335.0520 [M+H]⁺, 337.0490 [M+H+2]⁺

mp: 201-205 °C

N⁴-(2-chloro-4-fluorophenyl)-6-nitro-7-propoxyquinazolin-4,6-diamine (35{2})



35{2}

Starting from *N*-(2-chloro-4-fluorophenyl)-6-nitro-7-propoxyquinazolin-4-amine (**29{2}**), the product **35{2}** (**LMF403**) was obtained as an off white solid with 95% yield.

Spectroscopic data:

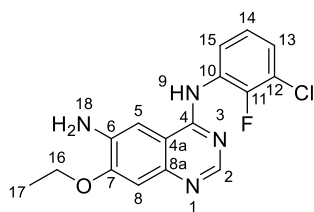
¹H NMR (400 MHz, DMSO-*d*₆) δ 9.09 (s, 1H, N₉-H), 8.18 (s, 1H, C₂-H), 7.63 (dd, $J = 8.8, 6.0$ Hz, 1H, C₁₅-H), 7.53 (dd, $J = 8.6, 2.8$ Hz, 1H, C₁₂-H), 7.32 (s, 1H, C₅-H), 7.27 (td, $J = 8.6, 2.8$ Hz, 1H, C₁₄-H), 7.05 (s, 1H, C₈-H), 5.31 (s, 2H, N₁₉-H₂), 4.11 (t, $J = 6.4$ Hz, 2H, C₁₆-H₂), 1.84 (*sextet*, $J = 7.1$ Hz, 2H, C₁₇-H₂), 1.10 (t, $J = 7.4$ Hz, 3H, C₁₈-H₃)

¹³C NMR (100 MHz, DMSO-*d*₆) δ 159.4 (d, $J = 244.9$ Hz, C₁₃), 156.4 (C₄), 152.0 (C₇), 150.7 (C₂), 144.6 (C_{8a}), 138.4 (C₆), 133.6 (C₁₀), 131.33 (d, $J = 10.7$ Hz, C₁₁), 130.2 (d, $J = 9.1$ Hz, C₁₅), 116.6 (d, $J = 25.8$ Hz, C₁₂), 114.5 (d, $J = 21.9$ Hz, C₁₄), 109.9 (C_{4a}), 106.3 (C₈), 100.9 (C₅), 69.6 (C₁₆), 21.9 (C₁₇), 10.5 (C₁₈).

IR (ATR), ν_{\max} (cm^{-1}) 3477, 3428 (*asym st* NH), 3293, 3149 (*sym st* NH), 1530 (*sc* N-H), 1512 (*st* C_{ar}-C_{ar}), 1450, 1270 (*st* C_{ar}-N), 1209 (*st* C-O), 836 (*oop b* C_{ar}-H).

HRMS (Q-TOF): m/z calculated for C₁₇H₁₆ClFN₄O [M]⁺: 346.0997. Found: 347.1013 [M+H]⁺, 349.0985 [M+H+2]⁺

mp: 211-215 °C

***N*⁴-(3-chloro-2-fluorophenyl)-7-ethoxy-6-nitroquinazolin-4,6-diamine (35{3})****35{3}**

Starting from *N*-(3-chloro-2-fluorophenyl)-7-ethoxy-6-nitroquinazolin-4-amine (**29{3}**), the product **35{3}** (**LMF341**) was obtained as a white solid with 95% yield.

Spectroscopic data:

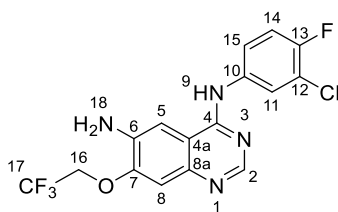
¹H NMR (400 MHz, DMSO-*d*₆) δ 9.29 (s, N9-H), 8.24 (s, C2-H), 7.54 (ddd, *J* = 8.4, 7.0, 1.7 Hz, C13*-H), 7.41 (ddd, *J* = 8.2, 6.7, 1.6 Hz, C15*-H), 7.31 (s, C5-H), 7.23 (td, *J* = 8.1, 1.4 Hz, C14-H), 7.07 (s, C8-H), 5.36 (s, N18-H₂), 4.22 (q, *J* = 6.9 Hz, C16-H₂), 1.44 (t, *J* = 6.9 Hz, C17-H₃).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 155.8 (C4), 152.1 (d, *J* = 248.7 Hz, C11), 152.0 (C8a), 150.6 (C2), 144.8 (C3), 138.6 (C6), 129.2 (d, *J* = 11.7 Hz, C10), 126.3 (*J* = 10.8 Hz, C15, C13), 124.7 (d, *J* = 4.6 Hz, C14), 120.0 (d, *J* = 16.7 Hz, C12), 110.1 (C4a), 106.2 (C5), 100.9 (C6), 63.8 (C16), 14.4 (C17).

IR (ATR), *v*_{max} (cm⁻¹) 3413 (*asym st* NH), 3250 (*asym st* NH₂), 3154 (*sym st* NH₂), 2986 (*st* C-H), 1610, 1575, 1509 (*sc* N-H), 1440 (*st* C_{ar}-C_{ar}), 1388, 1261 (*st* C_{ar}-N), 1207 (*st* C-O), 1040, 934, 823 (*oop b* C_{ar}-H), 756 (*w* N-H).

HRMS (Q-TOF): *m/z* calculated for C₁₆H₁₄ClFN₄O [M]⁺: 332.084. Found: 333.0900 [M+H]⁺, 335.0880 [M+H+2]⁺

mp: 225-228 °C

***N*⁴-(3-chloro-4-fluorophenyl)-6-nitro-7-(2,2,2-trifluoroethoxy)quinazolin-4,6-diamine (35{4})****35{4}**

Starting from *N*-(3-chloro-4-fluorophenyl)-6-nitro-7-(2,2,2-trifluoroethoxy)quinazolin-4-amine (**29{4}**), the product **35{4}** (**LMF359**) was obtained as a pale-yellow solid with 94% yield.

Spectroscopic data:

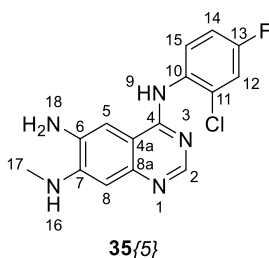
¹H NMR (400 MHz, DMSO-*d*₆) δ 8.64 (s, 1H, N9-H), 7.57 (s, 1H, C2-H), 7.35 (dd, *J* = 6.9, 2.6 Hz, 1H, C11-H), 6.97 (ddd, *J* = 8.9, 4.3, 2.7 Hz, 1H, C15-H), 6.63 (s, 1H, C8-H), 6.57 (t, *J* = 9.1 Hz, 1H, C14-H), 6.47 (s, 1H, C5-H), 4.53 (s, 2H, N18-H₂), 4.17 (q, *J* = 8.8 Hz, 2H, C16-H₂).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 155.2 (C4), 152.8 (d, *J* = 242.1 Hz, C13), 150.6 (C2), 150.0 (C7), 144.3 (C8a), 138.1 (C6), 137.3 (d, *J* = 3.0 Hz, C10), 123.9 (q, *J* = 277.7 Hz, C17), 122.7 (C11), 121.6 (d, *J* = 6.6 Hz, C15), 118.6 (d, *J* = 18.2 Hz, C12), 116.4 (d, *J* = 21.5 Hz, C14), 111.4 (C4a), 107.9 (C5), 101.9 (C8), 64.9 (q, *J* = 34.7 Hz, C16).

IR (ATR), ν_{max} (cm⁻¹) 3412 (*asym st* NH), 3312 (*sym st* NH), 1611 (*sc* N-H), 1576 (*st* C_{ar}-C_{ar}), 1495 (*st* C_{ar}-C_{ar}), 1431, 1401, 1215 (*st* C-N), 1167 (*st* C-O), 1148, 840 (*oop b* C_{ar}-H), 798 (*w* N-H).

HRMS (Q-TOF): *m/z* calculated for C₁₆H₁₁ClF₄N₄O [M]⁺: 386.0558. Found: 387.0620 [M+H]⁺, 389.0590 [M+H+2]⁺

mp: 190-193 °C

***N*⁴-(2-chloro-4-fluorophenyl)-*N*⁷-methyl-6-nitroquinazoline-4,6,7-triamine (35{5})**

Starting from *N*⁴-(2-chloro-4-fluorophenyl)-*N*⁷-methyl-6-nitroquinazoline-4,7-diamine (**29{5}**), the product **35{5}** (**LMF360**) was obtained as light pale-yellow solid with 99% yield.

Spectroscopic data:

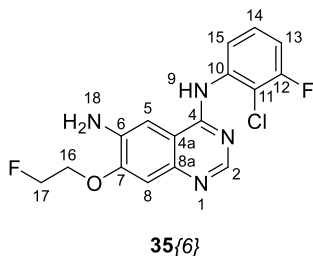
¹H NMR (400 MHz, DMSO-*d*₆) δ 8.77 (s, 1H, N9-H), 8.12 (s, 1H, C2-H), 7.71 (dd, *J* = 8.9, 5.9 Hz, 1H, C15-H), 7.51 (dd, *J* = 8.7, 2.9 Hz, 1H, C12-H), 7.25 (ddd, *J* = 9.0, 8.2, 3.0 Hz, 1H, C14-H), 7.20 (s, 1H, C5-H), 6.53 (s, 1H, C8-H), 5.85 (q, *J* = 4.7 Hz, 1H, N16-H), 5.13 (s, N18-H₂), 2.86 (d, *J* = 4.7 Hz, C17-H₃).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 158.9 (d, *J* = 244.9 Hz, C13), 155.7 (C4), 150.6 (C3), 146.1 (C8a), 143.8 (C7), 136.3 (C6), 133.8 (d, *J* = 3.2 Hz, C10), 130.4 (d, *J* = 10.8 Hz, C11), 129.5 (d, *J* = 9.1 Hz, C15), 116.5 (d, *J* = 25.7 Hz, C12), 114.4 (d, *J* = 21.9 Hz, C14), 106.6 (C4a), 101.9 (C8), 100.9 (C5), 29.7 (C17).

IR (ATR), *v*_{max} (cm⁻¹) 3442 (*asym st* NH), 3325, 3290 (*sym st* NH), 3073 (*st* C_{ar}-H), 2927 (*st* C-H), 1619 (*sc* N-H), 1573, 1528 (*st* C_{ar}-C_{ar}), 1440 (*b* C-H), 1186 (*st* C-N), 848 (*oop b* C_{ar}-H), 683 (*w* N-H).

HRMS (Q-TOF): *m/z* calculated for C₁₅H₁₃ClFN₅ [M]⁺: 317.0844. Found: 318.0900 [M+H]⁺, 320.0880 [M+H+2]⁺

mp: 244-246 °C

***N*⁴-(2-chloro-3-fluorophenyl)-7-(2-fluoroethoxy)-6-nitroquinazolin-4,6-diamine (35{6})**

Starting from *N*-(2-chloro-3-fluorophenyl)-7-(2-fluoroethoxy)-6-nitroquinazolin-4-amine (**29{6}**), the product **35{6}** (**LMF370**) was obtained as an ocher-brownish solid with 80% yield.

Spectroscopic data:

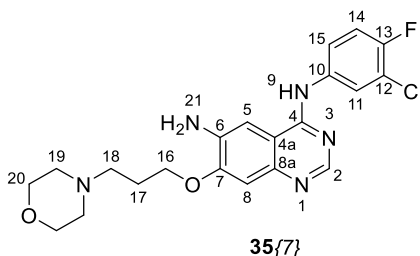
¹H NMR (400 MHz, DMSO-*d*₆) δ 9.23 (s, 1H, N9-H), 8.24 (s, 1H, C2-H), 7.53 (d, 1H, *J* = 8.1, 1.7 Hz C15-H), 7.44 – 7.35 (m, 1H, C14-H), 7.34 (s, 1H, C8-H), 7.28 (td, *J* = 8.7, 1.6 Hz, 1H, C13-H), 7.14 (s, 1H, C5-H), 5.42 (s, 2H, N18-H₂), 4.93 – 4.76 (m, 2H, C17-H₂), 4.49 – 4.37 (m, 2H, C16-H₂).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 158.0 (d, *J* = 245.1 Hz, C12), 156.0 (C4), 151.6 (C7), 150.6 (C2), 144.6 (C8a), 139.0 (C10), 138.6 (C6), 127.8 (d, *J* = 9.4 Hz, C14), 123.5 (d, *J* = 3.0 Hz, C15), 116.9 (d, *J* = 17.6 Hz, C11), 112.9 (d, *J* = 21.0 Hz, C13), 110.5 (C4a), 106.7 (C5), 101.0 (C8), 82.0 (d, *J* = 166.5 Hz, C17), 67.8 (d, *J* = 19.2 Hz, C16).

IR (ATR), ν_{max} (cm⁻¹) 3445 (*asym st* NH), 3315 (*sym st* NH), 1600 (*sc* N-H), 1532, 1512, 1466 (*st* C_{ar}-C_{ar}), 1438 (*st* C_{ar}-C_{ar}), 1399, 1249 (*st* C_{ar}-N), 1200 (*st* C-O), 1045 (*st* C-N), 984, 774 (*w* N-H), 699 (*oop b* C_{ar}-H).

HRMS (Q-TOF): *m/z* calculated for C₁₆H₁₃ClF₂N₄O [M]⁺: 350.0746. Found: 351.0800 [M+H]⁺, 353.0780 [M+H+2]⁺

mp: 180-185 °C

N⁴-(3-chloro-4-fluorophenyl)-7-(3-morpholinopropoxy)-6-nitroquinazolin-4,6-diamine (35{7})

Starting from *N*-(3-chloro-4-fluorophenyl)-7-(3-morpholinopropoxy)-6-nitroquinazolin-4-amine (**29{7}**), the product **35{7}** (**LMF404**) was obtained as a yellow solid with 82% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 9.38 (s, 1H, N9-H), 8.37 (s, 1H, C2-H), 8.19 (dd, *J* = 6.9, 2.6 Hz, 1H, C11-H), 7.81 (ddd, *J* = 8.8, 4.2, 2.7 Hz, 1H, C15-H), 7.42 – 7.36 (m, 2H, C5-H, C14-H), 7.08 (s, 1H, C8-H), 5.35 (s, 2H, N21-H₂), 4.20 (t, *J* = 6.2 Hz, 2H, C16-H₂), 3.58 (t, *J* = 4.4 Hz, 4H, C20-H₂, C20'-H₂), 2.50 (t, *J* = 6.5 Hz, 2H, C18-H₂), 2.39 (s br., 4H, C19-H₂, C19'-H₂), 1.98 (p, *J* = 6.5 Hz, 3H, C17-H₃).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 155.0 (C4), 152.6 (d, *J* = 241.7 Hz, C13), 152.0 (C7), 150.3 (C2), 144.8 (C8a), 138.6 (C6), 137.6 (C10), 122.4 (C11), 121.4 (d, *J* = 6.7 Hz, C15), 118.6 (d, *J* = 18.2 Hz, C12), 116.4

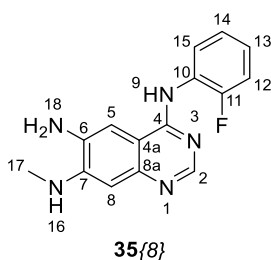
(d, $J = 21.5$ Hz, C14), 110.3 (C4a), 106.5 (C8), 100.8 (C5), 66.5 (C16), 66.2 (C20, C20'), 54.9 (C18), 53.4 (C19, C19'), 25.6 (C17).

IR (ATR), ν_{max} (cm^{-1}) 3404 (*asym st* NH), 3312 (*sym st* NH), 3098 (*st* $\text{C}_{\text{ar}}\text{-H}$), 2946 (*st* C-H), 2846, 1619 (*sc* N-H), 1584 (*st* $\text{C}_{\text{ar}}\text{-C}_{\text{ar}}$), 1534, 1512, 1495 (*st* $\text{C}_{\text{ar}}\text{-C}_{\text{ar}}$), 1444, 1212 (*st* C-O), 1114 (*st* C-N), 838, 803 (*w* N-H), 776 (*oop b* $\text{C}_{\text{ar}}\text{-H}$).

HRMS (Q-TOF): m/z calculated for $\text{C}_{21}\text{H}_{23}\text{ClFN}_5\text{O}_2$ $[\text{M}]^+$: 431.1524. Found: 432.1555 $[\text{M}+\text{H}]^+$, 434.1528 $[\text{M}+\text{H}+2]^+$

mp: 156-159 °C

N^4 -(2-fluorophenyl)- N^7 -methyl-6-nitroquinazoline-4,6,7-triamine (35{8})



Starting from N^4 -(2-fluorophenyl)- N^7 -methyl-6-nitroquinazoline-4,7-diamine (**29{8}**), the product **35{8}** (**LMF373**) was obtained as a pale-yellow solid with 95% yield.

Spectroscopic data:

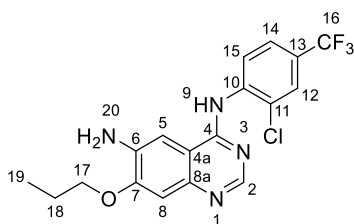
^1H NMR (400 MHz, DMSO- d_6) δ 8.81 (s, 1H, N9), 8.14 (s, 1H, C2-H), 7.65 (ddd, $J = 8.1, 5.7, 2.7$ Hz, 1H, C14-H), 7.25 (ddt, $J = 9.4, 5.8, 3.2$ Hz, 1H, C12-H), 7.21 (s, 1H, C5-H), 7.22 – 7.15 (m, 2H, C13-H, C15-H), 6.53 (s, 1H, C8-H), 5.84 (d, $J = 4.8$ Hz, 1H, N16-H), 5.11 (s, 2H, N18- H_2), 2.87 (d, $J = 4.7$ Hz, 3H, C17- H_3).

^{13}C NMR (100 MHz, DMSO- d_6) δ 156.2 (d, $J = 245.7$ Hz, C11), 155.5 (C4), 150.7 (C2), 146.2 (C7), 143.8 (C8a), 136.3 (C6), 127.9 (d, $J = 11.8$ Hz, C10), 127.0 (d, $J = 1.9$ Hz, C14), 125.4 (d, $J = 7.2$ Hz, C13), 124.1 (d, $J = 3.4$ Hz, C15), 115.6 (d, $J = 20.1$ Hz, C12), 106.9 (C4a), 101.9 (C8), 101.2 (C5), 29.7 (C17).

IR (ATR), ν_{max} (cm^{-1}) 3458 (*asym st* NH), 3347, 3239 (*sym st* NH), 3027 (*st* $\text{C}_{\text{ar}}\text{-H}$), 2808 (*st* C-H), 1614 (*sc* N-H), 1525 (*st* $\text{C}_{\text{ar}}\text{-C}_{\text{ar}}$), 1494 (*st* $\text{C}_{\text{ar}}\text{-C}_{\text{ar}}$), 1454, 1435, 1402, 1283 (*st* $\text{C}_{\text{ar}}\text{-N}$), 1256 (*st* C-N), 921, 820 (*oop b* $\text{C}_{\text{ar}}\text{-H}$), 760 (*w* N-H).

HRMS (Q-TOF): m/z calculated for $\text{C}_{15}\text{H}_{14}\text{FN}_5$ $[\text{M}]^+$: 283.1233. Found: 284.1285 $[\text{M}+\text{H}]^+$

mp: 225-227 °C

N*⁴-(2-chloro-4-(trifluoromethyl)phenyl)-6-nitro-7-propoxyquinazolin-4,6-diamine **35{9}***35{9}**

Starting from *N*-(2-chloro-4-(trifluoromethyl)phenyl)-6-nitro-7-propoxyquinazolin-4-amine (**29{9}**), the product **35{9}** (**LMF353**) was obtained as a pale-yellow solid with 97% yield.

