

الجمهورية الجزائرية الديمقراطية الشعبية

وزارة التعليم العالي والبحث العلمي

جامعة 20 اوت 1955 - سكيكدة

**REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE**

**MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE  
SCIENTIFIQUE**

**UNIVERSITE 20 AOUT 1955- SKIKDA**



**Faculté des Sciences**

**Département des Sciences de la Nature et de la Vie**

**Mémoire Présenté en Vue de l'Obtention du Diplôme de Master**

**Filière : Science Biologique**

**Option : Microbiologie Appliquée**

**Intitulé**

Etude théorique de l'inhibition de l'enzyme BTK (Bruton Tyrosine Kinase) par la méthode de conception de médicament à base de ligand

**Présenté Par : Bouleghlem Rokaya**

**Moumene Fouzia**

**Nettour Kawther**

**Nouiouat Marwa**

**Membre de Jury:**

Mr. Boulaknafet fouzi (MCA) Président Université du 20 Août 1955 – Skikda

Mr. Khairedine Kraim (MCB) Promoteur Enset – Skikda

Mme. Laib Imen (MCA) Examinatrice Université du 20 Août 1955 – Skikda

**Année universitaire 2021/2022**

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



## REMERCIEMENTS

Avant tout, nous remercions **DIEU** le tout puissant de nous avoir accordé la force et le courage pour réaliser ce modeste travail, atteindre notre but et réaliser ainsi un rêve.

Nos vifs remerciements et notre profonde gratitude s'adressent à notre promotrice **DOCTEURE KHAIRELINE KRAIM**, pour avoir accepté de diriger ce travail, pour son dévouement, ses précieux conseils, ses encouragements, sa patience, sa disponibilité et sa gentillesse.

Merci aussi à la présidente du jury de soutenance **BOULEKNAFET FOUZI** et à l'examinatrice **LAIB IMEN**, tous nos remerciements pour avoir accepté d'examiner et d'évaluer ce travail.

Nous remercions enfin tous ceux qui nous ont rendu service et qui ont contribué de près ou de loin pour accomplir ce travail.

## اهراء

بسم الله الوهاب الذي يهب بغير حساب، بسم الله على تمام النعم و  
جميل الكرم ان يطيب شكر البشر قبل شكر خالق البشر، فالحمد لله و  
الشكر له على فضله و منه أن بلغني هذه اللحظة، كان توفيق ربي  
يغمرنى طول هذه السنين، لم يخيبني يوما دعوته فيه، شكرا لك يارب.  
بعد شكر ربي أشكر نور عيناى أمي و أبي الذين تعبوا و شهرا ولم  
يشتكيا يوما و كان دعائهما يحيطني و يضمني أينما سارت خطواتي، لم  
يقصرا يوما في دعمي ماديا و معنويا، كان النصيح و التوجيه منجهما،  
أسأل الله العلي العظيم أن يحفظهما لي و يديم عليهما صحتهما و  
عافيتهما. بعد ذلك أود شكر زوجي الذي كان سندا لي طول هذه الفترة،  
و لا أنسى شكر الأستاذة إيمان بيشكر الأستاذة الرائعة التي كان  
الداعمة و المشجعة و المساندة في أصعب أوقاتي، و أشكر أهلي و  
أحبتى و صديقاتي، و أخيرا أشكر

نفسى على التعب و الجهد و السهر و التضحيات، و أطلب منها

الاستمرار للأمام للوصول إلى أعلى المراتب.

رقية

## اهداء

لحمد لله الذي وفقني ولم اكن لأصل لولا فضل الله علي و الحمد لله  
اهدي ثمرة جهدي وتعبي الى كل من تعب من اجل ان احقق هذا النجاح لكل  
من دعمني ورافقتي طول هذه الفترة وكان مصدر شجاعة ودعم لي الى كل  
من ساعدني طيلة مشواري الدراسي وكل الفضل لما انا عليه الان الى  
الوالدين الكريمين امي و ابي حفظكما الله ورعاكما و ادام عليكم الصحة و  
العافية لولاكم ماكنت لأصل هنا  
لكل العائلة الكريمة التي ساندتني ولا تزال من اخوتي سلمى سارة محمد الى  
الصديقة امال التي يسرت لي الصعاب  
الى الصديقة والأخت هذى و التي كانت ترسم الامل في طريقي ولم ثانية الى  
رفيقة دربي غزلان التي رافقتني طيلة مسيرتي الجامعية مصدر طاقتي  
الإيجابية ما كان ليحدث هذا لولا تشجيعها المستمر لن انسى وقوفها بجانبني  
الى صديقات التخرج كوثر مروى رقية و الى كفة أصدقائي و احبتي و كافة  
زميلاتي وزملاء دفعة ماستر 2021-2022 تخصص ميكروبيولوجي تطبيقي

فوزية

## اقراء

قال تعالى: لمن شكرتم ازددكم

الحمد لله الذي وهب لنا نعمة العقل و العلم

الحمد لله الذي يسر لنا امورنا و عززنا بالفهم

الحمد لله الذي وفقنا و سهل لنا التقدم للامام

الحمد لله و الصلاة على محمد اعظم النعم

قال صلى الله عليه و سلم: من لا يشكر الناس لا يشكر الله

تعجز كل كلمات الشكر امام عظمة الوالدين الذين دفعوا سنين عمرهم ليقظو ثمار نجاحنا...الى  
صاحبة القلب الكبير و فرحتي و سندي و عوني بعد الله سبحانه و تعالى و التي حملت طموجي و

اميّ شغفي على اكتافها نعمتي الحلوة

الى الذي احمل اسمه بكل شكر و اعتزاز و من بدل جهده ووقته لكي اصل الى ماوصلت اليه

ابيّ

لكل العائلة الكريمة التي ساندتني و لا تزال كل من زوجي و من اخوة و اخوات و الى رفيقات

المشوار اللاتي قاسمتني لحظاته رعاهم الله ووفقه

و في الاخير...الى نفسي التي تعبت...و سهرت و تحملت...رفع قبعة الفخر و الاعتزاز...و الحب

و الشكر

مروى

## اهراء

الحمد لله الذي ماتم جهد ولا ختم سعى الا بفضلله و ماتخطى العبد من صعوبات الا بتوفيقه و معونته..  
في آخر خطوة السُّلم ، وفي قِمة الحُلم الكبير ، وفي اللحظة الأكثر فُخْرًا ، عندما كان التعب والجهد  
يغسلنا بماء النَّجَاحِ لِيَبْقَى الواقع أيضًا أجمل .

ها أنا اليوم أهدي تخرجي إلى الرضا والنور وطهر الحياة والطريق المختصر للجنة ، صاحبة القلب  
الكبير ، نعمتي الحلوة ، فرحي وسندي وعَونِي بعد الله سبحانه وتعالى ، إلى من حملت طموحي  
وشغفي على أكتافها ، وكانت خطايا بها خفيفة " حبيبتي أمي " ونزعت الأشواك من طريقي بدعائها ،  
من احترقت كي تنير دربي أصل إلى ما وصلت إليه " أبي الغالي " إلى من أحمل اسمه بكل فخر  
وإعتزاز ومن بذل جهده ووقته لكي أصل إلى ماوصلتُ إليه.

إلى أخواتي: أميمة و ريم وإبنتها رُوح قلب خالنتها جوري أسيل وإخوتي : محمود و عبد الغاني، أنتم  
أجنتي والضوء الذي يوقظني للحياة وإكليل الزهور حين يضح كل ما حولي بالقتامة .

إلى صديقات عمري وأخوات قلبي إكرام ، وفاء ، بتول،رقية ، فوزية مروة التي إذا رأوا جدار روحي  
على وشك أن ينقضَ أقمنه ولم تفكرن أن تتخذن عليه أجرا .

وفي الأخير إلى نفسي التي أحب أن أوجه الشكر الكبير والامتنان العظيم ، نعم نفسي التي تعبت ،

سهرت ، تحملت وليس أعز من نفسي ، أحب

أن أرفع إليها قبعة الفخر والاعتزاز .

كوثر



# LE SOMMAIRE

Résumé

Abstract

ملخص

Remerciement

Dédicace

Sommaire

Liste des figures

Liste des abréviations

Liste des annexes

**INTRODUCTION GÉNÉRALE ..... 01**

**CHAPITRE I : Le cancer l'enzyme tyrosine kinase de bruton**

**I. Le cancer..... 02**

**I.1.Définition .....02**

**I.2. La classification du cancer ..... 02**

**I.3. Les types de cancer .....02**

**I.3.1. Le cancer du sein .....03**

**I.3.2. Le cancer de la prostate .....03**

**I.3.3. Le cancer du poumon .....03**

**I.3.4. Le cancer de l'estomac .....03**

**I.3.5. Le cancer de l'ovaire .....03**

**I.4. Les causes du cancer.....03**

**I.5. Le traitement.....03**

<b>I.5.1 la chirurgie .....</b>	<b>04</b>
<b>I.5.2 la chimiothérapie .....</b>	<b>04</b>
<b>I.5.3 La radiothérapie .....</b>	<b>04</b>
<b>I.5.4. L'immunothérapie.....</b>	<b>04</b>
<b>I.5.5 L'hormonothérapie.....</b>	<b>04</b>
<b>II. L'ENZYME BTK (BRUTONTYROSINE KINASE).....</b>	<b>04</b>
<b>II.1. Les enzymes comme cibles thérapeutiques .....</b>	<b>04</b>
<b>II.2. L'activité enzymatique.....</b>	<b>05</b>
<b>II.3. L'inhibition enzymatique .....</b>	<b>06</b>
<b>II.3.1. Inhibition réversibles .....</b>	<b>06</b>
<b>II.3.1.1. Inhibition compétitive .....</b>	<b>06</b>
<b>II.3.1.2. Inhibition non compétitive .....</b>	<b>06</b>
<b>II.3.1.3. Inhibition in compétitive .....</b>	<b>06</b>
<b>II.3.2. Inhibiteur irréversibles .....</b>	<b>07</b>
<b>II.4. Enzyme BTK (Burton tyrosine kinase) .....</b>	<b>07</b>
<b>II.4.1. Une tyrosine kinase de Bruton .....</b>	<b>07</b>
<b>II.4.2.Le gène de BTK .....</b>	<b>08</b>
<b>II.4.3.Structure de BTK .....</b>	<b>08</b>
<b>II.4.4. Le rôle de BTK dans les cellules B.....</b>	<b>09</b>
<b>III. INHIBITEURS DE BTK.....</b>	<b>10</b>
<b>III.1. Les inhibiteurs de BTK approuvée .....</b>	<b>10</b>
<b>III.1.1. L'ibrutinib : l'inhibiteur de BTK la première génération.....</b>	<b>10</b>
<b>III.1.2. L'inhibiteur de BTK de deuxième génération .....</b>	<b>11</b>
<b>III.2. Inhibiteur récemment approuvé .....</b>	<b>12</b>

<b>III.3.Inhibiteur de BTK en cours d’investigation clinique .....</b>	<b>12</b>
<b>III.3.1.Les inhibiteurs de BTK irréversibles.....</b>	<b>12</b>
<b>III.3.2.Les inhibiteurs de BTK réversibles.....</b>	<b>12</b>
<b>CONCLUSION.....</b>	<b>13</b>