Spectroscopic data:

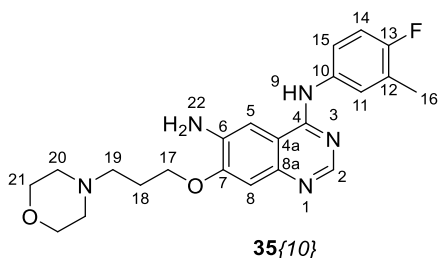
¹H NMR (400 MHz, DMSO-*d*₆) δ 9.11 (s, 1H, N9-H), 8.30 (s, 1H, C2-H), 8.13 (d, *J* = 8.5 Hz, 1H, C15-H), 7.94 (s, 1H, C12-H), 7.75 (d, *J* = 8.7 Hz, 1H, C14-H), 7.30 (s, 1H, C8-H), 7.10 (s, 1H, C5-H), 5.45 (s, 2H, N20-H₂), 4.13 (t, *J* = 6.4 Hz, 2H, C17-H₂), 1.85 (sextet, *J* = 6.5 Hz, 2H, C18-H₂), 1.06 (t, *J* = 7.4 Hz, 3H, C19-H₃).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 155.2 (C4), 152.3 (C8a), 150.2 (C2), 145.0 (C7), 140.8 (C10), 139.0 (C6), 128.4 (C11), 126.7 (C15), 126.5 (C12), 125.5 (q, *J* = 32.9 Hz, C13), 124.6 (C14), 123.6 (q, *J* = 273.4 Hz, C16), 110.4 (C4a), 106.3 (C5), 100.2 (C8), 69.7 (C17), 21.8 (C18), 10.5 (C19).

IR (ATR), ν_{max} (cm⁻¹) 3466 (*asym st* NH), 3420, 3298 (*sym st* NH), 3155 (*st* C_{ar}-H), 2963 (*st* C-H), 1609 (*sc* N-H), 1513 (*st* C_{ar}-C_{ar}), 1398 (*b* C-H), 1318 (*rock* C-H), 1268 (*st* C_{ar}-N), 1202 (*st* C-N), 1121 (*st* C-O), 1078 (*st* C-N), 965, 841 (*oop b* C_{ar}-H).

HRMS (Q-TOF): *m/z* calculated for C₁₈H₁₆ClF₃N₄O [M]⁺: 396.0965. Found: 397.1020 [M+H]⁺, 399.1000 [M+H+2]⁺

mp: 192-200 °C

***N*⁴-(4-fluoro-3-methylphenyl)-7-(3-morpholinopropoxy)-6-nitroquinazolin-4,6-diamine (35{10})**

Starting from *N*-(4-fluoro-3-methylphenyl)-7-(3-morpholinopropoxy)-6-nitroquinazolin-4-amine (**29{10}**), the product **35{10}** (**LMF371**) was obtained as an ochre-greenish solid with 91% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 9.18 (s, 1H, N9-H), 8.30 (s, 1H, C2-H), 7.69 (dd, *J* = 7.3, 2.6 Hz, 1H, C11-H), 7.62 (ddd, *J* = 9.1, 4.4, 2.6 Hz, 1H, C15-H), 7.39 (s, 1H, C5-H), 7.10 (t, *J* = 9.2 Hz, 1H, C14-H), 7.05 (s, 1H, C8-H), 5.27 (s, 2H, N22-H₂), 4.19 (t, *J* = 6.2 Hz, 2H, C17-H₂), 3.59 (t, *J* = 4.3 Hz, 4H, C21-H₂, C21'-H₂), 2.50 (t, *J* = 7.4 Hz, 2H, C19-H₂), 2.39 (s, 4H, C20-H₂, C20'-H₂), 2.25 (d, *J* = 1.8 Hz, 3H, C16-H₃), 1.98 (p, *J* = 6.5 Hz, 2H, C18-H₂).

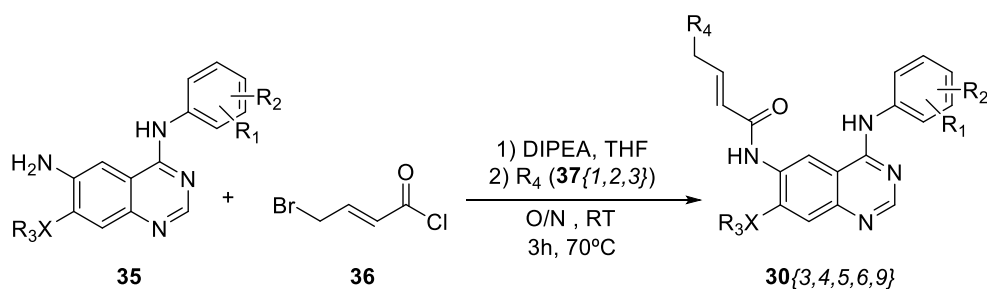
¹³C NMR (100 MHz, DMSO-*d*₆) δ 156.5 (d, *J* = 238.5 Hz, C13), 155.4 (C4), 151.8 (C7), 150.6 (C2), 144.7 (C8a), 138.3 (C6), 136.1 (d, *J* = 2.7 Hz, C10), 124.7 (d, *J* = 4.6 Hz, C11), 123.6 (d, *J* = 18.1 Hz, C12), 120.9 (d, *J* = 7.7 Hz, C15), 114.5 (d, *J* = 22.9 Hz, C14), 110.2 (C4a), 106.5 (C8), 101.1 (C5), 66.5 (C17), 66.2 (C21, C21'), 54.9 (C19), 53.4 (C20, C20'), 25.6 (C18), 14.5 (C16).

IR (ATR), ν_{max} (cm⁻¹) 3385 (*asym st* NH), 3306 (*sym st* NH), 3141 (*st* C_{ar}-H), 2814 (*st* C-H), 1630 (*sc* N-H), 1580 (*st* C_{ar}-C_{ar}), 1497 (*st* C_{ar}-C_{ar}), 1441 (*b* C-H), 1207 (*st* C-N), 1113 (*st* C-O), 847 (*w* N-H), 810 (*oop b* C_{ar}-H).

HRMS (Q-TOF): *m/z* calculated for C₂₂H₂₆FN₅O₂ [M]⁺: 411.2071. Found: 412.2130 [M+H]⁺

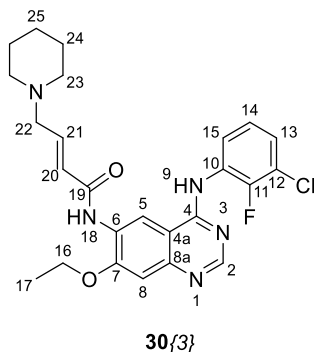
mp: 141-145 °C

5.2.3.9. Synthesis of (*E*)-*N*-(4-(phenylamino)-7substituted-quinazolin-6-yl)-4-(Het)but-2-enamide (method A)



(*E*)-4-bromobut-2-enoyl chloride (**35**) was freshly prepared before each reaction. To obtain the necessary amount of substrate, 1,30 mmol (2.0 eq.) of bromo-but-2-enoic acid were dissolved with 5 ml of anhydrous DCM. Later, 1.43 mmol (2.2 eq) of oxalyl chloride were added along with 1 drop of DMF. The solution was stirred under argon atmosphere overnight at room temperature turning the colorless transparent solution yellowish. The mixture was evaporated then to dryness yielding compound **36** as a yellow oil which was directly used.

0.65 mmol (1.0 eq.) of the corresponding **35** intermediate were placed in a 10 ml round-bottomed flask, dissolved in 5 ml of anhydrous THF and 0.6 ml of *N,N*-diisopropyl-ethylamine (DIPEA) were added along with the previously prepared reagent **35**. The mixture was stirred at room temperature for 2 hours. Later, 1.43 mmol (2.2 eq.) of the corresponding heterocycle (**37**) were added and the mixture was stirred at room temperature overnight. The following day, additional 3.7 mmol (5.7 eq.) of the compound **37** were added and the mixture was heated to 70 °C. After 3 hours, the solvents were removed *under vacuo* and the residue was extracted with ethyl acetate, washed with water and brine and dried over Na₂SO₄. The desired product was finally obtained after its purification by automated flash chromatography (silica gel 0-4% DCM: MeOH).

(E)-N-((3-chloro-2-fluorophenyl)amino)-7-ethoxyquinazolin-6-yl)-4-(piperidin-1-yl)but-2-enamide (30{3})

Starting from *N*⁴-(3-chloro-2-fluorophenyl)-7-ethoxy-6-nitroquinazolin-4,6-diamine (**35{3}**), freshly prepared (*E*)-4-bromobut-2-enoyl chloride (**36**) and piperidine (**37{1}**) the product **30{3}** (**LMF348**) was obtained as a powdery light beige solid with 27% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 9.84 (s, 1H, N9-H), 9.49 (s, 1H, N18-H), 8.93 (s, 1H, C5-H), 8.40 (s, 1H, C2-H), 7.47 (dtd, *J* = 8.3, 6.7, 1.7 Hz, 2H, C13-H, C15-H), 7.29 – 7.24 (m, 2H, C8-H, C14-H), 6.80 (dt, *J* = 15.4, 6.1 Hz, 1H, C21-H), 6.57 (d, *J* = 15.4 Hz, 1H, C20-H), 4.30 (q, *J* = 7.0 Hz, 2H, C16-H₂), 3.10 (dd, *J* = 6.1, 1.5 Hz, 2H, C22-H₂), 2.36 (m, 4H, C23-H₂, C23'-H₂), 1.51 (qt, *J* = 5.6 Hz, 4H, C24-H₂, C24'-H₂), 1.46 (t, *J* = 7.0 Hz, 3H, C17-H₃), 1.43 – 1.28 (m, 2H, C25-H₂).

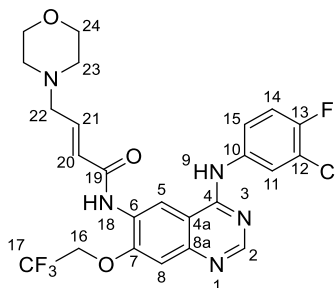
¹³C NMR (100 MHz, DMSO-*d*₆) δ 163.5 (C19), 157.5 (C4), 154.3 (C7), 154.1 (C2), 152.4 (d, *J* = 249.3 Hz, C11), 148.7 (C8a), 142.0 (C21), 128.6 (d, *J* = 11.7 Hz, C10), 127.4 (C6), 127.1 (C15*), 126.8 (C13*), 125.9 (C20), 124.8 (d, *J* = 4.55 Hz, C14), 120.0 (d, *J* = 16.5 Hz, C12), 115.5 (C5), 108.6 (C4a), 107.0 (C8), 64.5 (C16), 59.3 (C22), 54.1 (C23), 25.5 (C24), 23.9 (C25), 14.3 (C17).

IR (ATR), ν_{max} (cm⁻¹) 3456 (*st* NH), 3297 (*st* NH), 2943 (*st* C-H), 1616, 1527 (*sc* N-H), 1449 (*st* C_{ar}-C_{ar}), 1422 (*b* C-H), 1387 (*rock* C-H), 1206 (*st* C-N), 1116 (*st* C-O), 941 (*trans wag* =C-H), 776 (*oop b* C_{ar}-H).

HRMS (Q-TOF): *m/z* calculated for C₂₅H₂₇ClF_N₅O₂ [M]⁺: 483.1837. Found: 484.1890 [M+H]⁺, 486.1890 [M+H+2]⁺

mp: 198-202 °C

(E)-N-(4-((3-chloro-4-fluorophenyl)amino)-7-(2,2,2-trifluoroethoxy)quinazolin-6-yl)-4-morpholinobut-2-enamide (30{4})



30{4}

Starting from *N*⁴-(3-chloro-4-fluorophenyl)-6-nitro-7-(2,2,2-trifluoroethoxy) quinazolin-4,6-diamine (**35{4}**), freshly prepared (*E*)-4-bromobut-2-enoyl chloride (**36**) and morpholine (**37{2}**) the product **30{4}** (**LMF362**) was obtained as laminated white solid with 16% yield.

Spectroscopic data:

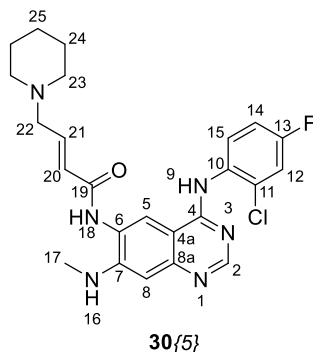
¹H NMR (400 MHz, DMSO-*d*₆) δ 9.85 (s, 1H, N9-H), 9.67 (s, 1H, N18-H), 8.81 (s, 1H, C5-H), 8.57 (s, 1H, C2-H), 8.15 (dd, *J* = 6.9, 2.6 Hz, 1H, C11-H), 7.81 (ddd, *J* = 9.1, 4.4, 2.7 Hz, 1H, C15-H), 7.50 (s, 1H, C8-H), 7.43 (t, *J* = 9.1 Hz, 1H, C14-H), 6.79 (dt, *J* = 15.4, 6.1 Hz, 1H, C21-H), 6.51 (d, *J* = 15.4 Hz, 1H, C20-H), 5.05 (q, *J* = 8.8 Hz, 2H, C16-H₂), 3.61 (t, *J* = 4.6 Hz, 4H, C24-H₂, C24'-H₂), 3.16 (dd, *J* = 6.2, 1.5 Hz, 2H, C22-H₂), 2.40 (t, *J* = 4.6 Hz, 4H, C23-H₂, C23'-H₂).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 163.7 (C19), 156.8 (C4), 154.3 (C2), 153.8 (C7), 153.27 (d, *J* = 243.0 Hz, C13), 149.0 (C8a), 141.1 (C21), 136.6 (d, *J* = 3.1 Hz, C10), 126.8 (C6), 126.0 (C20), 123.8 (q, *J* = 278.3 Hz, C17), 123.6 (C11), 122.4 (d, *J* = 6.8 Hz, C15), 118.7 (d, *J* = 18.4 Hz, C12), 118.6 (C5), 116.5 (d, *J* = 21.6 Hz, C14), 109.8 (C4a), 108.8 (C8), 66.2 (C24), 65.2 (q, *J* = 34.9 Hz, C16), 58.8 (C22), 53.3 (C23).

IR (ATR), ν_{max} (cm⁻¹) 3407 (*st* NH), 3274 (*st* NH), 2927 (*st* C-H), 2822, 1741 (*st* C=O) 1625, 1542, 1526 (*sc* N-H), 1498 (*st* C_{ar}-C_{ar}), 1457 (*b* C-H), 1424, 1275 (*st* C_{ar}-N), 1172, 1154 (*st* C-N), 1109 (*st* C-O), 961 (*trans wag* =C-H).

HRMS (Q-TOF): *m/z* calculated for C₂₄H₂₂ClF₄N₅O₃ [M]⁺: 539.1347. Found: 540.1400 [M+H]⁺, 542.1390 [M+H+2]⁺

mp: 204-207 °C

(E)-N-(4-((2-chloro-4-fluorophenyl)amino)-7-(methylamino)quinazolin-6-yl)-4-(piperidin-1-yl)but-2-enamide (30{5})

Starting from *N*⁴-(2-chloro-4-fluorophenyl)-*N*⁷-methyl-6-nitroquinazoline-4,6,7-triamine (**35{5}**), freshly prepared (*E*)-4-bromobut-2-enoyl chloride (**36**) and piperidine (**37{1}**) the product **30{5}** (**LMF361**) was obtained as crystalline ochre-brownish solid with 15% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 8.21 (s, 1H, C2-H), 8.18 (s, 1H, C5-H), 7.58 (dd, *J* = 8.9, 5.7 Hz, 1H, C15-H), 7.36 (dd, *J* = 8.4, 2.9 Hz, 1H, C12-H), 7.15 (ddd, *J* = 8.74, 8.01, 2.90, 1H, C14-H), 6.98 (dt, *J* = 15.5, 6.5 Hz, 1H, C21-H), 6.76 (s, 1H, C8-H), 6.38 (d, *J* = 15.3 Hz, 1H, C20-H), 3.21 (d, *J* = 6.5 Hz, 2H, C22-H₂), 2.94 (s, 3H, C17-H₃), 2.50 (s br., 4H, C23-H₂, C23'-H₂), 1.65 (p, *J* = 5.5 Hz, 4H, C24-H₂, C24'-H₂), 1.50 (s br., 2H, C25-H₂).

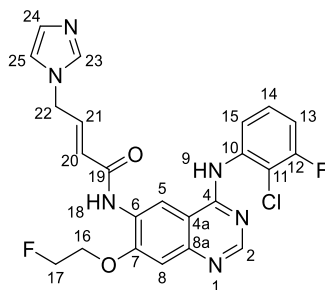
¹³C NMR (100 MHz, DMSO-*d*₆) δ 167.4 (C19), 162.2 (d, *J* = 247.5 Hz, C13), 160.1 (C4), 155.4 (C2), 151.2 (C8a), 151.0 (C7), 142.4 (C21), 133.6 (C10), 131.7 (C11), 131.6 (C15), 127.6 (C20), 125.8 (C6), 121.3 (C5), 118.1 (d, *J* = 26.0 Hz, C12), 115.7 (d, *J* = 22.2 Hz, C14), 106.5 (C4a), 103.2 (C8), 60.9 (C22), 55.6 (C23), 30.1 (C17), 26.6 (C24), 25.0 (C25).

IR (ATR), ν_{max} (cm⁻¹) 3250 (*st* NH), 2933 (*st* C-H), 2854, 1621 (*sc* N-H), 1573, 1513 (*st* C_{ar}-C_{ar}), 1460, 1419 (*st* C_{ar}-C_{ar}), 1386 (*rock* C-H), 1255 (*st* C_{ar}-N), 1209, 1186 (*st* C-N), 1110, 823 (*oop* C_{ar}-H).

HRMS (Q-TOF): *m/z* calculated for C₂₄H₂₆ClFN₆O [M]⁺: 468.9614. Found: 469.1900 [M+H]⁺, 471.1890 [M+H+2]⁺

mp: 174-180 °C

(E)-N-(4-((2-chloro-3-fluorophenyl)amino)-7-(2-fluoroethoxy)quinazolin-6-yl)-4-(1H-imidazol-1-yl)but-2-enamide (30{6})



30{6}

Starting from *N*⁴-(2-chloro-3-fluorophenyl)-7-(2-fluoroethoxy)-6-nitroquinazolin-4,6-diamine (**35{6}**), freshly prepared (*E*)-4-bromobut-2-enoyl chloride (**36**) and imidazole (**37{3}**) the product **30{6}** (**LMF374**) was obtained as an off-white slurry with 38% yield.

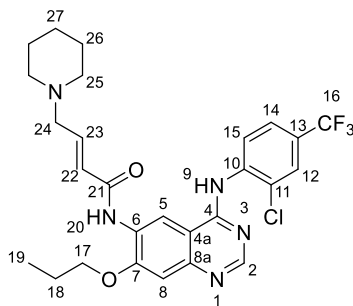
Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 10.28 (s, 1H, N₉-H), 10.12 (s, 1H, N₁₈-H), 9.31 (s, 1H, C₅-H), 8.80 (s, 1H, C₂-H), 8.12 (s, 1H, C₂₃-H), 7.90 – 7.79 (m, 2H, C₁₅-H, C₁₄-H), 7.80 – 7.71 (m, 2H, C₁₃-H, C₈-H), 7.62 (d, *J* = 1.2 Hz, 1H, C₂₅-H), 7.40 (d, *J* = 1.2 Hz, 1H, C₂₄-H), 7.37 (dt, *J* = 15.4, 4.9 Hz, 1H, C₂₁-H), 6.75 (d, *J* = 15.4 Hz, 1H, C₂₀-H), 5.32 (dd, *J* = 5.3, 1.9 Hz, 2H, C₂₂-H₂), 5.37 – 5.19 (m, 2H, C₁₇-H₂), 4.93 (dt, *J* = 22.8, 3.6 Hz, 2H, C₁₆-H₂).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 163.2 (C₁₉), 158.0 (d, *J* = 245.6 Hz, C₁₁), 157.9 (C₄), 154.4 (C₇), 154.3 (C₂), 148.8 (C_{8a}), 140.3 (C₂₁), 138.6 (C₁₀), 137.5 (C₂₃), 128.7 (C₂₄), 127.9 (d, *J* = 9.0 Hz, C₁₄), 127.1 (C₆), 125.1 (C₂₀), 124.6 (d, *J* = 2.5 Hz, C₁₅), 119.7 (C₂₅), 118.0 (d, *J* = 17.5 Hz, C₁₁), 116.5 (C₅), 113.8 (d, *J* = 21.3 Hz, C₁₃), 108.8 (C_{4a}), 107.6 (C₈), 81.9 (d, *J* = 166.9 Hz, C₁₇), 68.4 (d, *J* = 19.5 Hz, C₁₆), 46.6 (C₂₂).

IR (ATR), ν_{max} (cm⁻¹) 3222 (*st* NH), 2968 (*st* C-H), 1581, 1528 (*st* C_{ar}-C_{ar}), 1465, 1446 (*st* C_{ar}-C_{ar}), 1385 (*st* C_{ar}-N), 781 (*oop* C_{ar}-H).