**CHAPITRE II :**

<b>I. Relation Quantitative Structure-Activité (QSAR) .....</b>	<b>14</b>
<b>I.1. Définition .....</b>	<b>14</b>
<b>I.2.L’importance du QSAR en biologie.....</b>	<b>14</b>
<b>I.3.Principe de QSAR.....</b>	<b>14</b>
<b>I.4.L’application du modèle QSAR .....</b>	<b>15</b>
<b>I.5.Méthodologie générale d’une étude QSAR .....</b>	<b>16</b>
<b>I.5.1. Sélection de la chimio thèque.....</b>	<b>17</b>
<b>I.5.2. Génération de descripteurs moléculaires .....</b>	<b>17</b>
<b>I.5.3. Algorithmes de sélection des descripteurs moléculaires .....</b>	<b>17</b>
<b>I.5.3.1 Sélection objective.....</b>	<b>17</b>
<b>I.5.3.2 Sélection subjective .....</b>	<b>18</b>
<b>I.5.3.2.1. Introduction progressive .....</b>	<b>18</b>
<b>I.5.3.2.2. Élimination progressive.....</b>	<b>18</b>
<b>I.5.3.2.3. Sélection pas à pas (Forward Stepwise).....</b>	<b>18</b>
<b>I.5.4. Algorithmes d’apprentissages.....</b>	<b>18</b>
<b>I.5.4.1. la régression linéaire multiple (MLR).....</b>	<b>18</b>
<b>I.5.4.2. la régression aux moindres carrés partiels (PLS).....</b>	<b>19</b>
<b>I.5.4.3. réseaux neuronal artificiel (ANL) .....</b>	<b>19</b>
<b>I.5.5. Interprétation et validation d’un modèle QSAR .....</b>	<b>20</b>

<b>I.5.5.1. validations internes</b> .....	<b>20</b>
<b>I.5.5.2. validations externes</b> .....	<b>20</b>
<b>I.5.6. Domaine d'applicabilité</b> .....	<b>21</b>
<b>CONCLUSION</b> .....	<b>21</b>
<b>CHAPITRE III</b>	
<b>I. Matériel et Méthode</b> .....	<b>22</b>
<b>I.1. Préparation de chimio thèque</b> .....	<b>22</b>
<b>II. Génération de descripteurs moléculaires</b> .....	<b>23</b>
<b>II.1. Répartition de la chimio thèque</b> .....	<b>23</b>
<b>II.2. Prétraitement de la matrice de descripteur</b> .....	<b>24</b>
<b>II.3. Construction des modèles QSAR</b> .....	<b>25</b>
<b>II.4. Validation externes des modèles</b> .....	<b>25</b>
<b>CHAPITRE IV</b>	
<b>I. Résultats et discussion</b> .....	<b>26</b>
<b>I.1. L'étude quantitative par l'utilisation de MLR</b> .....	<b>26</b>
<b>I.1.1. Interprétation des résultats calculés par le groupe par l'apprentissage</b> .....	<b>27</b>
<b>I.1.2. Interprétation des résultats calculés par la validation externe</b> .....	<b>28</b>
<b>I.2. L'étude quantitative par l'utilisation de PLS</b> .....	<b>28</b>
<b>I.2.1. Interprétation des résultats calculés par le groupe d'apprentissage</b> .....	<b>29</b>
<b>I.2.2. Interprétation des résultats calculés par le groupe test</b> .....	<b>29</b>
<b>I.3. Application de l'algorithme d'apprentissage Boosting de Gradient</b> .....	<b>30</b>
<b>I.3.1. XG-BOOST Régression</b> .....	<b>30</b>
<b>CONCLUSION</b> .....	<b>32</b>

## LISTE DES FIGURES

<b>Figure 01</b> : Deux cellules cancéreuses .....	1
<b>Figure 02</b> : Les causes principales de cancers.....	2
<b>Figure 03</b> : Le complexe enzyme-substrat .....	5
<b>Figure 04</b> : Les types d'inhibition réversibles.....	7
<b>Figure 05</b> : La structure de BTK .....	8
<b>Figure 06</b> : La structure d'inhibiteurs de la deuxième génération .....	10
<b>Figure 07</b> : Modèle de l'étude de QSAR.....	15
<b>Figure 08</b> : La modélisation du QSAR.....	16
<b>Figure 09</b> : Schéma représentatif d'un réseau de neurones artificiel .....	19
<b>Figure10</b> : Base de données ChEMBL.....	22
<b>Figure11</b> : Préparation des molécules par le logiciel Discovery studio (Biovia).....	23
<b>Figure12</b> : Répartition de la chimio thèque en deux groupes training et test.....	24
<b>Figure13</b> : Le logiciel Molegro Virtuel.....	24
<b>Figure 14</b> : Représente la valeur calculée de IC50 mesurée par MLR (training) $R=0,53$ ..	27
<b>Figure 15</b> : Représente la valeur calculée de IC50 mesurée par MLR (Test) $Q=0,40$ .....	27
<b>Figure 16</b> : Représente la valeur calculée de IC50 mesurée par PLS (training) $R=0,53$ ...	29
<b>Figure 17</b> : Représente la valeur calculée de IC50 mesurée par PLS (Test) $Q=0,40$ .....	29
<b>Figure18</b> : Projection de la variation des valeurs d $pIC_{50}$ expérimentales en fonction de $pIC_{50}$ calculés pour l'ensemble d'apprentissage. ....	30
<b>Figure19</b> : Projection de la variation des valeurs d $pIC_{50}$ expérimentales en fonction de $pIC_{50}$ calculés pour l'ensemble de validation externe.....	31

## Le résumé

Dans ce travail nous avons réalisé une étude in silico, dont l'objectif est de lutter contre le **cancer** à travers **l'inhibition** de la **BTK** comme **cible thérapeutique**. L'approche adoptée dans ce projet est la méthode **QSAR**, dont plusieurs modèles ont été élaborés, on se basant sur les **descripteurs moléculaires** générés par **Dragon** et un ensemble de molécules constitués de 1317 molécules ayant une valeur d'**IC50** (nM). Les modèles construits sont obtenus avec les algorithmes d'apprentissage : **MLR**, **PLS**, et **XGBOOST**. Meilleur modèle est obtenu avec une analyse qualitative par l'algorithme **XGBOOST**. Ce dernier est utilisé pour cribler une **chimio thèque** composé de 1064 produits naturels tirés de la base **ZINC**. Le modèle nous a fournie une série de 667 molécules actives contre l'enzyme **BTK**.

**Les mots clé :** Cancer, l'inhibition, BTK, Cible thérapeutique, QSAR, Descripteurs moléculaires, Dragon, IC50, MLR, PLS, XGBOOST, Chimio thèque, ZINC.

## Abstract

In this work we carried out an in silico study, whose objective is to fight against **cancer** through the **inhibition** of **BTK** as a **therapeutic target**. The approach adopted in this project is the **QSAR** method, several models of which have been developed, based on the **molecular descriptors** generated by **Dragon** and a set of molecules consisting of 1317 molecules with an **IC50** (nM) value. The built models are obtained with the learning algorithms: **MLR**, **PLS**, and **XGBOOST**. Best model is obtained with qualitative analysis by **XGBOOST** algorithm. The latter is used to screen a **chemical library** composed of 1064 natural products from the **ZINC** base. The model provided us with a series of 667 molecules active against the **BTK** enzyme.

**Keywords** : Cancer, Inhibitor, BTK, Therapeutic target, QSAR, Molecular descriptors, Dragon, IC50, MLR, PLS, XGBOOST, Chemical library, ZINC.

## ملخص

لقد أنجزنا بفضل الله في هذا العمل دراسة افتراضية و التي كان هدفها محاربة مرض السرطان من خلال تثبيط إنزيم بروتون تيروزين كيناز (BTK) و ذلك باستعماله كهدف علاجي. المنهج المتبع في هذا المشروع هو طريقة QSAR والتي تم تطوير العديد من النماذج من خلالها، من خلال التركيز على المؤشرات الجزيئية المنشأة عن طريق برمجة دراغون (Dragon) ومجموعة جزيئات مكونة من 1317 جزيئة كل جزيئة لها قيمة IC50 (nM). النماذج المشكلة تحصلنا عليها من خلال خوارزميات التعلم الآتية، MLR, PLS, et XGBOOST : أفضل نموذج تحصلنا عليه عن طريق التحليل الوصفي بخوارزمية XGBOOST، هذا الأخير استعمل لغرلة مكتبة كيميائية مكونة من 1460 منتج طبيعي مأخوذ من قاعدة زنك (ZINC). النموذج قدم لنا سلسلة مكونة من 667 جزيئة فعالة ضد إنزيم BTK.

الكلمات المفتاحية : تثبيط، بروتون تيروزين كيناز (BTK)، هدف علاجي، QSAR، المؤشرات الجزيئية، دراغون (Dragon)، IC50، PLS، XGBOOST، MLR، زنك (ZINC).

## ABRIVIATION

ATP=Adénosine-triphosphate

BCR= B cell receptor

BKC-B= protein kinas C-B

BTK=Bruton tyrosine kinase

DA= Domaine d'applicabilité

DAG=diaglycérol

ES=Enzyme-Substrat

F= Statistique de Fisher

FDA= Food and drug administration

HLA=humain leucocyte antigène

IC50=concentration inhibitrice demi maximale

LLC=Leucémie lymphoïde chronique

LMO= Leave-Many-Out

LOO= Leave-One-Out

LYN=Lck /yes novel tyrosine kinas

MCL= lymphome à cellule du manteau

NF-KB=facteur nucléaire kappa

NK= Naturel killer

PH= pleckstrine homologie domaine

PH=potentiel hydrogène

PIP PKC=

PRR=proline-rich regions

PSET= Ensemble de prédiction

QSAR=Relations Quantitatives Structure Activité

SA= Site actif

SH2 /SH3=Src homology 2/3 domain

SLL=Small lymphocyte lymphoma

SYK=Spleen tyrosine kinas

TH=Tec homology domain

TK=tyrosine kinas

TP3=inositol triphosphate

XLA=X-linked agammaglobulinemia

XGBOOST= Extreme Boosting de gradient

## Introduction

Le cancer est devenu l'un des grands dilemmes qui occupent l'esprit des chercheurs en raison de la force de sa propagation et la difficulté à le traiter, et c'est devenu la raison principale et première cause de mortalité dans le monde entier [2]. Les dernières statistiques dans le monde ont révélé le nombre énorme de personnes infectées par cette maladie chaque année. En Algérie Soixante cinq mille (65000) nouveaux cas de cancer tous types confondus ont été recensés depuis le début de l'année 2021 [1].

Les principales stratégies utilisées dans le traitement des cancers sont : la chirurgie, la radiothérapie, la chimiothérapie, et parmi d'autres traitements qui sont utilisés pour combattre cette maladie mortelle c'est l'utilisation d'enzymes comme cibles thérapeutiques, et essayer de trouver des inhibiteurs de ces enzymes afin de tuer les cellules cancéreuses et de limiter leur propagation. Parmi les enzymes les plus utilisées sont les kinases y compris l'enzyme de tyrosine kinase de bruton (BTK) [4, 5,6].

Plusieurs inhibiteurs de cette enzyme ont déjà été trouvés mais malheureusement, ils n'ont pas été utiles au degré souhaité en raison de la résistance du corps à celui-ci, ainsi que des effets secondaires indésirables qui l'accompagne [6, 8, 9].

Notre problématique est comment trouver de nouveaux inhibiteurs contre la BTK et qu'ils soient plus efficaces avec moins d'effets secondaires ?

En raison de la difficulté de mener des expériences en laboratoire du fait de longue apparition des résultats et manque de disponibilité de réactifs et leur cher coût, nous avons utilisé dans notre mémoire la méthode de relation quantitative structure-activité (QSAR) pour identifier les structures chimiques pouvant présenter de bons effets inhibiteurs sur des cibles spécifiques, nous avons exploiter ce modèle pour la découverte de nouvelles structures chimiques par le criblage virtuel des grandes chimio thèques [13, 14, 15].

Notre mémoire en générale est divisé en six parties, en commençant par l'introduction générale ensuite le premier chapitre qui discute les enzymes en générale et l'enzyme de BTK spécialement, et après le deuxième chapitre qui aborde la méthodologie du modèle quantitatif et qualitatif de QSAR, et la partie expérimentale qui est présenté dans le troisième chapitre, suivie par la partie des résultats qui présenté dans le quatrième chapitre et terminé par une conclusion.

## I. Le cancer

### I.1. Définition

Ce terme englobe un groupe de maladies qui caractérisent par le changement anormale des cellules, formant un amas appelé tumeur, C'est le nom de type de tumeurs maligne qui se multiple rapidement dans l'organisme, les cellules cancéreuses initiales peuvent se propager depuis leur localisation primitive. Le cancer reste l'une des principales causes de mortalité [1].

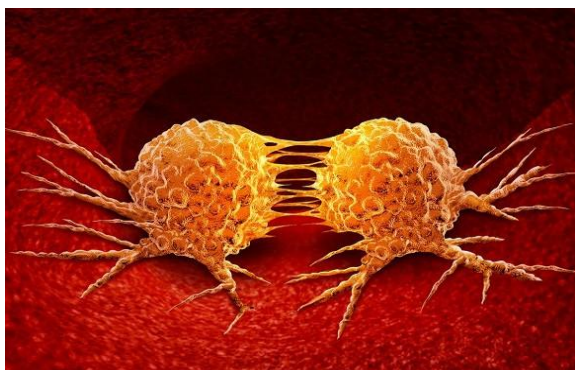


Figure 01 : Tumeurs malignes.

### I.2. La classification du cancer

Le cancer généralement classé en 3 catégories, à savoir les carcinomes, les sarcomes et les tumeurs hématopoïétiques [2].

**Le carcinome** : les cellules cancéreuses se développent dans les tissus qui recouvrent les organes.

**Le sarcome** : les cellules cancéreuses se développent dans les tissus de support que sont les os (ostéosarcome), la graisse (liposarcome) ou encore les muscles (rhabdomyosarcome). Les cellules cancéreuses se développent directement dans l'épithélium d'une glande.

**Le cancer hématopoïétique** : cette dernière typologie concerne le cancer du sang (leucémie).

### I.3. Les Types de Cancers

Il existe plusieurs types, nous citons les plus courants [3].

**I.3.1. Le cancer du sein** : plus fréquent chez la femme, il existe aussi chez l'homme résulte le dérèglement de certains cellules qui se multiplient et forment une tumeur, traiter chirurgicalement.

**I.3.2. Le cancer de la prostate** : ce cancer devient extrêmement fréquent, l'âge est un facteur de risque chez l'homme dans ce cas. Une surveillance peut être préférable à un traitement immédiat.

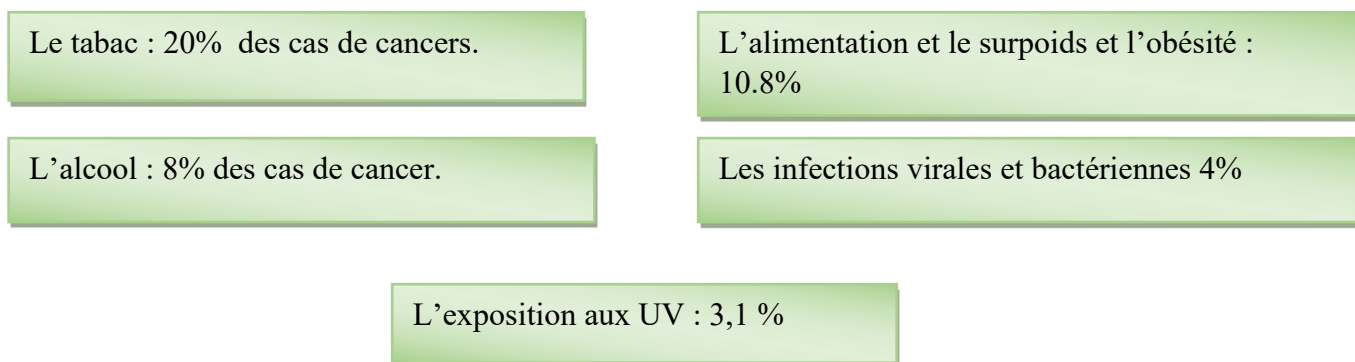
**I.3.3. Le cancer du poumon** : Il existe plusieurs types en fonction de type de cellules de branches, le tabagisme reste le facteur le plus important, les hommes restent deux fois plus touché que les femmes. Le traitement dépend au type de cancer bronchique.

**I.3.4 Le cancer de l'estomac** : trouve son origine au niveau de la muqueuse, l'alimentations parmi les facteurs de ce cancer.

**I.3.5. Le cancer d'ovaire** : Une grande variété de tumeurs peuvent se développer dans les ovaires appariant de la couche cellulaire externe.

## I.4. Les Causes du Cancer

Les causes du cancer sont multiples et difficiles à comprendre et à identifier, nous citons parmi eux : [1].



**Figure 02 : Les causes principales du cancer.**

## I.5. Le Traitement

Il agit soit localement, c'est-à-dire uniquement sur les cellules cancéreuses d'un organe atteint, soit par voie générale, c'est-à-dire sur l'ensemble des cellules cancéreuses présentes dans l'organisme. La chirurgie et la radiothérapie sont des traitements dits locaux, la chimiothérapie et l'hormonothérapie des traitements dits généraux. Les thérapies ciblées sont sélectives et s'attaquent à une cible précise dans la cellule cancéreuse [3].

**Il y a plusieurs types de traitement contre le cancer.**

### I.5.1. La chirurgie

Est la technique la plus ancienne de traitement de cancer, Elle est utilisée dans environ 80 % des cas, mais pour une meilleure efficacité peut être réalisée dans un second temps, voire parfois même refusée.

### **I.5.2. La chimiothérapie**

Est un traitement reposant sur la prise de médicaments plus ou moins associés, La chimiothérapie peut ainsi servir à réduire la taille d'une tumeur avant une opération chirurgicale, à limiter les risques de récives après l'intervention chirurgicale ou à traiter des localisations secondaires de tumeur, elle peut s'accompagner des effets indésirable.

### **I.5.3. La radiothérapie**

Seule la zone où se trouve la tumeur est exposée aux rayons afin d'éviter au maximum que des cellules saines ne soient elles aussi touchées. Cette exposition provoque une transformation des cellules qui perdent alors leur faculté à se multiplier.

### **I.5.4. L'immunothérapie**

Dans le cas du cancer, elle ne s'attaque pas directement à la tumeur, mais stimule les cellules immunitaires impliquées dans sa reconnaissance et sa destruction.

### **I.5.5. L'hormonothérapie**

Les cellules cancéreuses ont tendance à se multiplier plus vite en présence de ces hormones.

L'idée est donc de bloquer la production ou l'activité de ces hormones.

## **II. L'ENZYME BTK (BRUTON TYROSINE KINASE)**

### **II.1. Les enzymes comme cibles thérapeutiques**

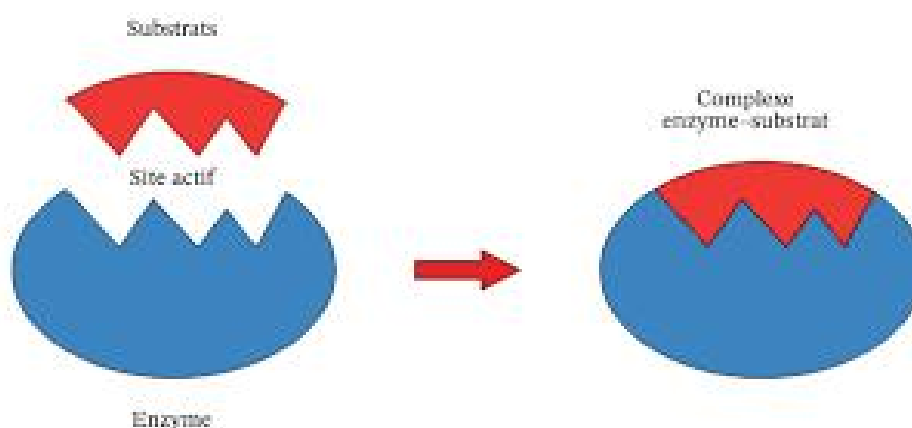
Les enzymes sont des molécules de nature protéique ou parfois des acides ribonucléiques fabriqués par l'organisme indispensables à l'activité biochimique cellulaire dont le rôle est de catalyser les réactions chimiques du vivant, sont de remarquables machines moléculaires qui déterminent le profil de certaines transformations chimique et elles assurent aussi la transformation d'une forme d'énergie en une autre [4]. Environ un quart des gènes du génome humain code pour des enzymes, ce qui témoigne de leur importance pour la vie, elles se retrouvent naturellement dans d'autre organismes vivants, tels que les végétaux, les animaux, et les microorganismes. Les enzymes qu'ils produisent par les microorganismes sont très utilisés par l'homme pour confectionner des aliments comme le pain, le fromage, la bière et le vin. Dernièrement les enzymes sont utilisés comme des cibles thérapeutiques pour plusieurs tumeurs malignes.

## II.2. L'Activité Enzymatique

Les enzymes se caractérisent par la spécificité réactionnelle donc chaque enzyme ne catalyse qu'une seule réaction ou un groupe de réaction de même type, ainsi par la spécificité de substrat par l'architecture spatiale de l'enzyme. L'activité de l'enzyme est dépendante des conditions physico-chimiques (température, pH, concentration du substrat, présence ou absence d'un inhibiteur..) régnant dans l'environnement immédiat de la réaction. Tout déséquilibre de l'un de ces conditions peut dénaturer la structure et la fonction du catalyseur. Quand la structure spatiale de l'enzyme n'est plus adaptée l'enzyme est dite dénaturée. La même façon si une mutation qui entraîne le remplacement d'un acide aminé par autre dans cette chaîne peut modifier la complémentarité enzyme-substrat et l'action de l'enzyme voire bloquer cette action. Une enzyme agit comme catalyseur en se liant au substrat et facilite sa réaction en stabilisant son état en transition vers un produit spécifique [5].

**SUBSTRAT(S) + ENZYME  $\longrightarrow$  COMPLEXE (ENZYME-SUBSTRAT).**

- **Substrat** : Une molécule entre dans la réaction pour y être transformée grâce à l'action catalytique de l'enzyme.
- **Produit** : Une molécule apparaît au cours de réaction catalysée par l'enzyme.
- **Le site actif** : (SA) est un privilégiée, qui a la forme d'une cavité, située dans la zone interne hydrophobe de la protéine, au niveau de laquelle s'exerce électivement le pouvoir catalytique de l'enzyme.



**Figure03 : Le complexe enzyme-substrat.**

### II.3. L'inhibition enzymatique

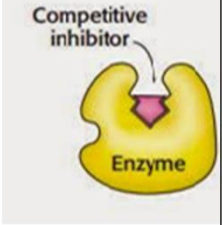
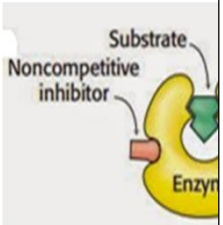
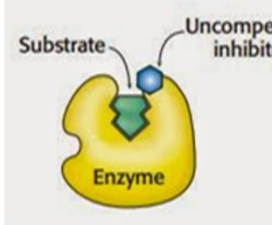
Quand la vitesse d'une réaction enzymatique diminue dans des conditions où l'enzyme n'est pas dénaturée, cela signifie que l'enzyme est inhibée. Il existe deux principaux types d'inhibition enzymatique: inhibition réversible et inhibition irréversible [6].

**II.3.1. Inhibition réversible** : l'activité enzymatique peut être retrouvée en enlevant l'inhibiteur, Pour déterminer le type d'inhibition, on mesure en l'absence d'inhibiteur, puis en présence d'un inhibiteur dont on fait varier la concentration. On distingue dans l'inhibition réversible trois catégories principales :

**II.3.1.1. Inhibition compétitive** : liaison au site actif implique blocage de l'activité enzymatique.

**II.3.1.2. Inhibition non compétitive** : Ils se lient de manière réversible ailleurs qu'au site actif de l'enzyme et n'en empêchent pas l'accès. Il s'agit donc d'inhibiteurs **allostériques**, c'est-à-dire agissant à un autre (*allo-*) site. Il existe deux types d'inhibition non compétitive : l'inhibition non compétitive pure et l'inhibition non compétitive mixte.

**II.3.1.3. Inhibition in compétitive** : appelé inhibition par blocage du complexe intermédiaire, l'enzyme et le substrat forment d'abord le complexe enzyme-substrat (complexe intermédiaire), puis l'inhibiteur se fixe à ce complexe. Il y a formation d'un complexe ternaire (ESI) inactif.