HRMS (Q-TOF): *m/z* calculated for C₂₃H₁₉ClF₂N₆O₂ [M]⁺: 484.1226. Found: 185.1290 [M+H]⁺, 487.1280 [M+H+2]⁺

(E)-N-(4-((2-chloro-4-(trifluoromethyl)phenyl)amino)-7-propoxyquinazolin-6-yl)-4-(piperidin-1-yl)but-2-enamide (30{9})**30{9}**

Starting from *N*⁴-(2-chloro-4-(trifluoromethyl)phenyl)-6-nitro-7-propoxyquinazolin-4,6-diamine (**35{9}**), freshly prepared (*E*)-4-bromobut-2-enoyl chloride (**36**) and piperidine (**37{1}**) the product **30{9}** (**LMF363**) was obtained as a white solid with 31% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 9.83 (s, 1H, N9-H), 9.47 (s, 1H, N20-H), 8.94 (s, 1H, C5-H), 8.41 (s, 1H, C2-H), 7.97 (s, 1H, C12-H), 7.87 (d, *J* = 8.5 Hz, 1H, C15-H), 7.77 (d, *J* = 8.4 Hz, 1H, C14-H), 7.28 (s, 1H, C8-H), 6.84 – 6.75 (m, 1H, C23-H), 6.55 (d, *J* = 15.4 Hz, 1H, C22-H), 4.19 (t, *J* = 6.4 Hz, 2H, C24-H₂), 3.10 (t, *J* = 6.1 Hz, 2H, C17-H₂), 2.35 (s, 4H, C25-H₂, C25'-H₂), 1.86 (sextet, *J* = 6.9 Hz, 2H, C18-H₂), 1.51 (p, *J* = 5.6 Hz, C26-H₂, C26'-H₂), 1.42 – 1.36 (m, 2H, C27-H₂), 1.08 – 1.00 (m, 3H, C19-H₂).

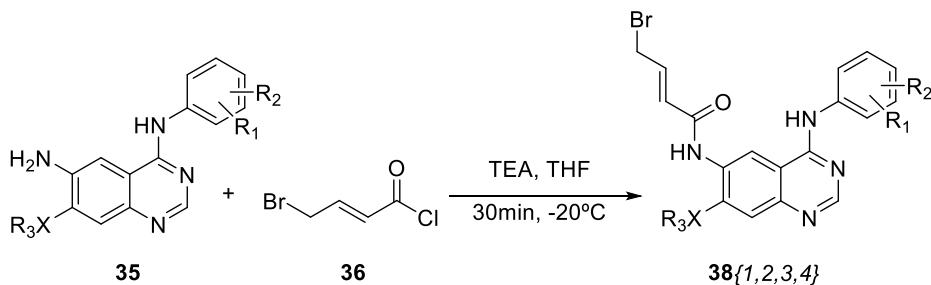
¹³C NMR (100 MHz, DMSO-*d*₆) δ 163.7 (C21), 157.5 (C4), 154.6 (C8a), 153.9 (C2), 148.9 (C7), 142.0 (C23), 140.6 (C10), 130.5 (C11), 128.9 (C15), 127.6 (C6), 126.8 (q, *J* = 30.7 Hz, C13), 126.7 (d, *J* = 2.7 Hz, C12), 125.9 (C22), 124.6 (C14), 123.5 (q, *J* = 272.3 Hz, C16), 115.5 (C5), 108.7 (C4a), 107.2 (C8), 70.3 (C17), 59.3 (C24), 54.1 (C25), 25.5 (C26), 23.9 (C27), 21.7 (C18), 10.4 (C19).

IR (ATR), ν_{max} (cm⁻¹) 3431 (*st* NH), 3285 (*st* NH), 2933 (*st* C-H), 2800, 1627 (*st* C=O), 1530 (*sc* N-H), 1424 (*st* C_{ar}-C_{ar}), 1384 (*rock* C-H), 1320 (*st* C_{ar}-N), 1308, 1276, 1173, 1114 (*st* C-O), 1078 1209, 1186 (*st* C-N), 1110, 845 (*oop* C_{ar}-H).

HRMS (Q-TOF): *m/z* calculated for C₂₇H₂₉ClF₃N₅O₂ [M]⁺: 547.1962. Found: 548.2010 [M+H]⁺, 550.2000 [M+H+2]⁺

mp: 180-184 °C

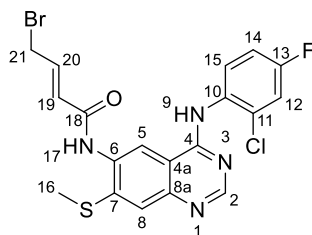
5.2.3.10. Synthesis of (*E*)-4-bromo-*N*-(4-(phenylamino)-7-substituted-quinazolin-6-yl)but-2-enamide (method B)



Firstly, (*E*)-4-bromobut-2-enoyl chloride (**35**) was freshly prepared before each reaction. To obtain the necessary amount of substrate, 1.93 mmol (1.31 eq.) of bromo-but-2-enoic acid were dissolved with 5 ml of anhydrous DCM. Later, 2.12 mmol (1.44 eq.) of oxalyl chloride were added along with 1 drop of DMF. The solution was stirred under argon atmosphere at room temperature for three hours turning the colorless transparent solution yellowish. The mixture was evaporated then to dryness yielding compound **36** as a yellow oil which was directly dissolved in 3 ml of anhydrous THF.

Secondly, 1.47 mmol (1.0 eq) of the corresponding compound **35** were dissolved in 30 ml of anhydrous THF in a 50 ml round-bottomed flask. 3.68 mmol (2.5 eq.) of triethylamine (TEA) were added and the flask was sealed under argon and cooled to -20 °C (ice-salt-water bath). The above THF solution containing the acid **36** was then added dropwise over 30 minutes and the reaction mixture was stirred for a further 30 minutes at this temperature. The mixture was diluted in 20 ml saturated NaHCO₃ solution and extracted twice with 20 ml of ethyl acetate. The combined organic fractions were washed with 20 ml brine, dried over Na₂SO₄, filtered, and the solvent was removed under reduced pressure to give a crude solid which was purified by automated flash chromatography (silica gel, 0-2% DCM: MeOH).

(E)-4-bromo-N-(4-((2-chloro-4-fluorophenyl)amino)-7-(methylthio)quinazolin-6-yl)but-2-enamide (38{1})



38{1}

Starting from *N*⁴-(2-chloro-4-fluorophenyl)-7-(methylthio)quinazoline-4,6-diamine (**35{1}**) and freshly prepared (*E*)-4-bromobut-2-enoyl chloride (**36**) the product **38{1}** (**LMF411**) was obtained as a slightly yellowish-white cotton-like solid with 44% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 9.88 (s, 2H, N₉-H, N₁₇-H), 8.40 (s, 1H, C₂-H), 8.32 (s, 1H, C₅-H), 7.57 (dd, *J* = 8.6, 2.9 Hz, 2H, C₁₂-H, C₁₅-H), 7.51 (s, 1H, C₈-H), 7.30 (td, *J* = 8.5, 2.8 Hz, 1H, C₁₄-H), 6.63 (d, *J* = 6.8 Hz, 1H, C₁₉-H), 6.50 (q, *J* = 6.6 Hz, 1H, C₂₀-H), 3.37 (d, *J* = 6.2 Hz, 2H, C₂₁-H₂), 2.57 (s, 3H, C₁₆-H₃).

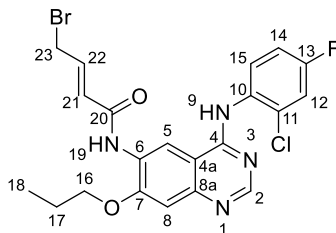
¹³C NMR (100 MHz, DMSO-*d*₆) δ 168.3 (C₁₈), 161.3 (d, *J* = 244.5 Hz, C₁₃), 158.7 (C₄), 155.0 (C₂), 148.4 (C_{8a}), 144.6 (C₇), 132.6 (C₁₀*), 132.2 (C₆), 131.2 (C₁₅, C₁₁*), 128.8 (C₂₀), 121.3 (C₅, C₈), 116.9 (d, *J* = 25.9 Hz, C₁₂), 114.8 (d, *J* = 22.5 Hz, C₁₄), 111.2 (C_{4a}), 110.6 (C₁₉), 36.7 (C₂₁), 13.9 (C₁₆).

IR (ATR), ν_{max} (cm⁻¹) 3439 (*st* NH), 3249 (*st* NH), 3028 (*st* C_{ar}-H), 2990 (*st* C-H), 2911, 1660 (*st* C=O), 1620, 1534 (*sc* N-H), 1490 (*st* C_{ar}-C_{ar}), 1421, 1392, 1300 (*st* C_{ar}-N), 1203 (*w* CH₂Br), 1091 (*st* C-N), 820 (*oop* C_{ar}-H), 663.

HRMS (Q-TOF): *m/z* calculated for C₁₉H₁₂BrClFN₄OS [M]⁺: 479.9823. Found: 480.9812 [M+H]⁺, 482.9789 [M+H+2]⁺

mp: 224-228 °C

(E)-4-bromo-N-(4-((2-chloro-4-fluorophenyl)amino)-7-propoxyquinazolin-6-yl)but-2-enamide (38{2})



38{2}

Starting from *N*⁴-(2-chloro-4-fluorophenyl)-6-nitro-7-propoxyquinazolin-4,6-diamine (**35{2}**) and freshly prepared (*E*)-4-bromobut-2-enoyl chloride (**36**) the product **38{1}** (**J**) was obtained as an other solid with 78% yield.

Spectroscopic data:

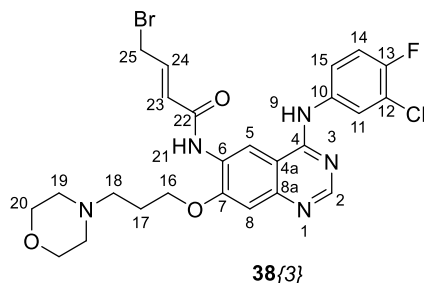
¹H NMR (400 MHz, DMSO-*d*₆) δ 9.73 (s, 1H, N9-H), 9.44 (s, 1H, N15-H), 8.76 (s, 1H, C2-H), 8.33 (s, 1H, C5-H), 7.63 – 7.49 (m, 2H, C12-H, C15-H), 7.29 (dd, *J* = 8.5, 3.1 Hz, 1H, C14-H), 7.23 (s, 1H C8-H), 6.65 (d, *J* = 6.9 Hz, 1H, C21-H), 6.53 (q, *J* = 6.7 Hz, 1H, C22-H), 4.17 (t, *J* = 6.5 Hz, 2H, C16-H₂), 3.44 (d, *J* = 6.2 Hz, 2H, C23-H₂), 1.86 (sextet, *J* = 6.7 Hz, 2H, C17-H₂), 1.05 (t, *J* = 7.4 Hz, 3H, C18-H₃).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 167.9 (C20), 159.9 (d, *J* = 244.8 Hz), 158.4 (C4), 154.9 (C2), 153.6 (C8a), 142.7 (C7), 132.8 (C10*), 132.1 (C6), 130.9 (C15), 128.8 (C22), 127.3 (C11*), 116.9 (d, *J* = 25.9 Hz, C12), 116.1 (C5), 114.7 (d, *J* = 21.7 Hz, C14), 110.9 (C21), 106.1 (C8), 70.3 (C16), 37.4 (C23), 21.7 (C17), 10.4 (C18).

IR (ATR), ν_{max} (cm⁻¹) 3426 (*st* NH), 3385 (*st* NH), 2927 (*st* C-H), 2884, 1695 (*st* C=O), 1622, 1524 (*sc* N-H), 1487 (*st* C_{ar}-C_{ar}), 1420, 1195 (*st* C-O), 964, 853 (*w* N-H).

HRMS (Q-TOF): *m/z* calculated for C₂₁H₁₉BrClFN₄O₂ [M]⁺: 492.0364. Found: 493.0380 [M+H]⁺, 495.0358 [M+H+2]⁺

mp: 158-162 °C

(E)-4-bromo-N-(4-((3-chloro-4-fluorophenyl)amino)-7-(3-morpholinopropoxy)quinazolin-6-yl)but-2-enamide (38{3})

Starting from *N*⁴-(3-chloro-4-fluorophenyl)-7-(3-morpholinopropoxy)-6-nitroquinazolin-4,6-diamine (**35{7}**) and freshly prepared (*E*)-4-bromobut-2-enoyl chloride (**36**) the product **38{3}** (**LMF408**) was obtained as a yellow crystalline solid with 29% yield.

Spectroscopic data:

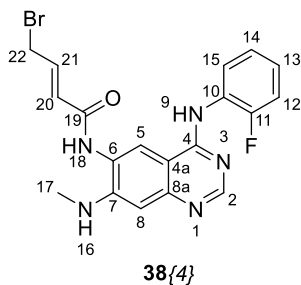
¹H NMR (400 MHz, DMSO-*d*₆) δ 9.78 (s, 1H, N9-H), 9.50 (s, 1H, N21-H), 8.75 (s, 1H, C5-H), 8.53 (s, 1H, C2-H), 8.13 (dd, *J* = 6.9, 2.6 Hz, 1H, C11-H), 7.80 (ddd, *J* = 9.0, 4.3, 2.5, 1H, C15-H), 7.41 (t, *J* = 9.1 Hz, 1H, C14-H), 7.27 (s, 1H, C8-H), 6.65 (d, *J* = 6.9 Hz, 1H, C23-H), 6.53 (q, *J* = 6.7 Hz, 1H, C24-H), 4.25 (t, *J* = 6.2 Hz, 2H, C16-H₂), 3.59 (s br., 4H, C20-H₂, C20'-H₂), 3.44 (d, *J* = 6.0 Hz, 2H, C25-H₂), 2.53 – 2.47 (m, 2H, C18-H₂), 2.40 (s br., 4H, C19-H₂, C19'-H₂), 2.00 (qt, *J* = 6.9 Hz, 2H, C17-H₂).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 167.9 (C22), 156.8 (C4), 155.0 (C7), 154.0 (C2), 153.2 (d, *J* = 242.5 Hz, C13), 149.2 (C8a), 136.8 (C10), 128.8 (C24), 127.1 (C6), 123.5 (C11), 122.4 (d, *J* = 7.2 Hz, C15), 118.7 (d, *J* = 18.3 Hz, C12), 116.5 (C5), 116.3 (d, *J* = 22.1 Hz, C14), 110.8 (C23), 108.7 (C4a), 107.3 (C8), 66.9 (C16), 66.1 (C20), 54.6 (C18), 53.3 (C19), 37.3 (C25), 25.4 (C17).

IR (ATR), ν_{max} (cm⁻¹) 3395 (*st* NH), 3287 (*st* NH), 2957 (*st* C-H), 2821, 1684 (*st* C=O), 1624, 1578, 1530 (*sc* N-H), 1496, 1451, 1423 (*st* C_{ar}-C_{ar}), 1386, 1332 (*st* C_{ar}-N), 1208 (*w* CH₂Br), 1113 (*st* C-O), 864.

HRMS (Q-TOF): *m/z* calculated for C₂₅H₂₆BrClFN₅O₃ [M]⁺: 577.0892. Found: 578.0914 [M+H]⁺, 580.0892ç [M+H+2]⁺

mp: 208-210 °C

(E)-4-bromo-N-(4-((2-fluorophenyl)amino)-7-(methylamino)quinazolin-6-yl)but-2-enamide (38{4})

Starting from *N*⁴-(2-fluorophenyl)-*N*⁷-methyl-6-nitroquinazolin-4,6,7-triamine (**35{4}**) and freshly prepared (*E*)-4-bromobut-2-enoyl chloride (**36**) the product **38{4}** (**LMF412**) was obtained as an ochre-greenish solid with 53% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 9.47 (s, 2H, N₉-H, N₁₈-H), 8.27 (s, 1H, C₂-H), 8.16 (s, 1H, C₅-H), 7.49 (t, *J* = 7.8 Hz, 1H, C₁₄-H), 7.27 (dd, *J* = 8.7, 3.3 Hz, 2H, C₁₃-H, C₁₂-H), 7.22 (dd, *J* = 6.9, 3.8 Hz, 1H, C₁₅-H), 6.63 (s, 1H, C₈-H), 6.61 (d, *J* = 6.8 Hz, 1H, C₂₂-H), 6.55 (q, *J* = 6.5 Hz, 1H, C₂₁-H), 6.20 (d, *J* = 4.4 Hz, 1H, N₁₆-H), 3.39 (d, *J* = 6.1 Hz, 2H, C₂₂-H₂), 2.83 (d, *J* = 4.8 Hz, 3H, C₁₇-H₃).

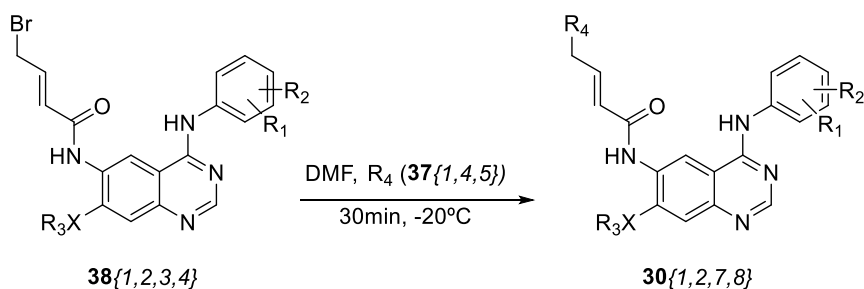
¹³C NMR (100 MHz, DMSO-*d*₆) δ 168.7 (C₁₉), 157.6 (C₄), 156.8 (d, *J* = 247.2 Hz, C₁₁), 154.4 (C₂), 150.2 (C_{8a}), 149.2 (C₇), 129.2 (C₂₁), 128.2 (C₁₄), 126.8 (d, *J* = 18.9 Hz, C₁₀), 126.7 (d, *J* = 6.8 Hz, C₁₃), 126.0 (C₆), 124.2 (d, *J* = 3.4 Hz, C₁₅), 121.1 (C₅), 115.8 (d, *J* = 20.1 Hz, C₁₂), 110.2 (C₂₀), 104.9 (C_{4a}), 102.1 (C₈), 36.9 (C₂₂), 29.5 (C₁₇).

IR (ATR), ν_{max} (cm⁻¹) 3203 (*st* NH), 2928 (*st* C-H), 1681 (*st* C=O), 1621, 1579 (*st* C_{ar}-C_{ar}), 1521 (*sc* N-H), 1500 (*st* C_{ar}-C_{ar}), 1453, 1417, 1387 (*rock* C-H), 1257 (*w* CH₂Br), 843 (*w* N-H), 749 (*oop* C_{ar}-H).

HRMS (Q-TOF): *m/z* calculated for C₁₉H₁₇BrFN₅O [M]⁺: 429.0601. Found: 430.0588 [M+H]⁺

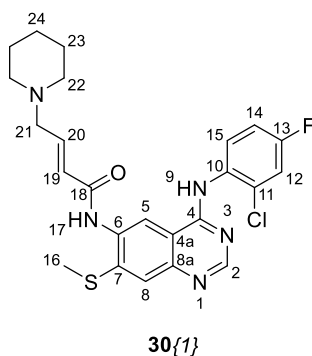
mp: 245-250 °C

5.2.3.11. Synthesis of (*E*)-*N*-(4-(phenyl)amino)-7-substituted-quinazolin-6-yl)-4-(Het)but-2-enamide (method B)



0.45 mmol (1.0 eq) of the corresponding compound **38** were dissolved in 7 ml of anhydrous DMF and cooled to -20°C (ice-salt-water bath). 3.05 mmol (6.8 eq.) of the corresponding **37** heterocycle were added dropwise and the reaction was set stirring for 2 hours at -20°C . The complete reaction was checked by TLC and this was quenched with 25 ml of water. Later, the aqueous mixture was poured into 50 ml saturated NaHCO_3 solution and extracted twice with 20 ml of ethyl acetate. The combined organic fractions were washed with 20 ml brine, dried over Na_2SO_4 , filtered, and the solvent was removed under reduced pressure to yield the desired and pure product **30**.

(E)-*N*-(4-((2-chloro-4-fluorophenyl)amino)-7-(methylthio)quinazolin-6-yl)-4-(piperidin-1-yl)but-2-enamide (**30{1}**)



Starting from (*E*)-4-bromo-*N*-(4-((2-chloro-4-fluorophenyl)amino)-7-(methylthio)quinazolin-6-yl)but-2-enamide (**38{1}**) and piperidine (**37{1}**) the product **30{1}** (**LMF415**) was obtained as a pale-yellow solid with 24% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 9.81 (s, 2H, N9-H, N17-H), 8.39 (s, 1H, C2-H), 8.37 (s, 1H, C5-H), 7.56 (dd, *J* = 12.1, 7.6 Hz, 2H, C12-H, C15-H), 7.52 (s, 1H, C8-H), 7.29 (t, *J* = 8.0 Hz, 1H, C14-H), 6.83 – 6.72 (m, 1H, C20-H), 6.37 (d, *J* = 14.5 Hz, 1H, C19-H), 3.10 (d, *J* = 5.2 Hz, 2H, C21-H₂), 2.57 (s, 3H, C16-H₃), 2.36 (s br., 4H, C22-H₂, C22'-H₂), 1.52 (s br., 4H, C23-H₂, C23'-H₂), 1.39 (s br., 2H, C24-H₂).

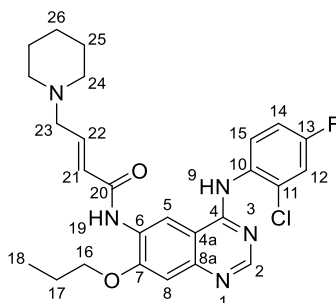
¹³C NMR (100 MHz, DMSO-*d*₆) δ 164.2 (C18), 161.3 (d, *J* = 246.1 Hz, C13), 158.6 (C4), 155.0 (C2), 148.5 (C8a), 144.3 (C7), 141.9 (C20), 132.6 (C10), 132.3 (C6), 131.3 (d, *J* = 10.2 Hz, C15), 125.1 (C21), 121.5 (C8), 121.1 (C21), 116.9 (d, *J* = 25.9 Hz, C12), 114.8 (d, *J* = 22.9 Hz, C14), 111.2 (C4a), 59.3 (C21), 54.1 (C22, C22'), 25.6 (C23, C23'), 23.9 (C24), 13.9 (C16).

IR (ATR), ν_{max} (cm⁻¹) 3428 (*st* NH), 3228 (*st* NH), 3025 (*st* C_{ar}-H), 2944 (*st* C-H), 2911, 1679, 1654 (*st* C=O), 1619, 1559, 1532 (*sc* N-H), 1472, 1413 (*st* C_{ar}-C_{ar}), 1199 (*st* C-N), 953, (*trans wag* =C-H), 814 (*w* N-H), 780 (*oop* C_{ar}-H).

HRMS (Q-TOF): *m/z* calculated for C₂₄H₂₅ClFN₅O₅ [M]⁺: 485.1452. Found: 486.1480 [M+H]⁺, 488.1465 [M+H+2]⁺

mp: 218-223 °C

(*E*)-*N*-(4-((2-chloro-4-fluorophenyl)amino)-7-propoxyquinazolin-6-yl)-4-(piperidin-1-yl)but-2-enamide (30{2})



30{2}

Starting from (*E*)-4-bromo-*N*-(4-((2-chloro-4-fluorophenyl)amino)-7-propoxyquinazolin-6-yl)but-2-enamide (**38{2}**) and piperidine (**37{1}**) the product **30{2}** (**LMF413**) was obtained as a pale-yellow solid with 71% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 9.70 (s, 1H, N9-H), 9.45 (s, 1H, N19-H), 8.85 (s, 1H, C5-H), 8.33 (s, 1H, C2-H), 7.65 – 7.47 (m, 2H, C15-H, C12-H), 7.32 – 7.25 (m, 1H, C14-H), 7.24 (s, 1H, C8-H), 6.79 (dt, *J* =

15.4, 6.1 Hz, 1H, C22-H), 6.53 (d, $J = 15.1$ Hz, 1H, C21-H), 4.17 (t, $J = 6.6$ Hz, 2H, C16-H₂), 3.10 (d, $J = 6.0$ Hz, 2H, C23-H₂), 2.35 (s br., 4H, C24-H₂, C24'-H₂), 1.86 (sextet, $J = 7.3$ Hz, 2H, C17-H₂), 1.51 (qt, $J = 5.3$ Hz, 4H, C25-H₂, C25'-H₂), 1.39 (d, $J = 2.8$ Hz, 2H, C26-H₂), 1.03 (t, $J = 7.4$ Hz, 3H, C18-H₃).

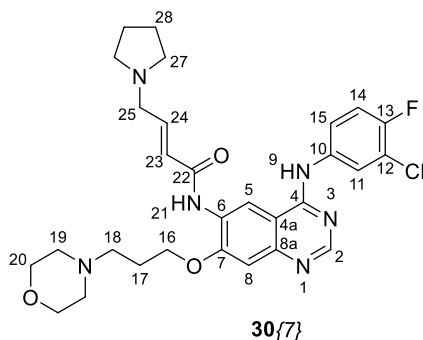
¹³C NMR (100 MHz, DMSO-*d*₆) δ 163.6 (C20), 162.3 (C4), 159.8 (d, $J = 246.0$ Hz, C13), 158.2 (C8a), 154.3 (C2), 148.8 (C7), 141.8 (C22), 133.1 (d, $J = 3.4$ Hz, C10), 132.1 (d, $J = 11.1$ Hz, C11), 130.9 (d, $J = 9.2$ Hz, C15), 127.2 (C6), 125.9 (C21), 116.8 (d, $J = 25.9$ Hz, C12), 116.3 (C5), 114.6 (d, $J = 22.1$ Hz, C14), 108.4 (C4a), 107.1 (C8), 70.1 (C16), 59.3 (C23), 54.1 (C24, C24'), 25.5 (C25, C25'), 23.9 (C26), 21.7 (C16), 10.4 (C17).

IR (ATR), ν_{max} (cm⁻¹) 3422 (st NH), 2927 (st C-H), 1692 (st C=O), 1630, 1576, 1527 (sc N-H), 1449, 1423 (st C_{ar}-C_{ar}), 1200 (st C-N), 965, (trans wag =C-H), 862, 738 (oop C_{ar}-H).

HRMS (Q-TOF): m/z calculated for C₂₆H₂₉ClFN₅O₂ [M]⁺: 497.1994. Found: 498.2013 [M+H]⁺, 500.2002 [M+H+2]⁺

mp: 145-155 °C

(*E*)-*N*-(4-((3-chloro-4-fluorophenyl)amino)-7-(3-morpholinopropoxy)quinazolin-6-yl)-4-(pyrrolidin-1-yl)but-2-enamide (30{7})



Starting from (*E*)-4-bromo-*N*-(4-((2-chloro-4-fluorophenyl)amino)-7-propoxyquinazolin-6-yl)but-2-enamide (**38{3}**) and pyrrolidine (**37{4}**) the product **30{7}** (LMF409) was obtained as a pale-brown solid with 97% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 9.79 (s, 1H, N9-H), 9.50 (s, 1H, N21-H), 8.86 (s, 1H, C5-H), 8.53 (s, 1H, C2-H), 8.14 (dd, $J = 6.9, 2.4$ Hz, 1H, C11-H), 7.81 (ddd, $J = 8.9, 4.3, 2.5$ Hz, 1H, C15-H), 7.42 (t, $J = 9.1$ Hz, 1H, C14-H), 7.27 (s, 1H, C8-H), 6.84 (dt, $J = 15.3, 5.8$ Hz, 1H, C24-H), 6.55 (d, $J = 15.4$ Hz, 1H, C23-H), 4.26 (t, $J = 6.2$ Hz, 2H, C16-H₂), 3.58 (t, $J = 4.6$ Hz, 4H, C20-H₂, C20'-H₂), 3.26 (d, $J = 5.7$ Hz, 2H, C25-

H₂), 2.51 – 2.41 (m, 6H, C27-H₂, C27'-H₂, C18-H₂), 2.41 – 2.34 (m, 4H, C19-H₂, C19'-H₂), 2.07 – 1.94 (m, 2H, C17-H₂), 1.72 (s, 4H, C28-H₂, C28'-H₂).

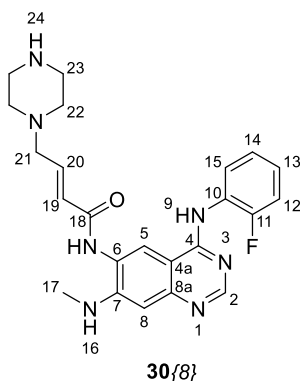
¹³C NMR (100 MHz, DMSO-*d*₆) δ 163.8 (C22), 156.7 (C4), 154.9 (C7), 153.9 (C2), 153.1 (d, *J* = 242.7 Hz, C13), 149.1 (C8a), 142.4 (C24), 136.8 (C10), 127.3 (C6), 125.0 (C23), 123.5 (C11), 122.3 (d, *J* = 6.8 Hz, C15), 118.7 (d, *J* = 18.2 Hz, C12), 116.4 (d, *J* = 21.6 Hz, C14), 116.4 (C5), 108.7 (C4a), 107.3 (C8), 67.0 (C16), 66.2 (C20), 56.2 (C25), 54.7 (C18), 53.6 (C27), 53.4 (C19), 25.4 (C17), 23.3 (C28).

IR (ATR), ν_{max} (cm⁻¹) 3547 (st NH), 3317 (st NH), 3114 (st C_{ar}-H), 2957 (st C-H), 2800, 1670 (st C=O), 1628, 1531 (sc N-H), 1498, 1451, 1427 (st C_{ar}-C_{ar}), 1397, 1344, 1208 (st C-N), 1117 (st C-O), 850 (*oop* C_{ar}-H).

HRMS (Q-TOF): *m/z* calculated for C₂₉H₃₄ClFN₆O₃ [M]⁺: 568.2365. Found: 569.2389 [M+H]⁺, 571.2380 [M+H+2]⁺

mp: 112-118 °C

(*E*)-*N*-(4-((2-fluorophenyl)amino)-7-(methylamino)quinazolin-6-yl)-4-(piperazin-1-yl)but-2-enamide (30{8})



Starting from (*E*)-4-bromo-*N*-(4-((2-chloro-4-fluorophenyl)amino)-7-propoxyquinazolin-6-yl)but-2-enamide (**38{4}**) and piperazine (**37{5}**) the product **30{8}** (**LMF414**) was obtained as an orange solid with 26% yield.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 9.46 (s, 1H, N₉-H), 9.36 (s, 1H, N₁*-H), 8.25 (s, 2H, C₂-H, C₅-H), 7.50 (t, *J* = 7.1 Hz, 1H, C₁₄-H), 7.31 – 7.16 (m, 3H, C₁₅-H, C₁₃-H, C₁₂-H), 6.73 (dt, *J* = 15.2, 6.0 Hz, 1H, C₂₀-H), 6.64 (s, 1H, C₈-H), 6.36 (d, *J* = 15.9 Hz, 1H, C₁₉-H), 6.12 (d, *J* = 5.3 Hz, 1H, C₁₇-H), 3.11 (d,

$J = 5.2$ Hz, 2H, C21-H₂), 2.83 (d, $J = 4.8$ Hz, 3H, C17-H₃), 2.76 – 2.73 (m, 4H, C23-H₂, C23'-H₂), 2.36 (s br., 4H, C22-H₂, C22'-H₂).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 164.3 (C18), 157.6 (C4), 156.8 (d, $J = 246.8$ Hz, C11), 154.5 (C2), 150.4 (C8a), 148.7 (C7), 140.5 (C20), 128.2 (d, $J = 19.9$ Hz, C10), 128.2 (C14), 126.6 (d, $J = 9.4$ Hz, C13), 126.2 (C19), 124.3 (C6), 124.2 (d, $J = 3.2$ Hz, C15), 120.3 (C5), 115.8 (d, $J = 20.0$ Hz, C12), 105.1 (C4a), 102.6 (C8), 59.2 (C21), 53.9 (C22, C22'), 45.4 (C23, C23'), 29.6 (C17).

IR (ATR), ν_{max} (cm⁻¹) 3428 (st NH), 3239 (st NH), 2922 (st C-H), 2854, 1673 (st C=O), 1646, 1619, 1559, 1530 (sc N-H), 1486, 1412 (st C_{ar}-C_{ar}), 1253 (st C_{ar}-N), 1183 (st C-O), 783 (oop C_{ar}-H).

HRMS (Q-TOF): m/z calculated for C₂₃H₂₆FN₇O [M]⁺: 435.2183. Found: 436.2210 [M+H]⁺

mp: 110-112 °C

5.3. Determination of the biological activity

5.3.1. *P. falciparum* *in vitro* growth inhibition assay

Intraerythrocytic strain *P. falciparum* 3D7A growth inhibition IC₅₀ was determined using a modification of the *in vitro* [³H]-hypoxanthine incorporation method (Desjardins *et al.*¹⁷⁷). Briefly, asynchronous cultures (ca. 70% rings) at 2% hematocrit and 0.5% parasitemia with 5 μ M hypoxanthine were exposed to 3-fold serial dilutions of the compounds in 96 well plates (Costar #3894). They were incubated 24 hours at 37 °C, 5% CO₂, 5% O₂, 95% N₂. After 24 hours of incubation, [³H]-hypoxanthine was added, and plates were incubated for an additional 24 hours. After that period, plates were harvested on glass fiber filters (Wallac #1450-421) using a cell harvester 96 (TOMTEC, Perkin Elmer). Filters were dried and melt-on scintillator sheets (MeltiLex A, PerkinElmer #1450-441) used to determine the incorporation of [³H]-hypoxanthine. Radioactivity was measured using a microbeta counter (Perkin Elmer). IC₅₀ values were determined using Excel and GrafIt 5 software.

5.3.2. EGFR enzymatic *in vitro* assay : measure of the residual activity

Determination of the effect of the compounds on the kinase activity of the 9 protein kinases *in vitro* was performed with a radiometric protein kinase assay (³³PanQinase® Activity Assay) at Reaction Biology Corp. (www.reactionbiology.com). Protein kinases used at Reaction Biology Corp. were purchased from Life Technologies (Invitrogen Corporation).

The kinase inhibition profile of 14 compounds was determined by measuring residual activity values at two concentrations each in singlicate in 9 protein kinase assays. Compounds were previously prepared as 1 x 10⁻⁰⁶ M/100% DMSO stock solutions. The compounds were tested at the following

final assay concentrations, in singlicate: 2.0E-09 M and 1.0E-08 M. The final DMSO concentration in all reaction cocktails (including high and low controls) was 1 %.

Protein Kinase Assay

A radiometric protein kinase assay (³³PanQinase® Activity Assay) was used for measuring activity of the 9 protein kinases. All kinase assays were performed in 96-well ScintiPlates™ from Perkin Elmer (Boston, MA, USA) in a 50 µl reaction volume. The reaction cocktail was pipetted in 4 steps in the following order: (a) 10 µl of non-radioactive ATP solution (in H₂O), (b) 25 µl of assay buffer/ [γ-³³P]-ATP mixture, (c) 5 µl of test sample in 10% DMSO, (d), 10 µl of enzyme/substrate mixture.

The assay for all protein kinases contained 70 mM HEPES-NaOH pH 7.5, 3 mM MgCl₂, 3 mM MnCl₂, 3 µM Na-orthovanadate, 1.2 mM DTT, 50 µg/ml PEG₂₀₀₀₀, ATP (variable concentrations, corresponding to the apparent ATP-K_m of the respective kinase, see Table 5.1), [γ-³³P]-ATP (approx. 8 x 10⁵ cpm per well), protein kinase (variable amounts; depending on the stock), and substrate (variable amounts; see Table 5.1 1).

All protein kinases provided by RBE were expressed in Sf9 insect cells or in *E.coli* as recombinant GST-fusion proteins or His-tagged proteins, either as full-length or enzymatically active fragments. All kinases were produced from human cDNAs. Kinases were purified by affinity chromatography using either GSH-agarose or immobilized metal. Affinity tags were removed from a number of kinases during purification. The purity of the protein kinases was examined by SDS-PAGE/Coomassie staining. The identity of the protein kinases was checked by mass spectroscopy.

Table 5.1 Assay parameters for the tested protein kinases

#	Kinase Name	Kinase PQ Lot	Kinase Concentration (ng/50µl)	Kinase Concentration (nM) *	ATP Concentration (µM)	Substrate Name	Substrate Lot	Substrate Concentration (µg/50µl)
1	EGFR	018,19	10	2,2	0,3	Poly(Glu,Tyr)4:1	SIG_20K5903	0,125
2	EGFR C797S	001,1	50	11,2	1	Poly(Glu,Tyr)4:1	SIG_20K5903	0,125
3	EGFR G719S	001,5	10	2,2	0,3	Poly(Glu,Tyr)4:1	SIG_20K5903	0,125
4	EGFR L858R	002,6	20	4,5	1	Poly(Glu,Tyr)4:1	SIG_20K5903	0,25
5	EGFR T790M	001,5	8	1,8	0,3	Poly(Glu,Tyr)4:1	SIG_20K5903	0,125
6	EGFR T790M/C797S/L858R	001,1	10	2,2	0,3	Poly(Glu,Tyr)4:1	SIG_20K5903	0,125
7	EGFR T790M/L858R	001,6	10	2,2	0,3	Poly(Glu,Tyr)4:1	SIG_20K5903	0,125
8	ERBB2	012,10	50	10,6	1	Poly(Glu,Tyr)4:1	SIG_20K5903	0,125
9	ERBB4	007,8	5	1,0	0,3	Poly(Glu,Tyr)4:1	SIG_20K5903	0,125

(*) Maximal molar enzyme assay concentrations, implying enzyme preparations exclusively containing 100 % active enzyme

The reaction cocktails were incubated at 30° C for 60 minutes. The reaction was stopped with 50 µl of 2 % (v/v) H₃PO₄, plates were aspirated and washed two times with 200 µl 0.9 % (w/v) NaCl. Incorporation of ³³P_i (counting of “cpm”) was determined with a microplate scintillation counter (Microbeta, Wallac).

Protein Kinase Assay

For each kinase, the median value of the cpm of three wells with complete reaction cocktails, but without kinase, was defined as "low control" (n=3). This value reflects unspecific binding of radioactivity to the plate in the absence of protein kinase but in the presence of the substrate. Additionally, for each kinase the median value of the cpm of three other wells with the complete reaction cocktail, but without any compound, was taken as the "high control", i.e. full activity in the absence of any inhibitor (n=3). The difference between high and low control was taken as 100 % activity for each kinase.

As part of the data evaluation the low control value of each kinase was subtracted from the high control value as well as from their corresponding "compound values". The residual activity (in %) for each compound well was calculated by using the following formula:

$$\text{Res. Activity (\%)} = 100 \times \left[\frac{(\text{cpm of compound} - \text{low control})}{(\text{high control} - \text{low control})} \right]$$

Quality controls

As a parameter for assay quality, the Z'-factor²⁵³ for the low and high controls of each assay plate (n = 8) was used. RBE's criterion for repetition of an assay is a Z'-factor below 0.4²⁵⁴.

As an additional quality control, a control inhibitor was tested in parallel, staurosporine (RBE). The inhibitor activity values were in the expected range for each kinase.

Chapter 6: Conclusions

All the reported results have proved that there is still space to be explored in the developed combinatorial libraries derived from the claimed Markush structures in the drug patents. Two proofs of concept have been developed to put in practice the suggested alternative strategy based in rational analysis with the study of the family of analogs of Tafenoquine and Dacomitinib. At the light of the results, we sustain that this methodology would highly benefit in terms of efficiency the early Drug Discovery process along with the database description of libraries of analogs for further reprofiling purposes. Consequently, considering the initial objectives stated in this thesis, the formulated conclusions are the following:

- 1.1. The initial study extracted from the latest FDA approved small molecules has allowed us to corroborate that the hit-to-lead procedure in the actual R&D process is highly localized compromising its efficiency.
- 1.2. Seven uncorrelated drug patents (e.g. Dacomitinib, Abemaciclib, Tafenoquine) have been analyzed to assess how well the chemical space claimed within is actually explored. It has been confirmed through the calculation of the respective space and population coverages that the reported molecules in literature (BD) do poorly represent the huge diversity involved in patent claims. Clear evidence is given with the space coverage, which is about 20% on average when clustering the chemical space in \sqrt{N} clusters (being N clusters the total number of analogs of the combinatorial library).
- 1.3. Neither the real explored space (BD) nor the fragment combination of the studied molecules (BCL) significantly represent the space defined by the combinatorial libraries derived from the Markush structure, even when compared with the coverage obtained by a statistically sampling of \sqrt{N} molecules via cherry-picking.
- 2.1. The Tafenoquine's chemical space study consisting in a family of 25,472 analogs have resulted in the synthesis of seven rationally selected analogs which have proven to be more representative (with a 70.0% SC and 74.8% PC) than the 58 reported analogs until date (with a 50.0% SC and 42.5% PC). The results obtained by the biological activity tests against *P. falciparum* asexual blood stages (which were conducted by GSK) showed that all candidates present interesting activities, being compound **15{2}** a potential lead candidate (which showed a IC₅₀ of 4.1 μ M). The study of the explored chemical space evidenced that many of the molecules reported in bibliography are placed in two regions with low IC₅₀ values indicating a biased selection and a poor/limited exploration of the chemical space available

as, after this study, both clusters would not be recommended to be further explored in optimization studies.