Type	Compétitive	Non compétitive	In compétitif
Les schémas	 <p>Inhibitor → Active site</p>	 <p>Inhibitor → Allosteric site</p>	 <p>Inhibitor → ES complex</p>
équation	$E + S \rightleftharpoons ES \rightarrow E + P$ $+ I \rightleftharpoons EI$	$E + S \rightleftharpoons ES \rightarrow E + P$ $+ I \rightleftharpoons EI$ $+ S \rightleftharpoons EIS$	$E + S \rightleftharpoons ES \rightarrow E + P$ $+ I \rightleftharpoons EIS$

<b>Descripti on</b>	<b>(I) se lie à libre (E) uniquement et entre en compétition avec(S) augmentant(S) surmonte l'inhibition par (I).</b>	<b>(I) lie le complexe (S) ou (ES) libre ; l'augmentatio n de (S) ne permet pas de vaincre (I) l'inhibition.</b>	<b>(I) se lie au complexe (ES) uniquement ; l'augmentation de (S) favorise l'inhibition par (I).</b>

**Figure 04 : Les types d'inhibition réversible.**

### **II.3.2. Inhibiteurs irréversibles :**

L'inhibiteur se lie de manière irréversible au site actif de l'enzyme ou à proximité, généralement par des liaisons covalentes, de sorte qu'il ne peut pas se dissocier de l'enzyme. Les inhibiteurs irréversibles se combinent avec les groupes fonctionnels de l'acide aminé dans le site actif, de manière irréversible. Les inhibiteurs irréversibles occupent ou détruisent définitivement le site actif de l'enzyme et diminuent la vitesse de réaction. L'activité enzymatique n'est pas retrouvée lors de la dialyse.

## **II.4. L'Enzyme BTK (Bruton Tyrosine Kinase)**

### **II.4.1. Une tyrosine kinase de Bruton**

La tyrosine kinase de Bruton (BTK) est une tyrosine kinase non réceptrice appartenant à la famille TEC. La BTK est exprimée dans les cellules de toutes les lignées hématopoïétiques à l'exception des cellules T et des plasmocytes, est une molécule critique en aval de la voie de signalisation du récepteur des cellules B (BCR) Et joue un rôle important dans le développement des cellules B. Etaient identifier entant que protéine défectueuse de l'agammaglobulinémie liée à (HLA) humain par vertie et ses collaborateurs en 1993. BTK joue un rôle essentiel dans le développement, la maturation et différenciation des cellules B mature en cellule plasmocytes producteurs des anticorps, et la prolifération et la survie de B mature. Cette enzyme est utilisée comme cible thérapeutique très

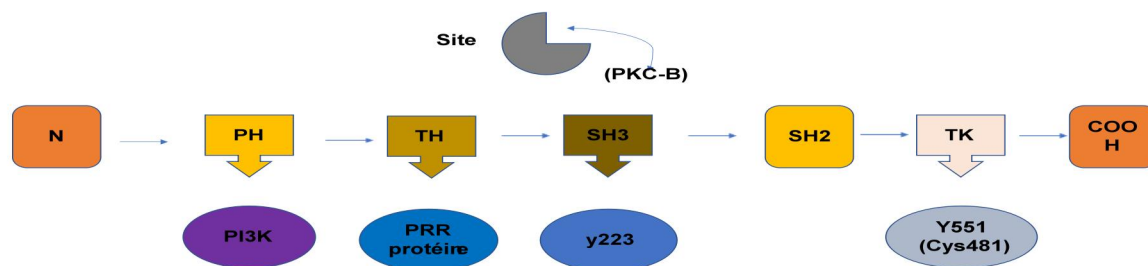
intéressante pour les tumeurs malignes à cellules B. L'importance de cette enzyme dans le développement des lymphocytes B a été mise en évidence lors de la découverte en 1952 par le pédiatre américain, le Dr Ogden Carr Bruton, d'un déficit immunitaire héréditaire, l'agammaglobulinémie liée à l'X (XLA), cette maladie orpheline se caractérise par une production faible d'immunoglobulines conduisant à une susceptibilité aux infections bactériennes, un déficit en lymphocytes B naïfs circulants par blocage de la maturation lymphoïde B, et à un non-développement des organes lymphoïdes secondaires [7,9].

#### **II.4.2. Le gène de BTK**

Il a été découvert en 1993, et plus de 800 mutations du gène BTK ont été à ce jour liées à l'XLA, la plupart ayant pour conséquence une absence de production de la protéine BTK

#### **II.4.3. Structure de BTK**

BTK est composé de 5 domaines : un domaine d'homologie pleckstrine (PH) situé dans le N-terminal et à l'extrémité possède le site pour la liaison des phosphatidylinositol 3Kinases (PI3K), un domaine d'homologie TEC (TH) constitué d'un motif BTK (BM) riche en proline (PRR), le domaine d'homologie SH3 constitué de deux domaines d'homologie SRC (SH3 suivi de SH2) et un domaine kinase C-terminal (BTK-KD). Et le domaine de kinase catalytique (TK) constitué d'un site de phosphorylation important (Y551), à l'extrémité de C-terminal compose le site de résidu qui représente le site de liaison covalente des BTK les plus étudiés. Des mutations dans tous les domaines de la BTK humaine se sont avérées conduire à XLA et des mutations faux-sens ont été trouvées dans tous les domaines à l'exception du domaine SH3. Des études structurales de la famille Src des tyrosines kinases ont révélé que ces protéines peuvent adapter deux conformations : un état auto-inhibiteur de la protéine, appelée conformation de « domaine régulateur assemblé », et une structure active, plus ouverte, où le domaine SH2 n'interagit pas avec la queue C-terminale non phosphorylée [8, 10,12].



**Figure 05 : La structure de BTK.**

#### **II.4.4. Le rôle de BTK dans les cellules B**

La Tyrosine kinase de Bruton a un rôle essentiel dans la voie de signalisation des lymphocytes B (BCR) il s'agit comme modulateur de signaux intracellulaire, lors de BTK est un composant important dans différents récepteurs de cellule B. Ce rôle est de réaliser une liaison antigène et BCR , le BTK relie le BCR par l'activation de NF - KB ( LYN ) et une nouvelle tyrosine kinase liée à ( LCK ) phosphorite le tyrosine des récepteurs CD79A et CD79B, après la première phosphorylation du BCR Le BCR phosphorylé incorporer la tyrosine kinase de la rate ( SYK ) sur la membrane, et le SYK phosphorite phosphoryle Le BTK, et le BTK phosphorylé peut phosphoryle PLCY<sub>2</sub> dans les résidu Y753 et Y 759, cette phosphorylation essentiel pour l'activité de lipase. le PLCY<sub>2</sub> activé hydrolyse pip<sub>2</sub> qui permet la génération d'inositol triphosphate ( IP<sub>3</sub> ) et la génération du dia glycérol ( DAG). Le rôle de ( TP<sub>3</sub> ) régule les niveaux de calcium intracellulaire, ainsi que le ( DAG ) permet de médie l'activation de la protéine kinase ( PK CB ) Aussi le SYK phosphorylé peut phosphoryle Pip<sub>3</sub>, et le PKC peut activé par Pi<sub>3</sub>k et BTK que se réaliser l'activation de la voie NF - KB , de cette façon , il régule l'activation , la prolifération et la différenciation des cellules B immature en cellule plasmocytes producteur d'anticorps[9,10, 12].

### **III. INHIBITEURS DE BTK**

Le BCR jouant un rôle clé dans la prolifération et la survie des cellules B, ses composants individuels pourraient être des cibles médicamenteuses potentielle dans les tumeurs malignes des cellules B. en tant qu'effecteur de l'activation de la voie de signalisation BCR BTK joue un rôle de médiateur dans la prolifération des cellules B dans lymphomagenèses. Le BTK, en tant que composant de la voie de signalisation BCR, est une cible médicamenteuse très intéressante pour les

lymphomes et les leucémies [1]. Les inhibiteurs de BTK ont attiré une attention accrue de la part des instituts pharmaceutiques et des industries depuis l'approbation du premier inhibiteur en 2013. Ces inhibiteurs peuvent interférer avec l'activité de la kinase soit en bloquant la liaison à l'ATP, soit interférant avec les interactions kinases-protéines ou en régulant la kinase par l'utilisation d'ARN stratégie de brouillage. Ces kinases peuvent être ciblées par des petits composés, capables d'inhiber la phosphorylation des protéines, ce qui empêche leur activation. Les inhibiteurs de BTK sont un moyen de cibler la voie BCR, et les sous-types de lymphomes à signalisation activée, comme la LLC, sont ceux qui répondent bien aux inhibiteurs de BTK [11].

### **III.1. Les inhibiteurs de BTK approuvés**

Les trois BTK actuellement approuvés par la FDA pour traiter les tumeurs malignes à cellule B sont l'ibrutinib, acalabrutinib et zanubrutinib.

#### **III.1.1. L'ibrutinib : L'inhibiteur de BTK de première génération**

Est un inhibiteur de BTK de première catégorie, après l'échec du LFM-A13 en 1999, l'ibrutinib a été initialement choisi pour le développement clinique de modèles in vivo en 2007. En 2010 Honiberg et ses collaborateurs ont rapporté l'efficacité de ce composé dans le lymphome à cellule B et par la suite, en 13 novembre 2013 et après une série de transaction commerciales, a été approuvé par l'agence européenne des médicaments (FDA). En plus de BTK, l'ibrutinib inhibe également de manière puissante d'autres kinases qui possèdent des cystéines réactives, telles que Csk, Fgr, Lck, Brk, Hck,...

Il s'agit d'un inhibiteur ATP-compétitif dont la valeur IC50 est de 0,5 nM contre BTK est qui présente une forte activité inhibitrice contre les kinases de la famille Tec et Src. L'ibrutinib est un inhibiteur irréversible qui agit de manière covalente sur la Cys481 dans le site de liaison ATP de la BTK. L'ibrutinib a provoqué une augmentation transitoire du taux de lymphocytes dans le sang, parallèlement à une réduction de la taille des ganglions lymphatiques ou la rate. Cette lymphocytose asymptomatique s'explique par le rôle de BTK dans la signalisation des récepteurs de chimiokines.[10,8].

#### **III.1.2. L'inhibiteur de BTK de deuxième génération**

Après l'apparition d'une résistance à l'ibrutinib et d'effets secondaires hors cible a conduit au développement actif d'inhibiteur de BTK de deuxième génération et plus spécifiques.

##### **Acalabrutinib et Zanubrutinib**

Sont deux nouveaux inhibiteurs irréversibles de BTK de deuxième génération. Ils sont plus puissants et plus sélectifs que l'ibrutinib et présentes moins d'effets secondaires hors cible. Ont été approuvés par la FDA, comme traitement de deuxième ligne pour le MCL. Conformément à leurs activités biologiques enzymatiques, ils ont montré d'excellents effets thérapeutiques contre multiple

tumeurs malignes à cellules B. Ils présentaient un meilleur profil de sécurité en raison de la prévention des effets secondaires induits hors cible. L'Acalabrutinib et le Zanubritinib partagent des modes de liaison irréversibles similaires avec le site de liaison du BTK mais ont présenté une sélectivité significativement améliorée par rapport au BTK contre d'autres kinases. Ces deux inhibiteurs bloquent la voie de signalisation BCR et induit de ce fait une baisse de la croissance et de la survie des cellules B malignes.[8,7]

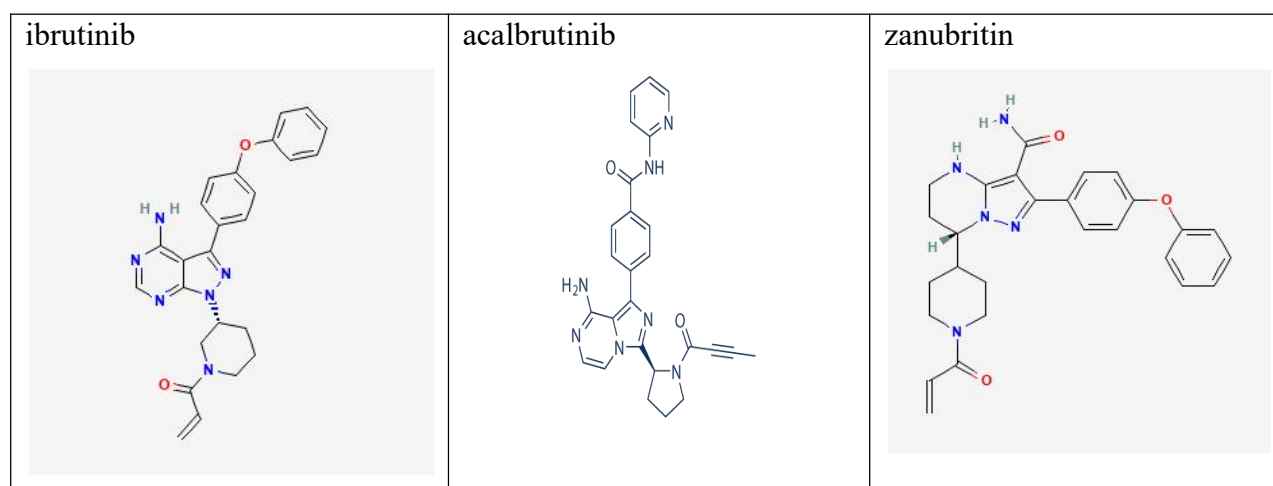


Figure 06 : La structure d'inhibiteurs de deuxième génération.