- 3.1. Dacomitinib's chemical space study has evidenced the vagueness of some of the Markush structure claimed in the patents, as the total enumeration of its derivatives comprehend more than 10 billion analogs. Through the library selection of the synthetic accessible space, it has been able to represent with only 8 compounds the same fraction represented by the actual 60 reported molecules (with 8.0% SC and 92.6% PC). Their *in vitro* biological assessment with a radiometric protein kinase assay against EGFR (wild-type and 6 of its mutations), HER2 and HER4, have demonstrated that all the molecules show some inhibitory activity at 10nM for all cases except for HER2. Compound **30{3}** stands out as the most interesting candidate for further studies as it has shown analogous inhibitory trend compared to Dacomitinib. As both are part of the same cluster (3), it would be suggested to deeply explore this region through further re-clustering steps in order to probably find a new lead candidate.
- 4.1. Finally, in both proofs of concept, molecule intermediates and alternative analogs have also shown interesting biological activities leaving the door opened for further hit-optimization processes outside of the Markush claimed space.

Therefore, through this study, our principal hypothesis is confirmed as the rational selection of a small sample of analogs has proven to be an efficient exploratory methodology suitable for the early drug discovery stages such as the hit-to-lead process or for High-Throughput Screening purposes.

Chapter 7: Bibliography

- (1) Sussman, A. *Million-Card Monte: Reforming the Markush Claim Post-AIA to Save Synthetic Chemical Innovation*; 12 J. MARSHALL REV. INTELL. PROP. L., 2013.
- (2) Markush, E. A. Pyrazolone Dye & Process of Making the Same. U.S. Patent No. 1,506,316, 1924.
- (3) Duelen, R.; Corvelyn, M.; Tortorella, I.; Leonardi, L.; Chai, Y. C.; Sampaolesi, M. *Medicinal Biotechnology for Disease Modeling, Clinical Therapy, and Drug Discovery and Development*; 2019. https://doi.org/10.1007/978-3-030-22141-6_5.
- (4) McKinsey. First-time launchers in the pharmaceutical industry <https://www.mckinsey.com/industries/life-sciences/our-insights/first-time-launchers-in-the-pharmaceutical-industry> (accessed Sep 5, 2022).
- (5) DiMasi, J. A.; Hansen, R. W.; Grabowski, H. G. The Price of Innovation: New Estimates of Drug Development Costs. *J. Health Econ.* **2003**, *22* (2), 151–185. [https://doi.org/10.1016/S0167-6296\(02\)00126-1](https://doi.org/10.1016/S0167-6296(02)00126-1).
- (6) Austin, D. Development in the Pharmaceutical Industry At a Glance. **2021**.
- (7) Stevens, E. *Medicinal Chemistry: The Modern Drug Discovery Process*; 2013.
- (8) Macalino, S. J. Y.; Gosu, V.; Hong, S.; Choi, S. Role of Computer-Aided Drug Design in Modern Drug Discovery. *Arch. Pharm. Res.* **2015**, *38* (9), 1686–1701. <https://doi.org/10.1007/s12272-015-0640-5>.
- (9) Maldonado, A. G.; Doucet, J. P.; Petitjean, M.; Fan, B. T. Molecular Similarity and Diversity in Chemoinformatics: From Theory to Applications. *Mol. Divers.* **2006**, *10* (1), 39–79. <https://doi.org/10.1007/s11030-006-8697-1>.
- (10) Beale, J. M.; Block, J. H. *Wilson and Gisvold's Textbook of Organic Medicinal and Pharmaceutical Chemistry*, 12th ed.; Wolters Kluwer Health, 2010.
- (11) Lombardino, J. G.; Lowe, J. A. The Role of the Medicinal Chemist in Drug Discovery - Then and Now. *Nat. Rev. Drug Discov.* **2004**, *3* (10), 853–862. <https://doi.org/10.1038/nrd1523>.
- (12) Grogan S, P. C. Pharmacokinetics <https://www.ncbi.nlm.nih.gov/books/NBK557744/> (accessed Aug 25, 2022).
- (13) Haschek, W. M.; Rousseaux, C. G.; Wallig, M. A.; Bolon, B.; Ochoa, R. *Haschek and Rousseaux's Handbook of Toxicologic Pathology*; Elsevier, 2013. <https://doi.org/10.1016/C2010-1-67850-9>.
- (14) Li, A. P. Screening for Human ADME/Tox Drug Properties in Drug Discovery. *Drug Discov. Today* **2001**, *6* (7), 357–366. [https://doi.org/10.1016/S1359-6446\(01\)01712-3](https://doi.org/10.1016/S1359-6446(01)01712-3).
- (15) Enzo Sciences. Innovative Tools for Drug Discovery www.enzolifesciences.com/browse/drug-discovery/ (accessed Sep 25, 2022).
- (16) Patrick, G. L. *An Introduction to Medicinal Chemistry*; Oxford University Press, 1995.
- (17) Hansch, C.; Muir, R. M.; Fujita, T.; Maloney, P. P.; Geiger, F.; Streich, M. Correlation of

- Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients. *Nature* **1962**, *194*, 170–180.
- (18) Hansch, C.; Muir, R. M.; Fujita, T.; Maloney, P. P.; Geiger, F.; Streich, M. The Correlation of Biological Activity of Plant Growth Regulators and Chloromycetin Derivatives with Hammett Constants and Partition Coefficients. *J. Am. Chem. Soc.* **1963**, *719*, 2817–2824.
- (19) Free, S. M.; Wilson, J. W. A Mathematical Contribution to Structure-Activity Studies. *J. Med. Chem.* **1964**, *7* (4), 395–399. <https://doi.org/10.1021/jm00334a001>.
- (20) Kopecký, J.; Boček, K.; Vlachová, D. Chemical Structure and Biological Activity on M-and p-Disubstituted Derivatives of Benzene. *Nature* **1965**, *207* (5000), 981.
- (21) Kopecký, J.; Boček, K. A Correlation between Constants Used in Structure-Activity Relationships. *Specialia* **1967**, 125.
- (22) Fujita, T.; Iwasa, J.; Hansch, C. A New Substituent Constant, σ_r , Derived from Partition Coefficients. *J. Am. Chem. Soc.* **1964**, *86* (23), 5175–5180. <https://doi.org/10.1021/ja01077a028>.
- (23) Kubinyi, H. Free Wilson Analysis. Theory, Applications and Its Relationship to Hansch Analysis. *Quant. Struct. Relationships* **1988**, *7* (3), 121–133. <https://doi.org/10.1002/qsar.19880070303>.
- (24) Craig, P. N. Comparison of Hansch and Free-Wilson Methods for Structure-Activity Correlation. *Cancer Chemother. Reports* **1974**, *4* (4), 39.
- (25) Todeschini, R.; Consonni, V.; Mannhold, R.; Kubinyi, H.; Folkers, G. *Molecular Descriptors for Chemoinformatics*; 2009.
- (26) Cruciani, G.; Pastor, M.; Mannhold, R. Suitability of Molecular Descriptors for Database Mining. A Comparative Analysis. *J. Med. Chem.* **2002**, *45* (13), 2685–2694. <https://doi.org/10.1021/jm0011326>.
- (27) Nilsson, I.; Polla, M. O. Composite Multi-Parameter Ranking of Real and Virtual Compounds for Design of MC4R Agonists: Renaissance of the Free-Wilson Methodology. *J. Comput. Aided. Mol. Des.* **2012**, *26* (10), 1143–1157. <https://doi.org/10.1007/s10822-012-9605-7>.
- (28) Goldberg, F. W.; Leach, A. G.; Scott, J. S.; Snelson, W. L.; Groombridge, S. D.; Donald, C. S.; Bennett, S. N. L.; Bodin, C.; Gutierrez, P. M.; Gyte, A. C. Free-Wilson and Structural Approaches to Co-Optimizing Human and Rodent Isoform Potency for 11 β -Hydroxysteroid Dehydrogenase Type 1 (11 β -HSD1) Inhibitors. *J. Med. Chem.* **2012**, *55* (23), 10652–10661. <https://doi.org/10.1021/jm3013163>.
- (29) Freeman-Cook, K. D.; Amor, P.; Bader, S.; Buzon, L. M.; Coffey, S. B.; Corbett, J. W.; Dirico, K. J.; Doran, S. D.; Elliott, R. L.; Esler, W.; Guzman-Perez, A.; Henegar, K. E.; Houser, J. A.; Jones, C. S.; Limberakis, C.; Loomis, K.; McPherson, K.; Murdande, S.; Nelson, K. L.; Phillion, D.; Pierce, B. S.; Song, W.; Sugarman, E.; Tapley, S.; Tu, M.; Zhao, Z. Maximizing Lipophilic Efficiency: The Use of Free-Wilson Analysis in the Design of Inhibitors of Acetyl-CoA Carboxylase. *J. Med. Chem.* **2012**, *55* (2), 935–942. <https://doi.org/10.1021/jm201503u>.
- (30) Polishchuk, P. Interpretation of Quantitative Structure-Activity Relationship Models: Past, Present, and Future. *J. Chem. Inf. Model.* **2017**, *57* (11), 2618–2639. <https://doi.org/10.1021/acs.jcim.7b00274>.

- (31) Chen, H.; Carlsson, L.; Eriksson, M.; Varkonyi, P.; Norinder, U.; Nilsson, I. Beyond the Scope of Free-Wilson Analysis: Building Interpretable QSAR Models with Machine Learning Algorithms. *J. Chem. Inf. Model.* **2013**, *53* (6), 1324–1336. <https://doi.org/10.1021/ci4001376>.
- (32) Craig, P. N. Comparison of the Hansch and Free-Wilson Approaches to Structure-Activity Correlation. *Biol. Correl. Hansch Approach* **1974**, *114*, 115–129.
- (33) EFPIA. European Federation of Pharmaceuticals Industries and Associations <https://www.efpia.eu/> (accessed Sep 25, 2022).
- (34) Barnard, J. M.; Downs, G. M. MSI Combinatorial Chemistry Consortium Meeting. In *Use of Markush Structure Techniques to Avoid Enumeration in Diversity Analysis of Large Combinatorial Libraries*; 1997.
- (35) Maclean, D.; Martin, E. J. On the Representation of Combinatorial Libraries. *J. Comb. Chem.* **2004**, *6* (1), 1–11. <https://doi.org/10.1021/cc0340325>.
- (36) Barnard, J. M. A Comparison of Different Approaches to Markush Structure Handling. *J. Chem. Inf. Comput. Sci.* **1991**, *31* (1), 64–68. <https://doi.org/10.1021/ci00001a010>.
- (37) Barnard, J. M.; Downs, G. M.; Von Scholley-Pfab, A.; Brown, R. D. Use of Markush Structure Analysis Techniques for Descriptor Generation and Clustering of Large Combinatorial Libraries. *J. Mol. Graph. Model.* **2000**, *18* (4–5), 452–463. [https://doi.org/10.1016/S1093-3263\(00\)00067-X](https://doi.org/10.1016/S1093-3263(00)00067-X).
- (38) Reymond, J.; Deursen, R. Van; Blum, L. C.; Ruddigkeit, L. Chemical Space as a Source for New Drugs. *Medchemcomm* **2010**, *1* (December 2005), 30–38. <https://doi.org/10.1039/c0md00020e>.
- (39) Dobson, C. M. Chemical Space and Biology. **2004**, *432* (December), 824–828.
- (40) Sorzano, C. O. S.; Vargas, J.; Montano, A. P. A Survey of Dimensionality Reduction Techniques. **2014**, 1–35.
- (41) Pascual, R.; Mateu, M.; Gasteiger, J.; Borrell, J. I.; Teixidó, J. Design and Analysis of a Combinatorial Library of HEPT Analogues: Comparison of Selection Methodologies and Inspection of the Actually Covered Chemical Space. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (1), 199–207. <https://doi.org/10.1021/ci0255681>.
- (42) Leach, A. R.; Hann, M. M. The in Silico World of Virtual Libraries. *Drug Discov. Today* **2000**, *5* (8), 326–336.
- (43) Johnson, M. A.; Maggiora, G. M. *Concepts and Applications of Molecular Similarity*; New York, NY, 1990. <https://doi.org/10.1002/jcc.540130415>.
- (44) Maggiora, G.; Vogt, M.; Stumpfe, D.; Bajorath, J. Molecular Similarity in Medicinal Chemistry. *J. Med. Chem.* **2014**, *57* (8), 3186–3204. <https://doi.org/10.1021/jm401411z>.
- (45) Medina-Franco, J. L.; Maggiora, G. M. MOLECULAR SIMILARITY ANALYSIS. In *Chemoinformatics for Drug Discovery*; John Wiley & Sons, Inc: Hoboken, NJ, 2013; pp 343–399. <https://doi.org/10.1002/97811118742785.ch15>.
- (46) Bender, A.; Glen, R. C. Molecular Similarity: A Key Technique in Molecular Informatics. *Org. Biomol. Chem.* **2004**, *2* (22), 3204. <https://doi.org/10.1039/b409813g>.

- (47) Bajorath, J. *Molecular Similarity Concepts for Informatics Applications*; 2017; pp 231–245. https://doi.org/10.1007/978-1-4939-6613-4_13.
- (48) Xue, L.; Godden, J. W.; Bajorath, J. Evaluation of Descriptors and Mini-Fingerprints for the Identification of Molecules with Similar Activity. *J. Chem. Inf. Comput. Sci.* **2000**, *40* (5), 1227–1234. <https://doi.org/10.1021/ci000327j>.
- (49) Bajorath, J. Integration of Virtual and High-Throughput Screening. *Nat. Rev. Drug Discov.* **2002**, *1* (11), 882–894. <https://doi.org/10.1038/nrd941>.
- (50) Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of MDL Keys for Use in Drug Discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (6), 1273–1280. <https://doi.org/10.1021/ci010132r>.
- (51) Bajusz, D.; Rácz, A.; Héberger, K. Chemical Data Formats, Fingerprints, and Other Molecular Descriptions for Database Analysis and Searching. In *Comprehensive Medicinal Chemistry III*; Elsevier, 2017; pp 329–378. <https://doi.org/10.1016/B978-0-12-409547-2.12345-5>.
- (52) Brown, R. D. Descriptors for Diversity Analysis. *Perspect. Drug Discov. Des.* **1997**, *7–8*, 31–49. <https://doi.org/10.1007/bf03380180>.
- (53) Willett, P.; Barnard, J. M.; Downs, G. M. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38* (6), 983–996. <https://doi.org/10.1021/ci9800211>.
- (54) Rose, J.; Gottfries, J.; Muresan, S.; Backlund, A.; Oprea, T. I. Novel Chemical Space Exploration via Natural Products. **2009**, 1953–1962.
- (55) Martin, E.; Cao, E. Euclidean Chemical Spaces from Molecular Fingerprints: Hamming Distance and Hempel's Ravens. *J. Comput. Aided. Mol. Des.* **2015**, *29* (5), 387–395. <https://doi.org/10.1007/s10822-014-9819-y>.
- (56) Van Der Maaten, L.; Hinton, G. Visualizing Data Using T-SNE. *Laurens. J. Mach. Learn. Res.* **2008**, *9*, 25792605.
- (57) Pearson, K. LIII. On Lines and Planes of Closest Fit to Systems of Points in Space. *London, Edinburgh, Dublin Philos. Mag. J. Sci.* **1901**, *2* (11), 559–572. <https://doi.org/10.1080/14786440109462720>.
- (58) Wold, S.; Esbensen, K.; Geladi, P. Principal Component Analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2* (1–3), 37–52. [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9).
- (59) Dunteman, G. *Principal Components Analysis*; SAGE Publications, Inc.: 2455 Teller Road, Newbury Park California 91320 United States of America, 1989. <https://doi.org/10.4135/9781412985475>.
- (60) Jolliffe, I. T. *Principal Component Analysis*; Springer Series in Statistics; Springer New York: New York, NY, 1986. <https://doi.org/10.1007/978-1-4757-1904-8>.
- (61) Kohonen, T. Self-Organized Formation of Topologically Correct Feature Maps. *Biol. Cybern.* **1982**, *43* (1), 59–69. <https://doi.org/10.1007/BF00337288>.
- (62) Kireeva, N.; Baskin, I. I.; Gaspar, H. A.; Horvath, D.; Marcou, G.; Varnek, A. Generative Topographic Mapping (GTM): Universal Tool for Data Visualization, Structure-Activity Modeling and Dataset Comparison. *Mol. Inform.* **2012**, *31* (3–4), 301–312. <https://doi.org/10.1002/minf.201100163>.

- (63) Oprea, T. I.; Gottfries, J. Chemography: The Art of Navigating in Chemical Space. *J. Comb. Chem.* **2001**, *3* (2), 157–166. <https://doi.org/10.1021/cc0000388>.
- (64) Zabolotna, Y.; Lin, A.; Horvath, D.; Marcou, G.; Volochnyuk, D. M.; Varnek, A. Chemography: Searching for Hidden Treasures. *J. Chem. Inf. Model.* **2021**, *61* (1), 179–188. <https://doi.org/10.1021/acs.jcim.0c00936>.
- (65) Lipinski, C.; Hopkins, A. Navigating Chemical Space for Biology and Medicine. *Nature* **2004**, *432* (7019), 855–861. <https://doi.org/10.1038/nature03193>.
- (66) Bohacek, R. S.; McMartin, C.; Guida, W. C. The Art and Practice of Structure-Based Drug Design : A Molecular Modeling Perspective. *Med. Res. Rev.* **1996**, *16* (1), 3–50.
- (67) Ertl, P. Cheminformatics Analysis of Organic Substituents : Identification of the Most Common Substituents , Calculation of Substituent Properties , and Automatic Identification of Drug-like Bioisosteric Groups. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 374–380.
- (68) Brown, R. D.; Martin, Y. C. Use of Structure-Activity Data to Compare Structure-Based Clustering Methods and Descriptors for Use in Compound Selection. *J. Chem. Inf. Comput. Sci.* **1996**, *36* (3), 572–584. <https://doi.org/10.1021/ci9501047>.
- (69) Miller, M. A. Chemical Database Techniques in Drug Discovery. *Nat. Rev. Drug Discov.* **2002**, *1* (3), 220–227. <https://doi.org/10.1038/nrd745>.
- (70) Vogt, M. How Do We Optimize Chemical Space Navigation? *Expert Opin. Drug Discov.* **2020**, *15* (5), 523–525. <https://doi.org/10.1080/17460441.2020.1730324>.
- (71) Dunbar, J. B. Cluster-Based Selection. *Perspect. Drug Discov. Des.* **1997**, *7/8*, 51–63. <https://doi.org/10.1007/BF03380181>.
- (72) Snarey, M.; Terrett, N. K.; Willett, P.; Wilton, D. J. Comparison of Algorithms for Dissimilarity-Based Compound Selection. *J. Mol. Graph. Model.* **1997**, *15* (6), 372–385. [https://doi.org/10.1016/S1093-3263\(98\)00008-4](https://doi.org/10.1016/S1093-3263(98)00008-4).
- (73) Rabal, O.; Pascual, R.; Borrell, J. I.; Teixidó, J. Cell-Integral-Diversity Criterion: A Proposal for Minimizing Cluster Artifact in Cell-Based Selections. *J. Chem. Inf. Model.* **2007**, *47* (5), 1886–1896. <https://doi.org/10.1021/ci600433c>.
- (74) Mason, J. S.; Pickett, S. D. Partition-Based Selection. *Perspect. Drug Discov. Des.* **1997**, *7--8*, 85–114. <https://doi.org/10.1007/bf03380183>.
- (75) Peralta, D. Diversity in Medicinal Chemistry. *ChemMedChem* **2006**, *6* (1), 3–6. <https://doi.org/10.1002/cmdc.201700803>.
- (76) Gilad, Y.; Nadassy, K.; Senderowitz, H. A Reliable Computational Workflow for the Selection of Optimal Screening Libraries. *J. Cheminform.* **2015**, *7* (61), 1–17. <https://doi.org/10.1186/s13321-015-0108-0>.
- (77) Gromski, P. S.; Henson, A. B.; Granda, J. M.; Cronin, L. How to Explore Chemical Space Using Algorithms and Automation. *Nat. Rev. Chem.* **2019**, *3* (2), 119–128. <https://doi.org/10.1038/s41570-018-0066-y>.
- (78) Waldman, M.; Li, H.; Hassan, M. Novel Algorithms for the Optimization of Molecular Diversity of Combinatorial Libraries. *J. Mol. Graph. Model.* **2000**, *18* (4–5), 412–426. [https://doi.org/10.1016/S1093-3263\(00\)00071-1](https://doi.org/10.1016/S1093-3263(00)00071-1).

- (79) Martin, Y. C. Challenges and Prospects for Computational Aids to Molecular Diversity. *Perspect. Drug Discov. Des.* **1997**, 7--8, 159–172. <https://doi.org/10.1007/bf03380186>.
- (80) Ashton, M.; Barnard, J.; Casset, F.; Charlton, M.; Downs, G.; Gorse, D.; Holliday, J.; Lahana, R.; Willett, P. Identification of Diverse Database Subsets Using Property-Based and Fragment-Based Molecular Descriptions. *Quant. Struct. Relationships* **2002**, 21 (6), 598–604. <https://doi.org/10.1002/qsar.200290002>.
- (81) Todeschini, R.; Consonni, V. *Molecular Descriptors for Chemoinformatics: Volume I: Alphabetical Listing / Volume II: Appendices, References, Volume 41*; 2009.
- (82) Willett, P. Similarity and Clustering in Chemical Information Systems. *Technometrics* **1990**, 32 (3), 359–360. <https://doi.org/10.1080/00401706.1990.10484713>.
- (83) Bratchell, N. Cluster Analysis. *Chemom. Intell. Lab. Syst.* **1989**, 6 (2), 105–125. [https://doi.org/10.1016/0169-7439\(87\)80054-0](https://doi.org/10.1016/0169-7439(87)80054-0).
- (84) Barnard, J. M.; Downs, G. M. Clustering of Chemical Structures on the Basis of Two-Dimensional Similarity Measures. *J. Chem. Inf. Comput. Sci.* **1992**, 32 (6), 644–649. <https://doi.org/10.1021/ci00010a010>.
- (85) Downs, G. M.; Willett, P. Clustering of Chemical Structure Databases for Compound Selection. In *Advanced Computer-Assisted Techniques in Drug Discovery*; van de Waterbeemd, H., Ed.; Methods and Principles in Medicinal Chemistry; Wiley, 2008; Vol. 3, pp 111–130. <https://doi.org/10.1002/9783527615674>.
- (86) Downs, G. M.; Barnard, J. M. *Clustering Methods and Their Uses in Computational Chemistry*; 2002; Vol. 18. <https://doi.org/10.1002/0471433519.ch1>.
- (87) Downs, G. M.; Willett, P.; Fisanick, W. Similarity Searching and Clustering of Chemical-Structure Databases Using Molecular Property Data. *J. Chem. Inf. Comput. Sci.* **1994**, 34 (5), 1094–1102. <https://doi.org/10.1021/ci00021a011>.
- (88) Jamois, E. A.; Hassan, M.; Waldman, M. Evaluation of Reagent-Based and Product-Based Strategies in the Design of Combinatorial Library Subsets. *J. Chem. Inf. Comput. Sci.* **2000**, 40 (1), 63–70. <https://doi.org/10.1021/ci990015k>.
- (89) Willett, P.; Winterman, V.; Bawden, D. Implementation of Nonhierarchical Cluster Analysis Methods in Chemical Information Systems: Selection of Compounds for Biological Testing and Clustering of Substructure Search Output. *J. Chem. Inf. Comput. Sci.* **1986**, 26 (3), 109–118. <https://doi.org/10.1021/ci00051a005>.
- (90) Pascual, R.; Borrell, J. I.; Teixidó, J. Analysis of Selection Methodologies for Combinatorial Library Design. *Mol. Divers.* **2003**, 6 (2), 121–133. <https://doi.org/10.1023/B:MODI.0000006836.76687.8b>.
- (91) Puig-De-La-Bellacasa, R.; Giménez, L.; Pettersson, S.; Pascual, R.; Gonzalo, E.; Esté, J. A.; Clotet, B.; Borrell, J. I.; Teixidó, J. Diverse Combinatorial Design, Synthesis and in Vitro Evaluation of New HEPT Analogues as Potential Non-Nucleoside HIV-1 Reverse Transcription Inhibitors. *Eur. J. Med. Chem.* **2012**, 54, 159–174. <https://doi.org/10.1016/j.ejmech.2012.04.038>.
- (92) Lin, A.; Horvath, D.; Afonina, V.; Marcou, G.; Reymond, J. L.; Varnek, A. Mapping of the Available Chemical Space versus the Chemical Universe of Lead-Like Compounds. *ChemMedChem* **2018**, 13 (6), 540–554. <https://doi.org/10.1002/cmdc.201700561>.

- (93) Fink, T.; Raymond, J. L. Virtual Exploration of the Chemical Universe up to 11 Atoms of C, N, O, F: Assembly of 26.4 Million Structures (110.9 Million Stereoisomers) and Analysis for New Ring Systems, Stereochemistry, Physicochemical Properties, Compound Classes, and Drug Discover. *J. Chem. Inf. Model.* **2007**, *47* (2), 342–353. <https://doi.org/10.1021/ci600423u>.
- (94) Ruffolo, R. R. Why Has R&D Productivity Declined in the Pharmaceutical Industry? *Expert Opin. Drug Discov.* **2006**, *1* (2), 99–102. <https://doi.org/10.1517/17460441.1.2.99>.
- (95) Scannell, J. W.; Blanckley, A.; Boldon, H.; Warrington, B. Diagnosing the Decline in Pharmaceutical R&D Efficiency. *Nat. Rev. Drug Discov.* **2012**, *11* (3), 191–200. <https://doi.org/10.1038/nrd3681>.
- (96) Mak, K. K.; Pichika, M. R. Artificial Intelligence in Drug Development: Present Status and Future Prospects. *Drug Discov. Today* **2019**, *24* (3), 773–780. <https://doi.org/10.1016/j.drudis.2018.11.014>.
- (97) Pushpakom, S.; Iorio, F.; Eysers, P. A.; Escott, K. J.; Hopper, S.; Wells, A.; Doig, A.; Williams, T.; Latimer, J.; McNamee, C.; Norris, A.; Sanseau, P.; Cavalla, D.; Pirmohamed, M. Drug Repurposing: Progress, Challenges and Recommendations. *Nat. Rev. Drug Discov.* **2018**, *18* (1), 41–58. <https://doi.org/10.1038/nrd.2018.168>.
- (98) Zitnik, M.; Nguyen, F.; Wang, B.; Leskovec, J.; Goldenberg, A.; Hoffman, M. M. Machine Learning for Integrating Data in Biology and Medicine: Principles, Practice, and Opportunities. *Inf. Fusion* **2019**, *50*, 71–91. <https://doi.org/10.1016/j.inffus.2018.09.012>.
- (99) Lavecchia, A. Deep Learning in Drug Discovery: Opportunities, Challenges and Future Prospects. *Drug Discov. Today* **2019**, *24* (10), 2017–2032. <https://doi.org/10.1016/j.drudis.2019.07.006>.
- (100) Jourdan, J. P.; Bureau, R.; Rochais, C.; Dallemagne, P. Drug Repositioning: A Brief Overview. *J. Pharm. Pharmacol.* **2020**, *72* (9), 1145–1151. <https://doi.org/10.1111/jphp.13273>.
- (101) Brown, D. G.; Wobst, H. J. A Decade of FDA-Approved Drugs (2010-2019): Trends and Future Directions. *J. Med. Chem.* **2021**, *64* (5), 2312–2338. <https://doi.org/10.1021/acs.jmedchem.0c01516>.
- (102) World Health Organization (WHO). ICD-11 for Mortality and Morbidity Statistics <https://icd.who.int/> (accessed Jul 20, 2021).
- (103) Rabal, O.; Oyarzabal, J. Biologically Relevant Chemical Space Navigator: From Patent and Structure-Activity Relationship Analysis to Library Acquisition and Design. *J. Chem. Inf. Model.* **2012**, *52* (12), 3123–3137. <https://doi.org/10.1021/ci3004539>.
- (104) Fakhoury, S. A.; Lee, H. T.; Reed, J. E.; Schlosser, K. M.; Sexton, K. E.; Teclé, H.; Winters, R. T. 4-PHENYLAMINO-QUINAZOLIN-6-YL-AMIDES. US 7,772,243 B2, 2010.
- (105) Mcchesney, J.; Nanayakkra, D. N.; Bartlerr, M.; Ager, A. L. 8-Aminoquinolines. US 2001/0007031 A1, 2001.
- (106) Mascitti, V.; Collman, B. M. DIOXA-BICYCLO[3.2.1]OCTANE-2,3,4-TRIOLE DERIVATIVES. US 2010/0056618 A1, 2010.
- (107) Coates, D. A.; Lawrence, M.; Knobloch, J. M.; De Dios Magna, A.; De Prado Gonzalez, A.; Del Prado Catalina, M. F.; Garcia Paredes, M. C.; Martin de la Nava, M.; Martin Ortega Finger, M.

- D.; Martinez Perez, J. A.; Mateo Herranz, A. I.; Perez Martinez, C.; Sanchez Martinez, C. PROTEIN KINASE INHIBITORS. US 2010/0160340 A1, 2010.
- (108) Heubach, G. 5-METHYL-ISOXAZOLE-4-CARBOXYLIC ACID ANILIDES. US 4,087,535, 1978.
- (109) Blumbergs, P.; LaMontagne, M. P. 4-Methyl-5-(Unsubstituted and Substituted Phenoxy)-2,6-Dimethoxy-8- (Aminoalkylamino) Quinolines. U.S. Patent 4,617,394, 1986.
- (110) Kuroita, T.; Sakamoto, H.; Ojima, M. BENZIMIDAZOLE DERIVATIVE AND USE AS A RECEPTOR ANTAGONIST. US 7,572,920 B2, 2009.
- (111) Meier, R. Fluorinated Benzyl Triazole Compounds. EP 0199262, 1986.
- (112) Kabi, F.; Pharma, G.; Bv, P.; Co, U.; Bv, P.; Gmbh, C. O.; Bv, P.; Gmbh, C. O.; Bv, P.; Hydrochloride, H.; Hydrochloride, H.; Hydrochloride, H.; Hydrochloride, H. Orange Book: Approved Drug Products with Therapeutic Equivalence Evaluations https://www.accessdata.fda.gov/scripts/cder/ob/index.cfm%0Ahttps://www.accessdata.fda.gov/scripts/cder/ob/index.cfm%0Ahttps://www.accessdata.fda.gov/scripts/cder/ob/patent_info.cfm?Product_No=001&Appl_No=022544&Appl_type=N (accessed Jul 20, 2021).
- (113) Farrar, J. J.; Schield, P. J.; Schmidt, W. K.; Carpenter, R. L. Methods for the Treatment and Prevention of Ileus. US 6,469,030 B2, 2002.
- (114) Brown, N. Chemoinformatics—An Introduction for Computer Scientists. *ACM Comput. Surv.* **2009**, *41* (2), 1–38. <https://doi.org/10.1145/1459352.1459353>.
- (115) Halkidi, M.; Vazirgiannis, M.; Balislakis, V. Quality Scheme Assessment in the Clustering Process. *Lect. Notes Comput. Sci.* **2000**, *1910*, 265–276. https://doi.org/10.1007/3-540-45372-5_26.
- (116) Rousseeuw, P. J. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *J. Comput. Appl. Math.* **1987**, *20* (C), 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
- (117) Mur, A.; Dormido, R.; Duro, N.; Dormido-Canto, S.; Vega, J. Determination of the Optimal Number of Clusters Using a Spectral Clustering Optimization. *Expert Syst. Appl.* **2016**, *65*, 304–314. <https://doi.org/10.1016/j.eswa.2016.08.059>.
- (118) Hruschka, E. R.; Covões, T. F. Feature Selection for Cluster Analysis: An Approach Based on the Simplified Silhouette Criterion. *Proc. - Int. Conf. Comput. Intell. Model. Control Autom. CIMCA 2005 Int. Conf. Intell. Agents, Web Technol. Internet* **2005**, *1*, 32–37. <https://doi.org/10.1109/cimca.2005.1631238>.
- (119) Nguyen, L. H.; Holmes, S. Ten Quick Tips for Effective Dimensionality Reduction. *PLOS Comput. Biol.* **2019**, *15* (6), e1006907. <https://doi.org/10.1371/journal.pcbi.1006907>.
- (120) Kaya, M.-F.; Schoop, M. Analytical Comparison of Clustering Techniques for the Recognition of Communication Patterns. *Gr. Decis. Negot.* **2022**, *31* (3), 555–589. <https://doi.org/10.1007/s10726-021-09758-7>.
- (121) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem in 2021: New Data Content and Improved Web Interfaces. *Nucleic Acids Res.* **2019**, *49*(D1), D1388–D1395. <https://doi.org/https://doi.org/10.1093/nar/gkaa971>.

- (122) Dutta, A. K.; Avery, B. A.; Wyandt, C. M. Development and Validation of a Stability-Indicating Reversed-Phase High Performance Liquid Chromatography Method for NPC 1161C, a Novel 8-Aminoquinoline Anti-Malarial Drug. *J. Chromatogr. A* **2006**, *1110* (1–2), 35–45. <https://doi.org/10.1016/j.chroma.2006.01.040>.
- (123) Strube, R. E.; LaMontagne, M. P. 4-Methyl-5-(Unsubstituted or Substituted Phenoxy)6-Methoxy-8-(Aminoalkylamino)Quinolines,. U.S. Patent 4,431,807, 1984.
- (124) Nodiff, E. A.; Tanabe, K.; Chen, E. H.; Saggiomo, A. J. Modifications of Primaquine as Antimalarials. 3. 5-Phenoxy Derivatives of Primaquine. *J. Med. Chem.* **1982**, *4816* (30), 1097–1101. <https://doi.org/10.1021/jm00351a018>.
- (125) Azad, C. S.; Saxena, M.; Siddiqui, A. J.; Bhardwaj, J.; Puri, S. K.; Dutta, G. P.; Anand, N.; Saxena, A. K. Synthesis of Primaquine Glyco-Conjugates as Potential Tissue Schizontocidal Antimalarial Agents. *Chem. Biol. Drug Des.* **2017**, *90* (2), 254–261. <https://doi.org/10.1111/cbdd.12944>.
- (126) LaMontagne, M. P.; Blumbergs, P.; Strube, R. E. Antimalarials. 14. 5-(Aryloxy)-4-Methylprimaquine Analogues. A Highly Effective Series of Blood and Tissue Schizonticidal Agents. *J. Med. Chem.* **1982**, *25* (9), 1094–1097. <https://doi.org/10.1021/jm00351a017>.
- (127) Ashley, E. A.; Phyto, A. P. Drugs in Development for Malaria. *Drugs* **2018**, *78* (9), 861–879. <https://doi.org/10.1007/s40265-018-0911-9>.
- (128) Mcchesney, J.; Nanayakkra, D. N.; Bartlerr, M.; Ager, A. L. 8-AMINOQUINOLINES. WO 97/36590, 1997.
- (129) Chaurasiya, N. D.; Ganesan, S.; Nanayakkara, N. P. D.; Dias, L. R. S.; Walker, L. A.; Tekwani, B. L. Inhibition of Human Monoamine Oxidase A and B by 5-Phenoxy 8-Aminoquinoline Analogs. *Bioorganic Med. Chem. Lett.* **2012**, *22* (4), 1701–1704. <https://doi.org/10.1016/j.bmcl.2011.12.108>.
- (130) World Health Organization (WHO). *World Malaria Report 2021*; 2021.
- (131) Ebstire, Y. A.; Abay, S. M.; Tadesse, W. T.; Ejigu, D. A. Tafenoquine and Its Potential in the Treatment and Relapse Prevention of Plasmodium Vivax Malaria: The Evidence to Date. *Drug Des. Devel. Ther.* **2016**, *10*, 2387–2399. <https://doi.org/10.2147/DDDT.S61443>.
- (132) Guerin, P. J.; Olliaro, P.; Nosten, F.; Druilhe, P.; Laxminarayan, R.; Binka, F.; Kilama, W. L.; Ford, N.; White, N. J. Malaria: Current Status of Control, Diagnosis, Treatment, and a Proposed Agenda for Research and Development. *Lancet Infect. Dis.* **2002**, *2* (9), 564–573. [https://doi.org/10.1016/S1473-3099\(02\)00372-9](https://doi.org/10.1016/S1473-3099(02)00372-9).
- (133) Popovici, J.; Ménard, D. Challenges in Antimalarial Drug Treatment for Vivax Malaria Control. *Trends Mol. Med.* **2015**, *21* (12), 776–788. <https://doi.org/10.1016/j.molmed.2015.10.004>.
- (134) Price, R. N.; Tjitra, E.; Guerra, C. A.; Yeung, S.; White, N. J.; Anstey, N. M. Vivax Malaria: Neglected and Not Benign. *Am. J. Trop. Med. Hyg.* **2007**, *77* (6 Suppl), 79–87.
- (135) Gringauz, A. *Introduction to Medicinal Chemistry: How Drugs Act and Why*; WILEY-VCH, 1996.
- (136) Singh, J. Centers for disease control and prevention <https://www.cdc.gov/malaria/about/biology/index.html> (accessed Jul 20, 2022). <https://doi.org/10.1097/jom.0000000000001045>.
- (137) Orphanet <https://www.orpha.net/> (accessed Sep 10, 2022).

- (138) WHO. Global Malaria Programme ; Elimination [https://www.who.int/teams/global-malaria-programme/case-management/diagnosis/rapid-diagnostic-tests/how-malaria-rdts-work#:~:text=Malaria rapid diagnostic tests \(RDTs, services cannot be readily provided.%0Ahttps://www.who.int/teams/global-malaria-programm](https://www.who.int/teams/global-malaria-programme/case-management/diagnosis/rapid-diagnostic-tests/how-malaria-rdts-work#:~:text=Malaria rapid diagnostic tests (RDTs, services cannot be readily provided.%0Ahttps://www.who.int/teams/global-malaria-programm) (accessed Sep 5, 2022).
- (139) Savirz, D. A.; Styka, A. N. *Assessment of Long-Term Health Effects of Antimalarial Drugs When Used for Prophylaxis*; 2020. <https://doi.org/10.17226/25688>.
- (140) Okombo, J.; Chibale, K. Recent Updates in the Discovery and Development of Novel Antimalarial Drug Candidates. *Medchemcomm* **2018**, *9* (3), 437–453. <https://doi.org/10.1039/c7md00637c>.
- (141) Adams, J. H.; Mueller, I. The Biology of *Plasmodium Vivax*. *Cold Spring Harb. Perspect. Med.* **2017**, *7* (9), a025585. <https://doi.org/10.1101/cshperspect.a025585>.
- (142) Shanks, G. D.; Oloo, A. J.; Aleman, G. M.; Ohrt, C.; Klotz, F. W.; Braitman, D.; Horton, J.; Brueckner, R. A New Primaquine Analogue, Tafenoquine (WR 238605), for Prophylaxis against *Plasmodium Falciparum* Malaria. *Clin. Infect. Dis.* **2001**, *33* (12), 1968–1974. <https://doi.org/10.1086/324081>.
- (143) Brueckner, R. P.; Coster, T.; Wesche, D. L.; Shmuklarsky, M.; Schuster, B. G. Prophylaxis of *Plasmodium Falciparum* Infection in a Human Challenge Model with WR 238605 , a New 8-Aminoquinoline Antimalarial. **1998**, *42* (5), 1293–1294.
- (144) Brueckner, R. P.; Lasseter, K. C.; Lin, E. T.; Schuster, B. G. First-Time-in-Humans Safety and Pharmacokinetics of WR238605 , a New Antimalarial. *Am. J. Trop. Med. Hyg.* **1998**, *58* (5), 645–649.
- (145) Gutteridge, W. E. Antimalarial Drugs Currently in Development. *J. R. Soc. Med.* **1989**, *82 Suppl 1*, 63–66; discussion 66-8.
- (146) Davidson, D. E.; Ager, A. L.; Brown, J. L.; Chapple, F. E.; Whitmire, R. E.; Rossan, R. N. New Tissue Schizontocidal Antimalarial Drugs. *Bull. World Health Organ.* **1981**, *59* (3), 463–479.
- (147) Peters, W. The Evolution of Tafenoquine—Antimalarial for a New Millennium? *J. R. Soc. Med.* **1999**, *92* (7), 345–352. <https://doi.org/10.1177/014107689909200705>.
- (148) Lacerda, M. V. G.; Llanos-Cuentas, A.; Krudsood, S.; Lon, C.; Saunders, D. L.; Mohammed, R.; Yilma, D.; Batista Pereira, D.; Espino, F. E. J.; Mia, R. Z.; Chuquiyaui, R.; Val, F.; Casapía, M.; Monteiro, W. M.; Brito, M. A. M.; Costa, M. R. F.; Buathong, N.; Noedl, H.; Diro, E.; Getie, S.; Wubie, K. M.; Abdissa, A.; Zeynudin, A.; Abebe, C.; Tada, M. S.; Brand, F.; Beck, H.-P.; Angus, B.; Duparc, S.; Kleim, J.-P.; Kellam, L. M.; Rousell, V. M.; Jones, S. W.; Hardaker, E.; Mohamed, K.; Clover, D. D.; Fletcher, K.; Breton, J. J.; Ugwuegbulam, C. O.; Green, J. A.; Koh, G. C. K. W. Single-Dose Tafenoquine to Prevent Relapse of *Plasmodium Vivax* Malaria . *N. Engl. J. Med.* **2019**, *380* (3), 215–228. <https://doi.org/10.1056/nejmoa1710775>.
- (149) Llanos-Cuentas, A.; Lacerda, M. V. G.; Hien, T. T.; Vélez, I. D.; Namaik-larp, C.; Chu, C. S.; Villegas, M. F.; Val, F.; Monteiro, W. M.; Brito, M. A. M.; Costa, M. R. F.; Chuquiyaui, R.; Casapía, M.; Nguyen, C. H.; Aruachan, S.; Papwijitsil, R.; Nosten, F. H.; Bancone, G.; Angus, B.; Duparc, S.; Craig, G.; Rousell, V. M.; Jones, S. W.; Hardaker, E.; Clover, D. D.; Kendall, L.; Mohamed, K.; Koh, G. C. K. W.; Wilches, V. M.; Breton, J. J.; Green, J. A. Tafenoquine versus Primaquine to Prevent Relapse of *Plasmodium Vivax* Malaria . *N. Engl. J. Med.* **2019**, *380* (3), 229–241. <https://doi.org/10.1056/nejmoa1802537>.