### III.2. Inhibiteurs récemment approuvés

#### Tirabrutinib et Orelabrutinib

Le tirabrutinib a récemment été approuvé pour le traitement du lymphome primaire du système nerveux central (PCNSL) au Japon en 2020. Il s'agit en outre du premier inhibiteur de BTK approuvé pour l'indication PCNSL. Son indication pour la MW et le lymphome lymphoplasmocytaire est actuellement en cours d'examen réglementaire au Japon. En outre, des essais cliniques sur les troubles auto-immunes, la LLC, le lymphome à cellules B, le syndrome de Sjogren, le pemphigus et la polyarthrite rhumatoïde sont en cours aux États-Unis, en Europe et au Japon. L'orelabrutinib est un autre inhibiteur de BTK promoteur qui a été approuvé le 25 décembre 2020 par l'administration nationale des produits médicaux (NMPA) en Chine pour le traitement de la LLC, de la SLL et de la MCL en tant que traitement de deuxième ligne. Il est actuellement évalué dans le cadre d'essais cliniques. Jusqu'à présent, les inhibiteurs de BTK sont autorisés pour le traitement des hémopathies malignes [10].

### III.3. Inhibiteurs de BTK en cours d'investigation clinique

Nombreuses molécules peuvent bloquer le BTK, de manière irréversible ou réversible.

### III.3.1. Les inhibiteurs de BTK irréversibles

Le spebrutinib, l'evobrutinib, l'olmutinib, le tirabrutinib, l'elsubrutinib et le tolebrutinib sont des inhibiteurs irréversibles de BTK actuellement en cours d'études cliniques, ces dérivés sont des composés basiques qui inhibent BTK à des concentrations nanomolaire (valeurs de IC50 de 0,5 nM, 37,9 nM, 1,0 nM et 2,2 nM pour spebrutinib, evobrutinib, olmutinib et tirabrutinib, respectivement). En mars 2020 le tirabrutinib a été approuvé au Japon pour le traitement du lymphome primaire du système nerveux et est désormais également pour le traitement du nombre de trouble auto-immuns. L'olmutinib inhibe également de manière irréversible un nombre des kinases qui partagent avec le BTK un résidu Cystéine réactif [8].

### III.3.2. Les inhibiteurs de BTK réversibles

Dits les inhibiteurs non covalents de BTK qui présentent plusieurs avantages par rapport aux inhibiteurs covalents existants. Ils peuvent conserver une inhibition puissante contre les variants C481S et C481R de BTK, En outre les inhibiteurs irréversibles présentent un risque de toxicité plus faible que les composés réversibles c'est pour quoi certains entre eux font l'objet d'études cliniques pour une administration à long terme dans le traitement des maladies auto-immunes.

## Conclusion

Ces dernières années la BTK est apparue comme une nouvelle cible en chimie médicinale pour le traitement des maladies auto-immunes et des tumeurs malignes à cellules B. L'industrie et monde universitaire ont mené développer des recherches intensives pour développer des inhibiteurs de BTK en tant qu'agent anti tumoraux. Seuls trois inhibiteurs irréversibles ont été mis sur le marché pour traiter les différents types de leucémies et lymphomes, tandis que les inhibiteurs de BTK réversibles font l'objet d'études cliniques pour une administration à long terme dans le traitement des maladies auto-immunes. L'étude des fonctions de BTK et ses applications thérapeutiques, ainsi que la découverte de nouveaux inhibiteurs de BTK et de formulations innovantes de composés approuvés, restent très fructueuses et d'un grand intérêt pour la communauté universitaire et l'industrie pharmaceutique.

## I. Relation quantitative structure-activité (QSAR)

Le développement de plusieurs techniques de modélisation est devenu très indispensable et parmi ces méthodes les plus utilisées, nous pouvons citer la méthode QSAR. Elle repose sur la recherche d'une relation quantitative ou qualitative d'une réponse biologique d'un ensemble de molécules et leurs propriétés physicochimiques à l'aide de l'utilisation de méthodes d'apprentissage statistiques et informatiques [12,13].

$$\text{Activité biologique} = f(\text{Paramètres})$$

### I.2. L'importance du QSAR en biologie

L'activité biologique des molécules est mesurée habituellement au moyen d'essais au laboratoire pour étudier le niveau d'inhibition d'une voie métabolique particulière. Mais la recherche de médicaments dans le cas où les échantillons non disponibles et les mesures expérimentales dangereux et long et très chers implique l'utilisation de modèles de types QSAR pour identifier les structures chimiques pouvant présenter de bons effets inhibiteurs sur des cibles spécifiques avant de passer au laboratoire (des études in-silico) [14].

### I.3. Principe de QSAR

L'étude de QSAR Consiste à trouver une relation mathématique reliant de manière quantitative une activité biologique mesurée pour une série de composés similaires dans les mêmes conditions expérimentales, avec des descripteurs moléculaires à l'aide des méthodes statistiques. Le modèle obtenu doit être robuste, stable et possède un pouvoir prédictif. Les descripteurs sélectionnés dans le modèle de QSAR aident les chercheurs à interpréter les principaux facteurs influençons sur la réponse biologique, comme on peut exploiter ce modèle pour la découverte de nouvelles structures chimiques par le criblage virtuel des grandes chimio-thèques [12, 14].

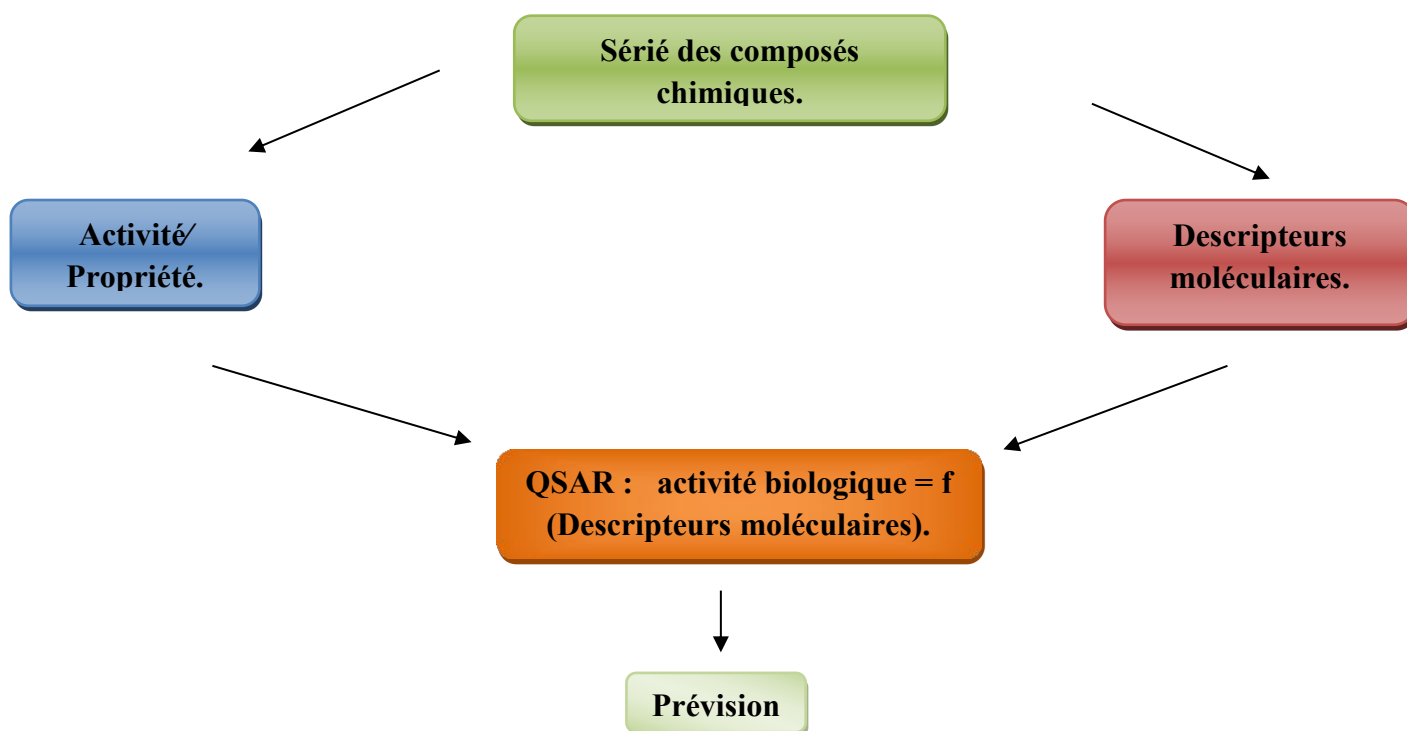


Figure 07: Modèle de l'étude de QSAR.

#### I.4. L'application du modèle QSAR

Il existe un grand nombre d'applications de l'approche QSAR, dont nous citons parmi lesquels :

- La prédiction des nouveaux inhibiteurs à partir des cibles thérapeutiques.
- La conception des médicaments et de nombreux autres produits.
- La prédiction d'une variété de propriétés physico-chimiques des molécules.
- L'optimisation de l'activité pharmacologique.
- La prédiction des effets secondaires de nouveaux composés.
- La prédiction de la toxicité pour l'homme par une exposition délibérée, occasionnelle et professionnelle.
- La conception des produits chimiques impliqués dans les processus industriels de laboratoire : solvants, réactif, etc. [16]

### I.5.Méthodologie générale d'une étude QSAR

La conception d'un modèle QSAR robuste, stable et prédictif, doit être passée par les étapes suivantes :

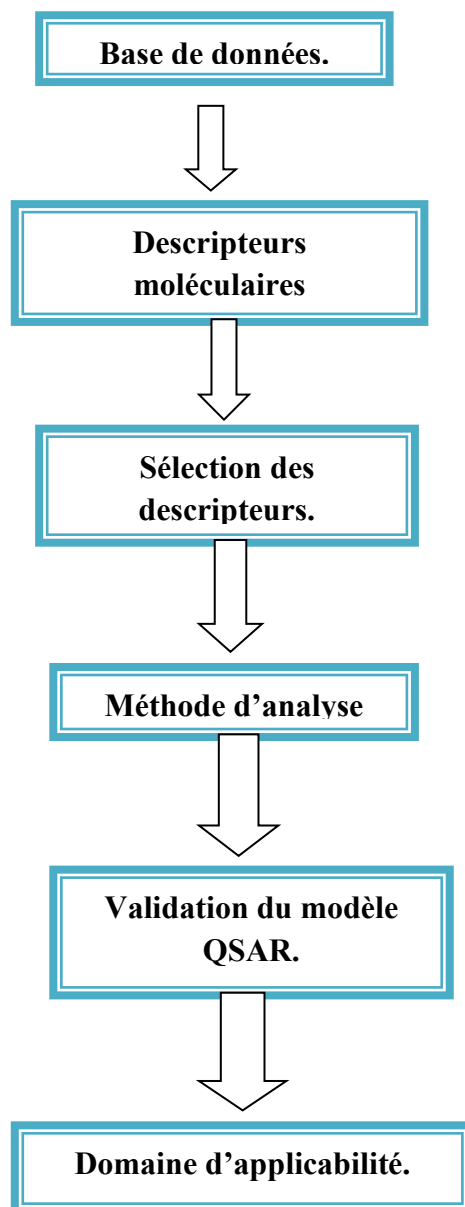


Figure 08: La modélisation du QSAR.