- (150) Medicines for Malaria Venture (MMV). US FDA Approves Krintafel (Tafenoquine) for the Radical Cure of *P. Vivax* Malaria 2018. *Press Release* **2018**, No. July.
- (151) Burrows, J. N.; Hooft van Huijsduijnen, R.; Möhrle, J. J.; Oeuvray, C.; Wells, T. N. Designing the next Generation of Medicines for Malaria Control and Eradication. *Malar. J.* **2013**, *12* (1), 187. <https://doi.org/10.1186/1475-2875-12-187>.
- (152) Val, F.; Costa, F. T. M.; King, L.; Brito-Sousa, J. D.; Bassat, Q.; Monteiro, W. M.; Siqueira, A. M.; Luzzatto, L.; Lacerda, M. V. G. Tafenoquine for the Prophylaxis, Treatment and Elimination of Malaria: Eagerness Must Meet Prudence. *Future Microbiol.* **2019**, *14* (15), 1261–1279. <https://doi.org/10.2217/fmb-2019-0202>.
- (153) Battle, K. E.; Baird, J. K. The Global Burden of Plasmodium Vivax Malaria Is Obscure and Insidious. *PLOS Med.* **2021**, *18* (10), e1003799. <https://doi.org/10.1371/journal.pmed.1003799>.
- (154) Rochford, R.; Ohrt, C.; Baresel, P. C.; Campo, B.; Sampath, A.; Magill, A. J. Humanized Mouse Model of Glucose 6-Phosphate Dehydrogenase deficiency for in Vivo Assessment of Hemolytic Toxicity. **2013**, *110* (43). <https://doi.org/10.1073/pnas.1310402110>.
- (155) McChesney, J.; Nanayakkara, D. N.; Barlett, M.; Ager, A. L. 8-Aminoquinolines. U.S. Patent 6,376,511 B2, 2002.
- (156) PubChem <https://pubchem.ncbi.nlm.nih.gov/> (accessed Jul 20, 2021).
- (157) LaMontagne, M. P.; Blumbergs, P. Antimalarials. 15. Side-Chain Analogues of 8-(4-Amino-1-Methylbutylamino)-6-Methoxy-4-Methyl-5-(3-Trifluoromethylphenoxy)Quinoline. *J. Heterocycl. Chem.* **1984**, *21*, 33–35.
- (158) Gamo, F.; Sanz, L. M.; Vidal, J.; Cozar, C. De; Alvarez, E.; Lavandera, J.; Vanderwall, D. E.; Green, D. V. S.; Kumar, V.; Hasan, S.; Brown, J. R.; Peishoff, C. E.; Cardon, L. R.; Garcia-bustos, J. F. Thousands of Chemical Starting Points for Antimalarial Lead Identification. *Nature* **2010**, *465* (7296), 305–310. <https://doi.org/10.1038/nature09107>.
- (159) Teixeira, C.; Vale, N.; Pe, B.; Gomes, A.; Gomes, J. R. B.; Gomes, P. “ Recycling ” Classical Drugs for Malaria. *Chem. Rev.* **2014**, *114* (Drug Discovery and Development for Neglected Diseases), 11164–11220. <https://doi.org/10.1021/cr500123g>.
- (160) Queener, S. F.; Bartlett, M. S.; Nasr, M.; Smith, J. W. 8-Aminoquinolines Effective against *Pneumocystis Carinii* In Vitro and In Vivo. **1993**, *37* (10), 2166–2172.
- (161) Singh, K.; Kaur, T. Pyrimidine-Based Antimalarials: Design Strategies and Antiplasmodial Effects. *Med. Chem. commun.* **2016**, *7* (2016), 749–768. <https://doi.org/10.1039/c6md00084c>.
- (162) Swain, M. ChemSpiPy <https://chemspipy.readthedocs.io/> (accessed Oct 2, 2022).
- (163) Royal Society of Chemistry. ChemSpider. Search and Share Chemistry <http://www.chemspider.com/> (accessed Aug 1, 2022).
- (164) Lipinski, C. A. Drug-like Properties and the Causes of Poor Solubility and Poor Permeability. *J. Pharmacol. Toxicol. Methods* **2000**, *44* (1), 235–249. [https://doi.org/10.1016/S1056-8719\(00\)00107-6](https://doi.org/10.1016/S1056-8719(00)00107-6).
- (165) Hirth, P. K.; Elaina, M.; Shawyer, L. K.; Ullrich, A.; Szekely, I.; Bajor, T.; Haimichael, J.; Orfi, L.;

- Levitcki, A.; Gazit, A.; Tang, P. C.; Lammers, R. Treatment of Platelet Derived Growth Factor Related Disorders Such as Cancers. U.S. Patent No. 6,331,555 B1, 2001.
- (166) WULFMAN, D. S.; COOPER, C. F. Monoreduction of Dinitroarenes with Iron/Acetic Acid. *Synth. Commun.* **1978**, *1978* (12), 924–925. <https://doi.org/10.1055/s-1978-24942>.
- (167) Jain, R.; Jain, S.; Gupta, R. C.; Anand, N. Synthesis of Amino Acid Derivatives of 8- [4-Amino-1-Methylbutylamino] -6- Schizontocidal Anti-Malarial Agents. **1994**, No. January.
- (168) Carroll, F. I.; Berrang, B.; Linn, C. P. 4-Substituted 5-[m -(Trifluoromethyl)Phenoxy]Primaquine Analogues as Potential Antimalarial Agents. *J. Med. Chem.* **1985**, *28* (11), 1564–1567. <https://doi.org/10.1021/jm00149a004>.
- (169) Fluorochem <http://www.fluorochem.co.uk/> (accessed Aug 4, 2022).
- (170) Bell, D.; Davies, B. J.; Kinsey, P. M. Process for the Preparation of Quinoline Derivatives. WO 03/093239, 2003.
- (171) Wacowich-Sgarbi, S. *CHEM 1114 – INTRODUCTION TO CHEMISTRY*, 1st ed.; Langara Chemistry Department, 2018.
- (172) Tekwani, B. L.; Walker, L. A. 8-Aminoquinolines: Future Role as Antiprotozoal Drugs. *Curr. Opin. Infect. Dis.* **2006**, *19* (6), 623–631. <https://doi.org/10.1097/QCO.0b013e328010b848>.
- (173) Leven, M.; Held, J.; Duffy, S.; Alves Avelar, L. A.; Meister, S.; Delves, M.; Plouffe, D.; Kuna, K.; Tschan, S.; Avery, V. M.; Winzeler, E. A.; Mordmüller, B.; Kurz, T. 8-Aminoquinolines with an Aminoxyalkyl Side Chain Exert in Vitro Dual-Stage Antiplasmodial Activity. *ChemMedChem* **2019**, *14* (4), 501–511. <https://doi.org/10.1002/cmdc.201800691>.
- (174) Maher, S. P.; Vantaux, A.; Chaumeau, V.; Chua, A. C. Y.; Cooper, C. A.; Andolina, C.; Péneau, J.; Rouillier, M.; Rizopoulos, Z.; Phal, S.; Piv, E.; Vong, C.; Phen, S.; Chhin, C.; Tat, B.; Ouk, S.; Doeurk, B.; Kim, S.; Suriyakan, S.; Kittiphanakun, P.; Awuku, N. A.; Conway, A. J.; Jiang, R. H. Y.; Russell, B.; Bifani, P.; Campo, B.; Nosten, F.; Witkowski, B.; Kyle, D. E. Probing the Distinct Chemosensitivity of Plasmodium Vivax Liver Stage Parasites and Demonstration of 8-Aminoquinoline Radical Cure Activity in Vitro. *Sci. Rep.* **2021**, *11* (1), 19905. <https://doi.org/10.1038/s41598-021-99152-9>.
- (175) Bermúdez, M.; Moreno-Pérez, D. A.; Arévalo-Pinzón, G.; Curtidor, H.; Patarroyo, M. A. Plasmodium Vivax in Vitro Continuous Culture: The Spoke in the Wheel. *Malar. J.* **2018**, *17* (1), 301. <https://doi.org/10.1186/s12936-018-2456-5>.
- (176) Russell, B.; Suwanarusk, R.; Borlon, C.; Costa, F. T. M.; Chu, C. S.; Rijken, M. J.; Sriprawat, K.; Warter, L.; Koh, E. G. L.; Malleret, B.; Colin, Y.; Bertrand, O.; Adams, J. H.; D’Alessandro, U.; Snounou, G.; Nosten, F.; Rénia, L. A Reliable Ex Vivo Invasion Assay of Human Reticulocytes by Plasmodium Vivax. *Blood* **2011**, *118* (13). <https://doi.org/10.1182/blood-2011-04-348748>.
- (177) Desjardins, R. E.; Canfield, C. J.; Haynes, J. D.; Chulay, J. D. Quantitative Assessment of Antimalarial Activity in Vitro by a Semiautomated Microdilution Technique. *Antimicrob. Agents Chemother.* **1979**, *16* (6), 710–718. <https://doi.org/10.1128/AAC.16.6.710>.
- (178) Sidhu, A. B. S.; Verdier-Pinard, D.; Fidock, D. A. Chloroquine Resistance in Plasmodium Falciparum Malaria Parasites Conferred by Pfcr1 Mutations. *Science (80-.)*. **2002**, *298* (5591), 210–213. <https://doi.org/10.1126/science.1074045>.

- (179) Ridley, R. G. Medical Need, Scientific Opportunity and the Drive for Antimalarial Drugs. *Nature* **2002**, *415* (6872), 686–693. <https://doi.org/10.1038/415686a>.
- (180) OMS. International Agency for Research on Cancer - World Health Organization <https://gco.iarc.fr/today/home> (accessed Aug 16, 2022).
- (181) Hanahan, D.; Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **2011**, *144* (5), 646–674. <https://doi.org/10.1016/j.cell.2011.02.013>.
- (182) Abraham, D. J. *Burger's Medicinal Chemistry and Drug Discovery*; Wiley, 2003. <https://doi.org/10.1002/0471266949>.
- (183) WHO. World Health Organization.. - World Health Organization <https://www.who.int/en/news-room/fact-sheets/detail/arsenic> (accessed Aug 16, 2022).
- (184) National Cancer Institute. Risk Factors for Cancer - National Cancer Institute <https://www.cancer.gov/about-cancer/causes-prevention/risk> (accessed Aug 16, 2022).
- (185) Anand, P.; Kunnumakara, A. B.; Sundaram, C.; Harikumar, K. B.; Tharakan, S. T.; Lai, O. S.; Sung, B.; Aggarwal, B. B. Cancer Is a Preventable Disease That Requires Major Lifestyle Changes. *Pharm. Res.* **2008**, *25* (9), 2097–2116. <https://doi.org/10.1007/s11095-008-9661-9>.
- (186) American Cancer society. Lung cancer statistics. How common is lung cancer.
- (187) de Groot, P. M.; Wu, C. C.; Carter, B. W.; Munden, R. F. The Epidemiology of Lung Cancer. *Transl. Lung Cancer Res.* **2018**, *7* (3), 220–233. <https://doi.org/10.21037/tlcr.2018.05.06>.
- (188) Cancer Research UK. Cancer Research UK - Lung Cancer <https://www.cancerresearchuk.org/about-cancer/lung-cancer> (accessed Aug 17, 2022).
- (189) GBD Compare | IHME Viz Hub - Data Visualizations <https://vizhub.healthdata.org/gbd-compare/> (accessed Aug 17, 2022).
- (190) Sato, M.; Shames, D. S.; Gazdar, A. F.; Minna, J. D. A Translational View of the Molecular Pathogenesis of Lung Cancer. *J. Thorac. Oncol.* **2007**, *2* (4), 327–343. <https://doi.org/10.1097/01.JTO.0000263718.69320.4c>.
- (191) Jiao, Q.; Bi, L.; Ren, Y.; Song, S.; Wang, Q.; Wang, Y. Advances in Studies of Tyrosine Kinase Inhibitors and Their Acquired Resistance. *Mol. Cancer* **2018**, *17* (1), 36. <https://doi.org/10.1186/s12943-018-0801-5>.
- (192) Arechaga-Ocampo, E.; Villegas-Sepulveda, N.; Lopez-Urrutia, E.; Ramos-Suzarte, M.; Lopez-Camarillo, C.; Perez-Plasencia, C.; la Rosa, C. H. G.; Cortes-Gonzalez, C.; A., L. Biomarkers in Lung Cancer: Integration with Radiogenomics Data. In *Oncogenomics and Cancer Proteomics - Novel Approaches in Biomarkers Discovery and Therapeutic Targets in Cancer*; InTech, 2013. <https://doi.org/10.5772/53426>.
- (193) Choong, N. W.; Salgia, R.; Vokes, E. E. Key Signaling Pathways and Targets in Lung Cancer Therapy. *Clin. Lung Cancer* **2007**, *8* (SUPPL. 2), S52–S60. <https://doi.org/10.3816/CLC.2007.s.002>.
- (194) Cohen, P.; Cross, D.; Jänne, P. A. Kinase Drug Discovery 20 Years after Imatinib. *Nat. Rev. Drug Discov.* **2022**, 41573. <https://doi.org/10.1038/s41573-022-00418-2>.
- (195) Vu, P.; Patel, S. P. Non-Small Cell Lung Cancer Targetable Mutations: Present and Future.

- Precis. Cancer Med.* **2020**, *3* (March). <https://doi.org/10.21037/pcm.2019.11.03>.
- (196) Gazdar, A. F. Activating and Resistance Mutations of EGFR in Non-Small-Cell Lung Cancer: Role in Clinical Response to EGFR Tyrosine Kinase Inhibitors. *Oncogene* **2009**, *28* (S1), S24–S31. <https://doi.org/10.1038/onc.2009.198>.
- (197) Rosell, R.; Morán, T.; Carcereny, E.; Quiroga, V.; Molina, M. Á.; Costa, C.; Benlloch, S.; Tarón, M. Non-Small-Cell Lung Cancer Harboring Mutations in the EGFR Kinase Domain. *Clin. Transl. Oncol.* **2010**, *12* (2), 75–80. <https://doi.org/10.1007/S12094-010-0473-0>.
- (198) Attwood, M. M.; Fabbro, D.; Sokolov, A. V.; Knapp, S.; Schiöth, H. B. Trends in Kinase Drug Discovery: Targets, Indications and Inhibitor Design. *Nat. Rev. Drug Discov.* **2021**, *20* (11), 839–861. <https://doi.org/10.1038/s41573-021-00252-y>.
- (199) Girard, N. Optimizing Outcomes in EGFR Mutation-Positive NSCLC: Which Tyrosine Kinase Inhibitor and When? *Futur. Oncol.* **2018**, *14* (11), 1117–1132. <https://doi.org/10.2217/fon-2017-0636>.
- (200) Kujtan, L.; Subramanian, J. Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitors for the Treatment of Non-Small Cell Lung Cancer. *Expert Rev. Anticancer Ther.* **2019**, *19* (7), 547–559. <https://doi.org/10.1080/14737140.2019.1596030>.
- (201) Castellanos, E. H.; Horn, L. Generations of Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitors: Perils and Progress. *Curr. Treat. Options Oncol.* **2015**, *16* (10), 51. <https://doi.org/10.1007/s11864-015-0365-1>.
- (202) Kohsaka, S.; Petronczki, M.; Solca, F.; Maemondo, M. Tumor Clonality and Resistance Mechanisms in EGFR Mutation-Positive Non-Small-Cell Lung Cancer: Implications for Therapeutic Sequencing. *Futur. Oncol.* **2019**, *15* (6), 637–652. <https://doi.org/10.2217/fon-2018-0736>.
- (203) Brabender, J.; Danenberg, K. D.; Metzger, R.; Schneider, P. M.; Park, J. M.; Salonga, D.; Hölscher, A. H.; Danenberg, P. V. Epidermal Growth Factor Receptor and HER2-Neu mRNA Expression in Non-Small Cell Lung Cancer Is Correlated with Survival. *Clin. Cancer Res.* **2001**, *7* (7), 1850–1855.
- (204) Leonetti, A.; Sharma, S.; Minari, R.; Perego, P.; Giovannetti, E.; Tiseo, M. Resistance Mechanisms to Osimertinib in EGFR-Mutated Non-Small Cell Lung Cancer. *Br. J. Cancer* **2019**, *121* (9), 725–737. <https://doi.org/10.1038/s41416-019-0573-8>.
- (205) Zeng, Y.; Yu, D.; Tian, W.; Wu, F. Resistance Mechanisms to Osimertinib and Emerging Therapeutic Strategies in Nonsmall Cell Lung Cancer. *Curr. Opin. Oncol.* **2022**, *34* (1), 54–65. <https://doi.org/10.1097/CCO.0000000000000805>.
- (206) Rewcastle, G. W.; Palmer, B. D.; Bridges, A. J.; Showalter, H. D. H.; Sun, L.; Nelson, J.; McMichael, A.; Kraker, A. J.; Fry, D. W.; Denny, W. A. Tyrosine Kinase Inhibitors. 9. Synthesis and Evaluation of Fused Tricyclic Quinazoline Analogues as ATP Site Inhibitors of the Tyrosine Kinase Activity of the Epidermal Growth Factor Receptor. *J. Med. Chem.* **1996**, *39* (4), 918–928. <https://doi.org/10.1021/jm950692f>.
- (207) Wu, J.; Chen, W.; Xia, G.; Zhang, J.; Shao, J.; Tan, B.; Zhang, C.; Yu, W.; Weng, Q.; Liu, H.; Hu, M.; Deng, H.; Hao, Y.; Shen, J.; Yu, Y. Design, Synthesis, and Biological Evaluation of Novel Conformationally Constrained Inhibitors Targeting EGFR. *ACS Med. Chem. Lett.* **2013**, *4* (10),

974–978. <https://doi.org/10.1021/ml4002437>.