### I.5.1. Sélection de la Chimio thèque

Le terme chimio thèque signifie un ensemble de molécules testées expérimentalement contre une cible biologique (ou réponse biologique). Le choix de la chimio thèque possédant une réponse biologique constitue une étape très importante lors de la construction d'un modèle QSAR. L'ensemble de molécules doit être composé de données fiables (IC50, EC50, ... etc.) et les données devraient provenir du même protocole expérimental [16].

### I.5.2. Génération de descripteurs moléculaires

Les descripteurs moléculaires sont des représentations mathématiques formelles d'une molécule. La nature des descripteurs moléculaires varie en fonction de leurs types et leurs procédés de calcul. Généralement on distingue plusieurs types de descripteurs moléculaires : les descripteurs constitutionnels, topologiques, et électroniques. Ils peuvent également être classés sur la base de dimensionnalité des structures comme structure bidimensionnelle (2D) ou tridimensionnelle (3D). Plusieurs logiciels ont été développés pour la génération de descripteurs moléculaires, dont on cite : Dragon, PaDEL, Discovery Studio, MOE, ... etc. [16]

### I.5.3. Algorithmes de sélection de descripteurs moléculaires

La sélection des descripteurs moléculaires ne se fait pas au hasard donc pour un bon choix il faut vérifier que les descripteurs moléculaires tout d'abord doivent avoir une relation directe et avoir des liens avec la réponse biologique. Ensuite les descripteurs doivent être variés, simples et efficaces puisque la complexité des descripteurs provoque une perturbation sur les résultats de prédiction. L'objectif de cette étape est de réduire le nombre de descripteurs et le choix des meilleurs sous ensemble de descripteurs qui décrivent la variation des données biologiques. La procédure de sélection et de réduction des descripteurs peut être effectuée en deux étapes: [16]

#### I.5.3.1. Sélection objective

Cette sélection consiste à réduire le nombre de descripteurs sans faire participer la réponse biologique. On commence par l'exclusion de tous les descripteurs ayant un pourcentage élevé de valeurs identiques pour l'ensemble des molécules. De même, On élimine l'un des deux descripteurs présentant une forte colinéarité (le coefficient de corrélation R supérieur à 0.95), seul celui présentant la plus grande variance est retenu. Ces procédures rendent la suite de l'analyse moins coûteuse en termes de temps de calcul, puisqu'elles réduisent le nombre de descripteurs restant à traiter [14].

### **I.5.3.2. Sélection subjective**

Cette sélection se fait par ces trois étapes respectivement [14].

#### **I.5.3.2.1. Introduction progressive**

C'est-à dire l'incorporation des descripteurs moléculaires au modèle un à un en sélectionnant à chaque étape le descripteur moléculaire dont la corrélation avec la réponse biologique (comme pIC50) est la plus élevée.

#### **I.5.3.2.2. Elimination progressive**

C'est la construction du modèle QSAR avec la totalité des descripteurs et on commence de les éliminer en gardant ceux qui permettent l'obtention d'un modèle ayant une bonne corrélation avec la réponse biologique.

La fonction d'évaluation soit d'origine statistique tels que : le coefficient de détermination  $R^2$ , l'écart type  $s$ , test de Fisher  $F$ , et le coefficient de corrélation issu du cross validation  $R^2_{cv-100}$ , ce qui reflètent la qualité de chaque solution candidate.

#### **I.5.3.2.3. Sélection pas à pas (Forward Stepwise)**

C'est la combinaison des deux méthodes précédentes. Les descripteurs moléculaires sont incorporés un à un dans le modèle par une sélection progressive. Cependant, à chaque étape, nous vérifions que les corrélations partielles des variables (descripteurs moléculaires) précédemment introduites sont encore significatives.

### **I.5.4. Algorithmes d'apprentissages**

Pour élaborer un modèle QSAR nous avons besoin d'un algorithme d'apprentissage (analyse quantitative ou qualitative). L'algorithme d'apprentissage nous permet de quantifier la relation entre l'activité biologique et les descripteurs moléculaires. Dans l'approche QSAR, plusieurs algorithmes d'apprentissage sont appelés, dont on cite [14] :

#### **I.5.4.1. La régression linéaire multiple (MLR)**

L'apprentissage par la régression linéaire multiple (MLR) est basé sur l'élimination des descripteurs non significatifs (un par un) jusqu'à l'obtention d'un modèle valide (incluant la probabilité critique  $< 0,05$  pour tous les descripteurs et le modèle complet). Mathématiquement, l'algorithme est utilisé pour trouver une fonction linéaire entre l'activité biologique exprimée par la grandeur biologique pIC50 et l'ensemble de descripteurs moléculaires sélectionnées dans l'étape précédente. La méthode MLR résout l'équation  $Y = X.A + B$ , où  $Y$ ,  $X$ ,  $A$  et  $B$  représentent respectivement le vecteur de la propriété, la matrice des descripteurs moléculaires, la matrice des coefficients et la matrice des erreurs de régression [12].

### I.5.4.2. La régression aux moindres carrés partiels (PLS)

Est une méthode rapide, efficace et optimale pour un critère de minimisation, son utilisation est recommandée dans le cas où un grand nombre de variables explicatives est utilisé, ou lorsqu'il y a de fortes colinéarité entre les variables [12].

### I.5.4.3. Réseau neuronal artificiel (ANN)

Généralement dans les études de QSAR, les chercheurs utilisent le réseau de neurones à base de Feed Forward à trois couches : une couche d'entrée pour les descripteurs moléculaires, une couche cachée de nombre de neurones variables et une couche de sortie représente la variable dépendante ou bien la réponse biologique. Avec une fonction de transfert non linéaire de la couche d'entrée vers la couche cachée et une fonction linéaire de la couche cachée vers la couche de sortie. L'évaluation de l'apprentissage se fait par la minimisation de l'erreur entre les résultats obtenus (pIC50 prédites) les valeurs expérimentales [12].

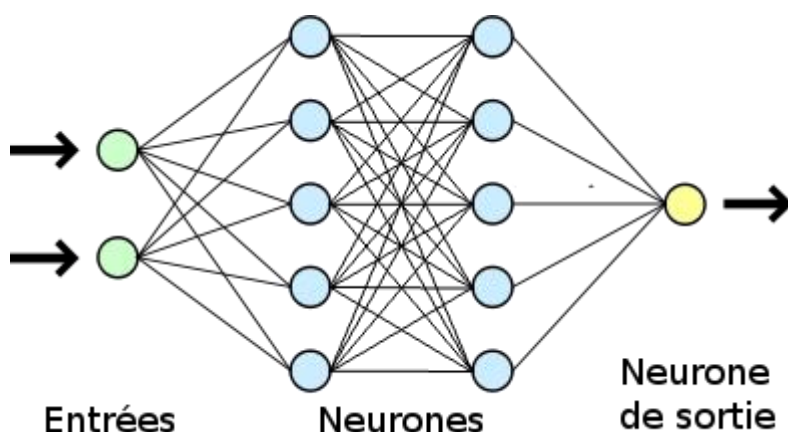


Figure 09: Schéma représentatif d'un réseau de neurones artificiels

## I.5.5. Interprétation et validation d'un modèle QSAR

La validation des modèles est nécessaire pour estimer leur fiabilité et déterminer son pouvoir explicatif et son pouvoir prédictif. Pour ce faire, trois méthodes sont employées : les méthodes de validation interne, le test de randomisation et la validation externe [14, 16].

### I.5.5.1. Validation interne

Validation croisée, Leave-One-Out (LOO) ou Leave-Many-Out (LMO) est la méthode la plus utilisée dans la plupart du temps dans QSAR. La procédure Leave-One-Out consiste à retirer successivement une molécule de l'ensemble d'apprentissage, alors que la procédure Leave-n-Out correspond à un découpage en plusieurs parties de l'ensemble de molécules de l'apprentissage.

La performance des modèles de régression est estimée avec les paramètres statistiques de la validation croisée  $s_{cv-loo}$  (l'écart type de la validation croisée), et  $Q^2_{cv-loo}$  (coefficient de détermination issus de la validation et représente la capacité de la prévision du modèle QSAR).

Parmi les méthodes utilisées pour faire la validation interne, la randomisation de la réponse biologique [14].

#### ❖ Test de randomisation

Le test de randomisation permet d'affirmer que la corrélation de chance ne joue aucun rôle durant le développement du modèle. Ces tests sont exécutés quantitativement avec les modèles de QSAR et qualitativement avec les modèles issus de la classification. La supposition sous-jacente du test de randomisation est la suivante : si les capacités prédictives du modèle ne sont pas dues aux corrélations de chance, alors la désorganisation aléatoire des observations conduira à des modèles (quantitatifs ou qualitatifs) de prévisions faibles [14, 15].

#### I.5.5.2 Validation Externe

Il n'est pas suffisant de développer un modèle de QSAR avec une excellente qualité d'ajustement et de prévision, mais il est également nécessaire de généraliser ces prévisions en l'appliquant sur un ensemble de molécules externe (généralement représente de 20 à 30 % de la totalité des molécules). Les prévisions résultantes pour PSET déterminent alors la validité externe du modèle. Les mesures quantitatives de la prévision par validation externe peuvent être identiques à celles utilisées pour la validation interne [14].

#### I.5.6. Domaine d'applicabilité

Un modèle QSAR ne peut pas être considéré comme un modèle universel, puisqu'il est développé sur un nombre limité de composés qui ne couvrent pas tout l'espace chimique. Par conséquent l'activité/propriété prédite d'un composé, chimiquement dissimilaire au jeu d'apprentissage, ne pourra pas être considérée fiable. Le domaine d'applicabilité (DA) permet de définir la zone dans laquelle un composé pourra être prédit avec confiance. Le DA correspond donc à la région de l'espace chimique incluant les composés de la série d'apprentissage et les composés similaires, proche dans ce même espace [12, 14, 15].

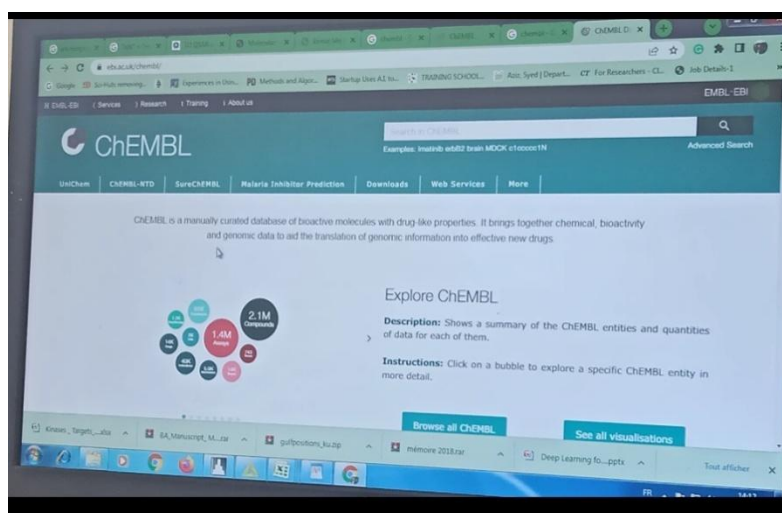
### Conclusion

Dans ce chapitre nous avons abordé la méthodologie QSAR en décrivant ses différentes caractéristiques, la mise en place de modèle QSAR n'est pas une chose aisée. Un des problèmes importants réside également dans le traitement de données en grande quantité. Un grand nombre de descripteurs et de molécules peuvent être à analyser, mais aucune règle stricte n'existe quand au choix des paramètres structuraux les plus importants parmi le jeu complet de ceux disponible.

## I. Matériels et Méthodes

### I.1. Préparation de la chimio thèque

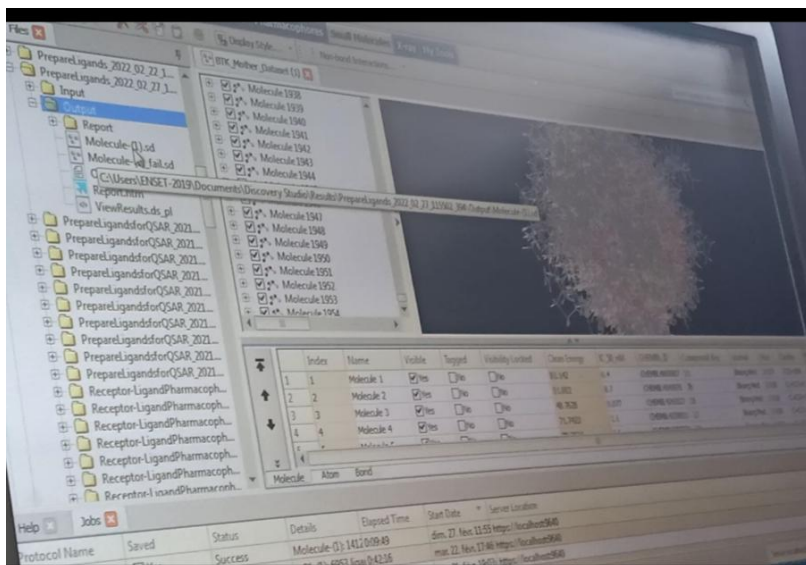
Nous avons pu télécharger plus de 1955 molécules ayant une activité inhibitrice contre l'enzyme de tyrosine kinase « BTK » à partir de la base de données ChEMBL (<https://www.ebi.ac.uk/chembl/>) dont les molécules possédant des valeurs non définies ont été éliminées.



**Figure10** : Base de données ChEMBL.

Nous avons collecté une chimio thèque constituée de 1317 molécules possédant une valeur d'IC<sub>50</sub> (grandeur biologique représente la concentration qui inhibe 50% de l'activité enzymatique).

La préparation de la base de données (1317 molécules) a été effectuée à l'aide d'un logiciel du Discovery Studio (Biovia), dont il a converti l'extension smiles (.smi) en extension .sdf (structure data file) avec la génération des coordonnées cartésiennes pour chaque atome, suivie par une optimisation structurale à l'aide du même logiciel.



**Figure11** : Préparation des molécules par le logiciel discovery studio (Biovia).

## II. Génération des descripteurs moléculaires

La matrice de descripteurs moléculaires a été générée par le logiciel Dragon version 06. Le nombre total de descripteurs moléculaires calculés par ce logiciel est 4860 descripteurs réparties en 20 blocks dont la plupart sont des descripteurs à base topologique et constitutionnels.

### II.1. Répartition de la chimio thèque

La chimio thèque représenté par le vecteur de la réponse biologique et la matrice de descripteurs moléculaires est divisé en deux sous-ensembles. Le premier sous – ensemble est réservé pour l'apprentissage et constitué de 988 molécules alors le reste qui représente 25% de la chimio thèque est réservé pour la validation externe des modèles obtenus.

La division de la chimio thèque est effectuée en fonction des valeurs ascendantes de pIC50 (-LOGIC50).

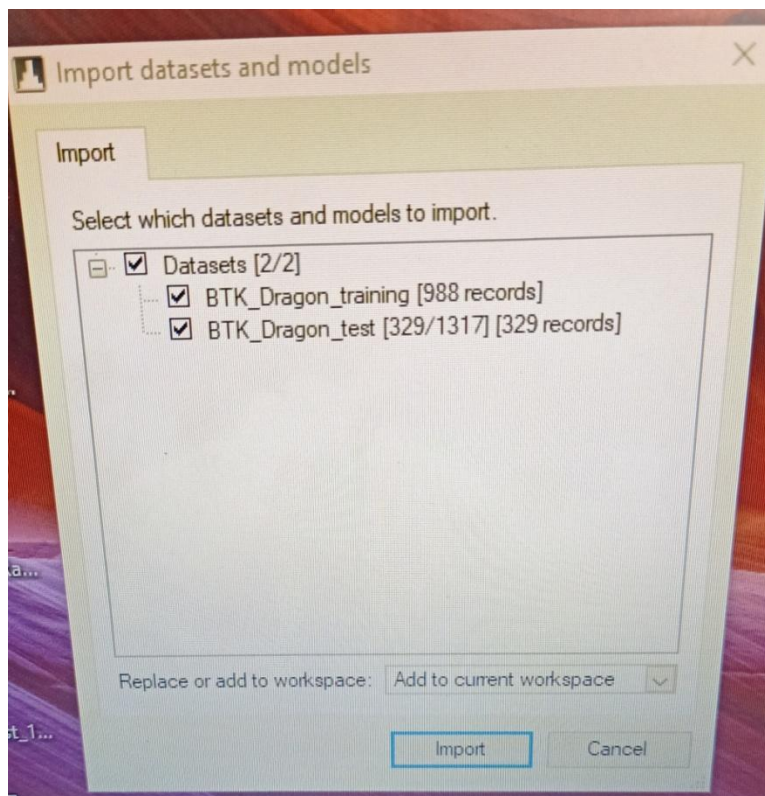


Figure12 : Répartition de la chimio thèque en deux groupes training et test.

## II.2. Prétraitement de la matrice de descripteurs

Afin d'éviter la redondance de l'information et minimiser le temps de calculs lors de la phase de sélection de descripteurs, une filtration a été effectuée à l'aide du Data Modeler (un module d'un logiciel Molegro Virtuel docker version 06).

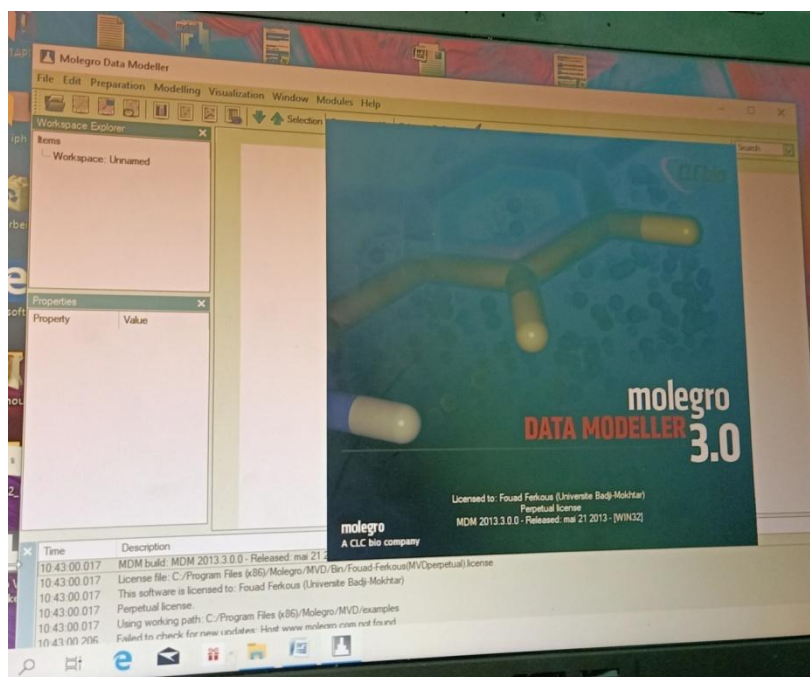


Figure13: Le logiciel Molegro Virtuel.

Dans cette étape, le logiciel travail uniquement sur la matrice de descripteurs moléculaires générés par Dragon et sans l'intervention du vecteur de la réponse biologique (IC50). On supprime tous les descripteurs ayant une valeur constante pour l'ensemble des molécules d'apprentissage.

Aussi, on élimine tous les colonnes de descripteurs présentant une ou plusieurs valeurs non définies.

La suppression de l'un des deux descripteurs présentant une forte colinéarité (le seuil du coefficient de corrélation est fixé à  $R=0.9$ ).

### II.3. Construction des modèles QSAR

Dans cette étape, plusieurs algorithmes d'apprentissage ont été appelés afin de sélectionner le meilleur algorithme d'apprentissage.

Nous avons sélectionné le Forward sélection comme algorithme de sélection de variables installé dans le même logiciel (Molegro – Data Modeler).

Les algorithmes d'apprentissage utilisés dans cette étude sont : la régression linéaire multiple, PLS et L'XGBOOST.

### II.4. Validation externe des modèles

Les modèles obtenus sont validés par les calculs de leur pouvoir prédictif en se basant sur la prédiction de l'activité biologique (pIC50) du sous-ensemble de validation externe.

Les paramètres statistiques suivants :

**R<sup>2</sup>** : coefficient de détermination qui mesure la variation entre les valeurs expérimentales (pIC50 exp.) et les valeurs calculées par le modèle.

**Q<sup>2</sup>** : coefficient de validation externe.

## I. Résultats et discussion

L'enzyme de tyrosine a une grande importance dans le domaine médicinale et industries alimentaire on précise un groupe de composants qui occupent l'information concernant le degré d'inhibition enzymatique. D'après la procédure de sélection et le tirage par l'utilisation de logiciel **Discovery studio** le nombre de molécule atteindre 1412, mais après le calcul des descripteurs moléculaire et le fais d'opération du triage et l'élimination de tous les descripteurs qui obtient des chiffres constants et qui ne possède aucune valeur nous avons obtenu 1317 molécules. Après l'opération de division en deux groupes l'un pour **l'apprentissage** (training) avec 988 molécule et l'autre pour **la validation externe** (test) avec 329 molécule, nous avons construire des modèles par l'utilisation de groupes des algorithmes d'apprentissage.

Ce qui suit une représentation des modèles essentiels obtenus par l'utilisation de logiciel Data Modeler sur le groupe « Apprentissage ».

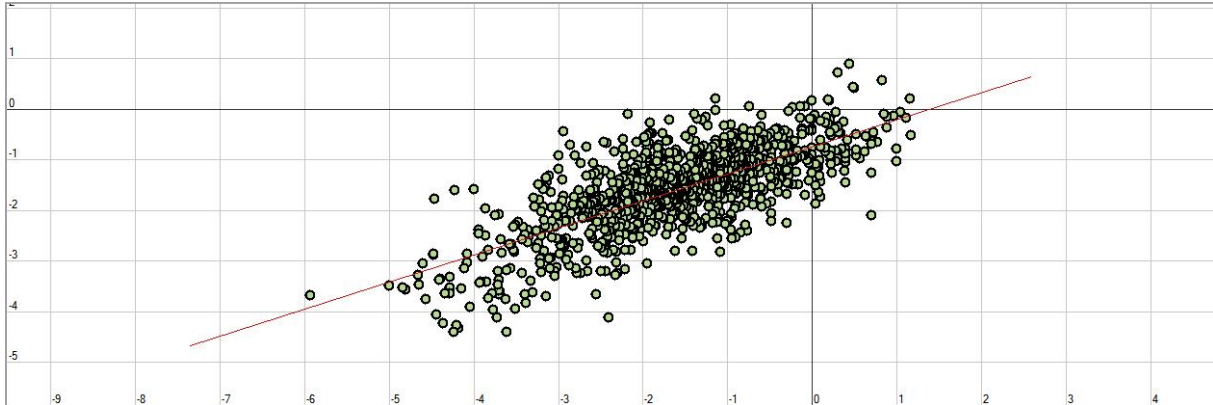
### I.1. L'étude quantitative par l'utilisation de MLR

Dans cette étape nous avons obtenu une matrice de 44 descripteurs et nous avons essayé de relier ces descripteurs avec l'activité biologique (pIC50), à partir des résultats obtenus par l'application de cette algorithmes, nous avons obtenu un bon modèle constitué de 44 descripteurs comme une variable indépendante influé à l'activité biologique et présenté par la valeur pIC50. Dans ce modèle il y a des descripteurs convient positivement avec la quantité biologique, c'est-à-dire plus la valeur de descripteur est grande, plus la valeur de l'activité biologique est augmentée. Alors que certains descripteurs sont affectés négativement à l'activité biologique. Le nombre de descripteurs qui convient positivement avec l'activité biologique est 23 et les descripteurs qui convient négativement est 21.