- (208) Smaill, J. B.; Gonzales, A. J.; Spicer, J. A.; Lee, H.; Reed, J. E.; Sexton, K.; Althaus, I. W.; Zhu, T.; Black, S. L.; Blaser, A.; Denny, W. A.; Ellis, P. A.; Fakhoury, S.; Harvey, P. J.; Hook, K.; McCarthy, F. O. J.; Palmer, B. D.; Rivault, F.; Schlosser, K.; Ellis, T.; Thompson, A. M.; Trachet, E.; Winters, R. T.; Tecele, H.; Bridges, A. Tyrosine Kinase Inhibitors. 20. Optimization of Substituted Quinazoline and Pyrido[3,4-d]Pyrimidine Derivatives as Orally Active, Irreversible Inhibitors of the Epidermal Growth Factor Receptor Family. *J. Med. Chem.* **2016**, *59* (17), 8103–8124. <https://doi.org/10.1021/acs.jmedchem.6b00883>.
- (209) Zhang, L.; Yang, Y.; Zhou, H.; Zheng, Q.; Li, Y.; Zheng, S.; Zhao, S.; Chen, D.; Fan, C. Structure-Activity Study of Quinazoline Derivatives Leading to the Discovery of Potent EGFR-T790M Inhibitors. *Eur. J. Med. Chem.* **2015**, *102*, 445–463. <https://doi.org/10.1016/j.ejmech.2015.08.026>.
- (210) Smaill, J. B.; Rewcastle, G. W.; Loo, J. A.; Greis, K. D.; Chan, O. H.; Reyner, E. L.; Lipka, E.; Showalter, H. D. H.; Vincent, P. W.; Elliott, W. L.; Denny, W. A. Tyrosine Kinase Inhibitors. 17. Irreversible Inhibitors of the Epidermal Growth Factor Receptor: 4-(Phenylamino)Quinazoline- and 4-(Phenylamino)Pyrido[3,2-d]Pyrimidine-6-Acrylamides Bearing Additional Solubilizing Functions. *J. Med. Chem.* **2000**, *43* (7), 1380–1397. <https://doi.org/10.1021/jm990482t>.
- (211) Tsou, H. R.; Mamuya, N.; Johnson, B. D.; Reich, M. F.; Gruber, B. C.; Ye, F.; Nilakantan, R.; Shen, R.; Discifani, C.; DeBlanc, R.; Davis, R.; Koehn, F. E.; Greenberger, L. M.; Wang, Y. F.; Wissner, A. 6-Substituted-4-(3-Bromophenylamino)Quinazolines as Putative Irreversible Inhibitors of the Epidermal Growth Factor Receptor (EGFR) and Human Epidermal Growth Factor Receptor (HER-2) Tyrosine Kinases with Enhanced Antitumor Activity. *J. Med. Chem.* **2001**, *44* (17), 2719–2734. <https://doi.org/10.1021/jm0005555>.
- (212) Smaill, J. B.; Palmer, B. D.; Rewcastle, G. W.; Denny, W. A.; McNamara, D. J.; Dobrusin, E. M.; Bridges, A. J.; Zhou, H.; Showalter, H. D. H.; Winters, R. T.; Leopold, W. R.; Fry, D. W.; Nelson, J. M.; Slintak, V.; Elliot, W. L.; Roberts, B. J.; Vincent, P. W.; Patmore, S. J. Tyrosine Kinase Inhibitors. 15. 4-(Phenylamino)Quinazoline and 4-(Phenylamino)Pyrido[d]Pyrimidine Acrylamides as Irreversible Inhibitors of the ATP Binding Site of the Epidermal Growth Factor Receptor. *J. Med. Chem.* **1999**, *42* (10), 1803–1815. <https://doi.org/10.1021/jm9806603>.
- (213) Carmi, C.; Cavazzoni, A.; Vezzosi, S.; Bordini, F.; Vacondio, F.; Silva, C.; Rivara, S.; Lodola, A.; Alfieri, R. R.; La Monica, S.; Galetti, M.; Ardizzone, A.; Petronini, P. G.; Mor, M. Novel Irreversible Epidermal Growth Factor Receptor Inhibitors by Chemical Modulation of the Cysteine-Trap Portion. *J. Med. Chem.* **2010**, *53* (5), 2038–2050. <https://doi.org/10.1021/jm901558p>.
- (214) Engelman, J. A.; Zejnullahu, K.; Gale, C. M.; Lifshits, E.; Gonzales, A. J.; Shimamura, T.; Zhao, F.; Vincent, P. W.; Naumov, G. N.; Bradner, J. E.; Althaus, I. W.; Gandhi, L.; Shapiro, G. I.; Nelson, J. M.; Heymach, J. V.; Meyerson, M.; Wong, K. K.; Jänne, P. A. PF00299804, an Irreversible Pan-ERBB Inhibitor, Is Effective in Lung Cancer Models with EGFR and ERBB2 Mutations That Are Resistant to Gefitinib. *Cancer Res.* **2007**, *67* (24), 11924–11932. <https://doi.org/10.1158/0008-5472.CAN-07-1885>.
- (215) U.S. Food and Drug Administration (FDA). FDA approves dacomitinib for metastatic non-small cell lung cancer <https://www.fda.gov/drugs/drug-approvals-and-databases/fda-approves->

- dacomitinib-metastatic-non-small-cell-lung-cancer-0 (accessed Aug 19, 2022).
- (216) Mok, T. S.; Cheng, Y.; Zhou, X.; Lee, K. H.; Nakagawa, K.; Niho, S.; Lee, M.; Linke, R.; Rosell, R.; Corral, J.; Migliorino, M. R.; Pluzanski, A.; Sbar, E. I.; Wang, T.; White, J. L.; Wu, Y.-L. Improvement in Overall Survival in a Randomized Study That Compared Dacomitinib With Gefitinib in Patients With Advanced Non-Small-Cell Lung Cancer and EGFR -Activating Mutations. *J. Clin. Oncol.* **2018**, *36* (22), 2244–2250. <https://doi.org/10.1200/JCO.2018.78.7994>.
- (217) Brzezniak, C.; Carter, C. A.; Giaccone, G. Dacomitinib, a New Therapy for the Treatment of Non-Small Cell Lung Cancer. *Expert Opin. Pharmacother.* **2013**, *14* (2), 247–253. <https://doi.org/10.1517/14656566.2013.758714>.
- (218) Gajiwala, K. S.; Feng, J.; Ferre, R.; Ryan, K.; Brodsky, O.; Weinrich, S.; Kath, J. C.; Stewart, A. Insights into the Aberrant Activity of Mutant EGFR Kinase Domain and Drug Recognition. *Structure* **2013**, *21* (2), 209–219. <https://doi.org/10.1016/j.str.2012.11.014>.
- (219) Kobayashi, Y.; Fujino, T.; Nishino, M.; Koga, T.; Chiba, M.; Sesumi, Y.; Ohara, S.; Shimoji, M.; Tomizawa, K.; Takemoto, T.; Mitsudomi, T. EGFR T790M and C797S Mutations as Mechanisms of Acquired Resistance to Dacomitinib. *J. Thorac. Oncol.* **2018**, *13* (5), 727–731. <https://doi.org/10.1016/j.jtho.2018.01.009>.
- (220) Editor, M. J. W. *Cancer II (Topics in Medicinal Chemistry - Volume 28)*; Waring, M. J., Ed.; Topics in Medicinal Chemistry; Springer International Publishing: Cham, 2018; Vol. 28. <https://doi.org/10.1007/978-3-319-75926-5>.
- (221) Oncology, P. Dacomitinib (PF-00299804) - Fact Sheet. 2012.
- (222) Fakhoury, S. A.; Lee, H. T.; Reed, J. E.; Schlosser, K. M.; Sexton, K. E.; Tecle, H.; Winters, R. T. 4-Phenylamino-Quinazolin-6-Yl Amides. US 2005/0250761 A1, 2005.
- (223) Huan, L. C.; Tran, P. T.; Phuong, C. V.; Duc, P. H.; Anh, D. T.; Hai, P. T.; Huong, L. T. T.; Thuan, N. T.; Lee, H. J.; Park, E. J.; Kang, J. S.; Linh, N. P.; Hieu, T. T.; Oanh, D. T. K.; Han, S. B.; Nam, N. H. Novel 3,4-Dihydro-4-Oxoquinazoline-Based Acetohydrazides: Design, Synthesis and Evaluation of Antitumor Cytotoxicity and Caspase Activation Activity. *Bioorg. Chem.* **2019**, *92* (April), 103202. <https://doi.org/10.1016/j.bioorg.2019.103202>.
- (224) Hansen, B. K.; Loveridge, C. J.; Thyssen, S.; Wørmer, G. J.; Nielsen, A. D.; Palmfeldt, J.; Johannsen, M.; Poulsen, T. B. STEFs: Activated Vinylogous Protein-Reactive Electrophiles. *Angew. Chemie - Int. Ed.* **2019**, *58* (11), 3533–3537. <https://doi.org/10.1002/anie.201814073>.
- (225) Zhao, L.; Fan, T.; Shi, Z.; Ding, C.; Zhang, C.; Yuan, Z.; Sun, Q.; Tan, C.; Chu, B.; Jiang, Y. Design, Synthesis and Evaluation of Novel ErbB/HDAC Multitargeted Inhibitors with Selectivity in EGFR T790M Mutant Cell Lines. *Eur. J. Med. Chem.* **2021**, *213*, 113173. <https://doi.org/10.1016/j.ejmech.2021.113173>.
- (226) Vasu, D.; Yorimitsu, H.; Osuka, A. Palladium-Assisted “Aromatic Metamorphosis” of Dibenzothiophenes into Triphenylenes. *Angew. Chemie Int. Ed.* **2015**, *54* (24), 7162–7166. <https://doi.org/10.1002/anie.201501992>.
- (227) Shao, J.; Chen, E.; Shu, K.; Chen, W.; Zhang, G.; Yu, Y. 6-Oxooxazolidine–Quinazolines as Noncovalent Inhibitors with the Potential to Target Mutant Forms of EGFR. *Bioorg. Med. Chem.* **2016**, *24* (16), 3359–3370. <https://doi.org/10.1016/j.bmc.2016.04.046>.

- (228) Sheehan, J. C.; Hess, G. P. A New Method of Forming Peptide Bonds. *J. Am. Chem. Soc.* **1955**, *77* (4), 1067–1068. <https://doi.org/10.1021/ja01609a099>.
- (229) Yang, Z.; Yang, N.; Ou, Q.; Xiang, Y.; Jiang, T.; Wu, X.; Bao, H.; Tong, X.; Wang, X.; Shao, Y. W.; Liu, Y.; Wang, Y.; Zhou, C. Investigating Novel Resistance Mechanisms to Third-Generation EGFR Tyrosine Kinase Inhibitor Osimertinib in Non-Small Cell Lung Cancer Patients. *Clin. Cancer Res.* **2018**, *24* (13), 3097–3107. <https://doi.org/10.1158/1078-0432.CCR-17-2310>.
- (230) Ou, S.-H. I.; Cui, J.; Schrock, A. B.; Goldberg, M. E.; Zhu, V. W.; Albacker, L.; Stephens, P. J.; Miller, V. A.; Ali, S. M. Emergence of Novel and Dominant Acquired EGFR Solvent-Front Mutations at Gly796 (G796S/R) Together with C797S/G and L792F/H Mutations in One EGFR (L858R/T790M) NSCLC Patient Who Progressed on Osimertinib. *Lung Cancer* **2017**, *108*, 228–231. <https://doi.org/10.1016/j.lungcan.2017.04.003>.
- (231) Jiang, J.; Greulich, H.; Jänne, P. A.; Sellers, W. R.; Meyerson, M.; Griffin, J. D. Epidermal Growth Factor–Independent Transformation of Ba/F3 Cells with Cancer-Derived Epidermal Growth Factor Receptor Mutants Induces Gefitinib-Sensitive Cell Cycle Progression. *Cancer Res.* **2005**, *65* (19), 8968–8974. <https://doi.org/10.1158/0008-5472.CAN-05-1829>.
- (232) Reaction Biology Europe GmbH <https://www.reactionbiology.com/> (accessed Jul 15, 2022).
- (233) Cherinka, B.; Andrews, B. H.; Sánchez-Gallego, J.; Brownstein, J.; Argudo-Fernández, M.; Blanton, M.; Bundy, K.; Jones, A.; Masters, K.; Law, D. R.; Rowlands, K.; Weijmans, A.-M.; Westfall, K.; Yan, R. Marvin: A Tool Kit for Streamlined Access and Visualization of the SDSS-IV MaNGA Data Set. *Astron. J.* **2019**, *158* (2), 74. <https://doi.org/10.3847/1538-3881/ab2634>.
- (234) ChEMBL Database <https://www.ebi.ac.uk/chembl/> (accessed Jul 20, 2021).
- (235) Tobergte, D. R.; Curtis, S. MOE Molecular Operating Environment. *Journal of Chemical Information and Modeling*. Montreal, QC, Canada, H3A 2R7 2013, pp 1689–1699.
- (236) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (237) Charrad, M.; Ghazzali, N.; Boiteau, V.; Niknafs, A. Nbclust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *J. Stat. Softw.* **2014**, *61* (6), 1–36. <https://doi.org/10.18637/jss.v061.i06>.
- (238) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; Vijaykumar, A.; Bardelli, A. Pietro; Rothberg, A.; Hilboll, A.; Kloeckner, A.; Scopatz, A.; Lee, A.; Rokem, A.; Woods, C. N.; Fulton, C.; Masson, C.; Häggström, C.; Fitzgerald, C.; Nicholson, D. A.; Hagen, D. R.; Pasechnik, D. V.; Olivetti, E.; Martin, E.; Wieser, E.; Silva, F.; Lenders, F.; Wilhelm, F.; Young, G.; Price, G. A.; Ingold, G.-L.; Allen, G. E.; Lee, G. R.; Audren, H.; Probst, I.; Dietrich, J. P.; Silterra, J.; Webber, J. T.; Slavič, J.; Nothman, J.; Buchner, J.; Kulick, J.; Schönberger, J. L.; de Miranda Cardoso, J. V.; Reimer, J.; Harrington, J.; Rodríguez, J. L. C.; Nunez-Iglesias, J.; Kuczynski, J.; Tritz, K.; Thoma, M.; Newville, M.; Kümmerer, M.; Bolingbroke, M.; Tartre, M.;

- Pak, M.; Smith, N. J.; Nowaczyk, N.; Shebanov, N.; Pavlyk, O.; Brodtkorb, P. A.; Lee, P.; McGibbon, R. T.; Feldbauer, R.; Lewis, S.; Tygier, S.; Sievert, S.; Vigna, S.; Peterson, S.; More, S.; Pudlik, T.; Oshima, T.; Pingel, T. J.; Robitaille, T. P.; Spura, T.; Jones, T. R.; Cera, T.; Leslie, T.; Zito, T.; Krauss, T.; Upadhyay, U.; Halchenko, Y. O.; Vázquez-Baeza, Y. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17* (3), 261–272. <https://doi.org/10.1038/s41592-019-0686-2>.
- (239) Ramón, P.; Rosalia. Estudios Sobre La Metodología Computacional Para El Diseño de Bibliotecas Combinatorias : Desarrollo Del Programa Pralins y Aplicaciones En El Campo Del HIV. **2003**.
- (240) Kaufman, L.; Rousseeuw, P. K. Partitioning Around Medoids (Program PAM); Kaufman, L., Rousseeuw, P. J., Eds.; Wiley Series in Probability and Statistics; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1990; pp 68–125. <https://doi.org/10.1002/9780470316801.ch2>.
- (241) Novikov, A. PyClustering: Data Mining Library. *J. Open Source Softw.* **2019**, *4* (36), 1230. <https://doi.org/10.21105/joss.01230>.
- (242) von Luxburg, U. A Tutorial on Spectral Clustering. *Stat. Comput.* **2007**, *17* (4), 395–416. <https://doi.org/10.1007/s11222-007-9033-z>.
- (243) Merris, R. Laplacian Matrices of Graphs: A Survey. In *Linear algebra and its applications*; 1994; pp 143–176.
- (244) Navas-Palencia, G. Optimal Binning: Mathematical Programming Formulation. **2020**.
- (245) Lewis, R. A.; Mason, J. S.; McLay, I. M. Similarity Measures for Rational Set Selection and Analysis of Combinatorial Libraries: The Diverse Property-Derived (DPD) Approach. *J. Chem. Inf. Comput. Sci.* **1997**, *37* (3), 599–614. <https://doi.org/10.1021/ci960471y>.
- (246) Kim, S.; Thiessen, P. A.; Cheng, T.; Yu, B.; Bolton, E. E. An Update on PUG-REST: RESTful Interface for Programmatic Access to PubChem. *Nucleic Acids Res.* **2018**, *46*(W1), W563-570. <https://doi.org/doi:10.1093/nar/gky294>.
- (247) Wattenberg, M.; Viégas, F.; Johnson, I. How to Use T-SNE Effectively. *Distill* **2016**, *1* (10). <https://doi.org/10.23915/distill.00002>.
- (248) Karlov, D. S.; Sosnin, S.; Tetko, I. V.; Fedorov, M. V. Chemical Space Exploration Guided by Deep Neural Networks. *RSC Adv.* **2019**, *9* (9), 5151–5157. <https://doi.org/10.1039/c8ra10182e>.
- (249) Pawar, G. G.; Brahmanandan, A.; Kapur, M. Palladium(II)-Catalyzed, Heteroatom-Directed, Regioselective C–H Nitration of Anilines Using Pyrimidine as a Removable Directing Group. *Org. Lett.* **2016**, *18* (3), 448–451. <https://doi.org/10.1021/acs.orglett.5b03493>.
- (250) ENAMINE Ltd www.enameine.net (accessed Aug 2, 2022).
- (251) McChesney, J. D.; Sarangan, S. Synthesis of Site Specifically Deuterated Primaquines II. N-Alkyl Deuterated Primaquines. *J. Label. Compd. Radiopharm.* **1984**, *21* (4), 293–298. <https://doi.org/10.1002/jlcr.2580210402>.
- (252) Igel, P.; Geyer, R.; Strasser, A.; Dove, S.; Seifert, R.; Buschauer, A. Synthesis and Structure–Activity Relationships of Cyanoguanidine-Type and Structurally Related Histamine H₄ Receptor Agonists. *J. Med. Chem.* **2009**, *52* (20), 6297–6313.

<https://doi.org/10.1021/jm900526h>.

- (253) Zhang, W.; Sommers, C. L.; Burshtyn, D. N.; Stebbins, C. C.; DeJarnette, J. B.; Tribble, R. P.; Grinberg, A.; Tsay, H. C.; Jacobs, H. M.; Kessler, C. M.; Long, E. O.; Love, P. E.; Samelson, L. E. Essential Role of LAT in T Cell Development. *Immunity* **1999**, *10* (3), 323–332. [https://doi.org/10.1016/S1074-7613\(00\)80032-1](https://doi.org/10.1016/S1074-7613(00)80032-1).
- (254) Iversen, P. W.; Eastwood, B. J.; Sittampalam, G. S.; Cox, K. L. A Comparison of Assay Performance Measures in Screening Assays: Signal Window, Z' Factor, and Assay Variability Ratio. *SLAS Discov.* **2006**, *11* (3), 247–252. <https://doi.org/10.1177/1087057105285610>.

The following annexes can be found in [PhDThesis_LeticiaManénFreixa](https://tinyurl.com/PhDThesisLMF) (<https://tinyurl.com/PhDThesisLMF>) :

- **Annex I:** Database of the latest approved drugs by the FDA in the period from 2008 to 2020.
- **Annex II:** List of 206 1D and 2D descriptors calculated using MOE2020.09
- **Annex III:** Fragments used to build the different combinatorial libraries of analogs.
- **Annex IV:** OV Binning partitioning representation of Tafenoquine's case study dataset for 2, 4, 8, 16, 32 and 64 bins.
- **Annex V:** Folder containing the spectroscopic data concerning the library of Tafenoquine analogs (NMR, IR, HRMS).
- **Annex VI:** Folder containing the spectroscopic data concerning the library of Dacomitinib analogs (NMR, IR, HRMS).
- **Annex VII:** Folder containing the databases of the combinatorial libraries.
- **Annex VIII:** Folder containing the clustered databases of the combinatorial libraries
- **Annex IX:** Folder containing PyLINS.py script and the exemplifying Jupyter Notebook file PyLINS_example.ipynb.
- **Annex X:** Deconstructing Markush: Improving the R&D Efficiency Using Library Selection in Early Drug Discovery (Published article with DOI: 10.3390/ph15091159)
- **Annex XI:** Contributions to scientific meetings.