- Dans ce qui suit l'affichage de l'équation mathématique présenté de ce modèle.

$$\begin{aligned} \text{pIC50} = & 0.79664 * \text{nR07} - 5.53299 * \text{ARR} + 3.2604 * \text{PCR} - 3.34916 * \text{MATS3m} + 30.4466 * \\ & \text{JGI7} + 5.62593 * \text{Eig13\_EA(dm)} - 7.683 * \text{VE2\_RG} - 0.0741734 * \text{Mor10u} + 0.168501 * \text{Mor18m} \\ & + 0.381819 * \text{Mor21m} - 0.219464 * \text{Mor06p} - 0.288832 * \text{Mor09p} + 0.62096 * \text{Mor24p} + 0.037248 \\ & * \text{Mor12s} - 4.06595 * \text{"R2m+"} + 1.49138 * \text{R5p} + 0.405553 * \text{nR=Cp} + 0.910201 * \text{nR\#CH/X} - \\ & 1.05768 * \text{nArCO} - 1.91283 * \text{nRNH2} + 0.303689 * \text{nPyrrolidines} - 0.928084 * \text{nPyrroles} - \\ & 0.615274 * \text{"C-011"} + 0.272104 * \text{"C-034"} + 0.398195 * \text{"N-073"} + 0.385391 * \text{CATS2D\_09\_DP} - \\ & 0.00294074 * \text{CATS2D\_02\_AL} + 0.0320982 * \text{CATS2D\_05\_AL} + 0.16021 * \text{CATS2D\_06\_PL} - \\ & 0.632281 * \text{CATS2D\_07\_NL} + 0.921122 * \text{"B03[O-Cl]"} + 0.417585 * \text{"B04[N-F]"} - 0.204757 * \\ & \text{"B05[O-O]"} + 1.10029 * \text{"B06[O-Cl]"} - 0.259246 * \text{"B07[N-O]"} + 0.783559 * \text{"B07[N-Cl]"} + \end{aligned}$$

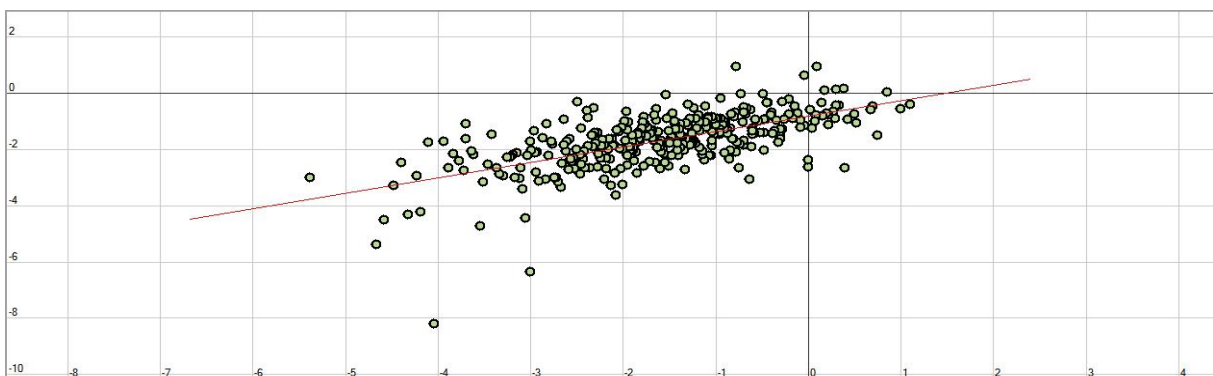
$$0.255469 * "B07[O-O]" - 0.740515 * "B07[O-S]" + 0.33836 * "B08[O-F]" + 0.356122 * "B10[N-S]" - 0.0736177 * "F06[N-N]" - 0.21125 * "F09[O-O]" - 0.342197 * LLS\_01 + 0.35288 * "Psychotic-80" - 5.32494$$



**Figure14** : Représente la variation de valeur d’IC50 mesurée par MLR de groupe d’apprentissage (training).

**I.1.1. Interprétation des résultats calculés par le groupe par l’apprentissage**

La figure représente la traduction de changement des valeurs calculées par le modèle de MLR et les valeurs mesurées au laboratoire. Pour mesurer la force des modèles nous avons abandonné les données de groupe d’apprentissage (training) par l’équation précédente et nous avons calculés la valeur statistique  $R^2=0,4$  dont la valeur était inférieure à la valeur requise.



**Figure15**: Représente la variation de valeur d’IC50 mesurée par MLR de groupe (Test)

### I.1.2. Interprétation des résultats calculés par la validation externe

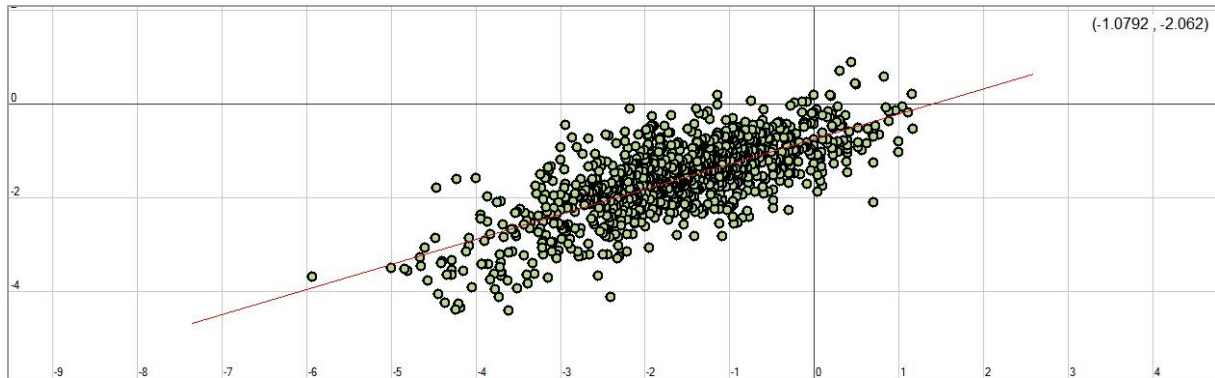
La figure représente la traduction de changement des valeurs calculées par le modèle de MLR et les valeurs mesurées au laboratoire. Pour mesurer la force des modèles nous avons abandonné les données de groupe de (test) par l'équation précédente et nous avons calculés la valeur statistique  $Q^2=0,53$  dont la valeur était supérieure à la valeur requise.

### **I.2. L'étude quantitative par l'utilisation de PLS**

À partir des résultats obtenus par l'application de cet algorithme, nous avons obtenu un bon modèle constitué de 15 descripteurs comme une variable indépendante influé à l'activité biologique et présenté par la valeur  $pIC_{50}$ . Dans ce modèle il y a des descripteurs convient positivement avec l'activité biologique, c'est-à-dire plus la valeur de descripteur est grande, plus la valeur de l'activité biologique est augmentée. Alors que certains descripteurs sont affectés négativement. Le nombre des descripteurs qui convient positivement avec la quantité biologique est 24 et les descripteurs qui convient négativement est 20.

**- Dans ce qui suit l'affichage de l'équation mathématique présenté de ce modèle.**

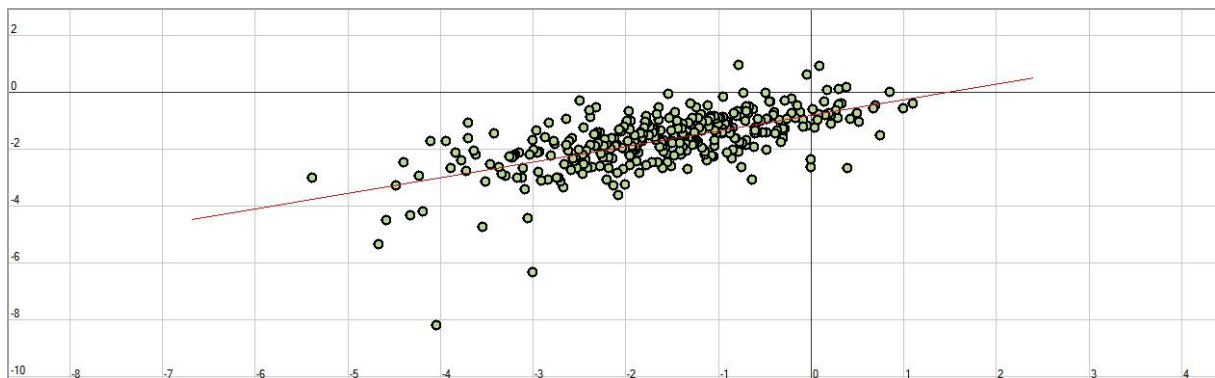
$$\begin{aligned}
 pIC_{50} = & 0.799758 * nR07 - 5.52415 * ARR + 3.20826 * PCR - 3.35827 * MATS3m + 30.1716 * JGI7 + \\
 & 5.64167 * Eig13\_EA(dm) - 7.62613 * VE2\_RG - 0.0748809 * Mor10u + 0.1573 * Mor18m + 0.392347 * \\
 & Mor21m - 0.219152 * Mor06p - 0.285564 * Mor09p + 0.62817 * Mor24p + 0.037646 * Mor12s - 4.04283 * \\
 & "R2m+" + 1.51275 * R5p + 0.403686 * nR=Cp + 0.90756 * nR#CH/X - 1.05063 * nArCO - 1.90186 * \\
 & nRNH2 + 0.305574 * nPyrrolidines - 0.928955 * nPyrroles - 0.614805 * "C-011" + 0.274154 * "C-034" + \\
 & 0.394343 * "N-073" + 0.39004 * CATS2D_09_DP - 0.0027229 * CATS2D_02_AL + 0.032143 * \\
 & CATS2D_05_AL + 0.159126 * CATS2D_06_PL - 0.639664 * CATS2D_07_NL + 0.919466 * "B03[O-Cl]" \\
 & + 0.411712 * "B04[N-F]" - 0.20631 * "B05[O-O]" + 1.10653 * "B06[O-Cl]" - 0.254331 * "B07[N-O]" + \\
 & 0.787535 * "B07[N-Cl]" + 0.257141 * "B07[O-O]" - 0.746923 * "B07[O-S]" + 0.344451 * "B08[O-F]" + \\
 & 0.354989 * "B10[N-S]" - 0.0747142 * "F06[N-N]" - 0.212047 * "F09[O-O]" - 0.349533 * LLS_01 + \\
 & 0.356857 * "Psychotic-80" - 5.27256
 \end{aligned}$$



**Figure16** : Représente la valeur calculée d'IC50 mesurée par PLS de groupe d'apprentissage (training).

### I.2.1. Interprétation des résultats calculés par le groupe d'apprentissage

La figure représente la traduction de changement des valeurs calculées par le modèle de PLS et les valeurs mesurées au laboratoire. Pour mesurer la force des modèles nous avons abandonné les données de groupe d'apprentissage (training) par l'équation précédente et nous avons calculés la valeur statistique  $R^2=0,4$  dont la valeur était inférieure à la valeur requise.



**Figure17** : Représente la valeur calculée d'IC50 mesurée par PLS de groupe (Test).

### I.2.2. Interprétation des résultats calculés par le groupe test

La figure représente la traduction de changement des valeurs calculées par le modèle de PLS et les valeurs mesurées au laboratoire. Pour mesurer la force des modèles nous avons abandonné les données de groupe de (test) par l'équation précédente et nous avons calculés la valeur statistique  $Q^2=0,53$  dont la valeur était supérieure à la valeur requise.

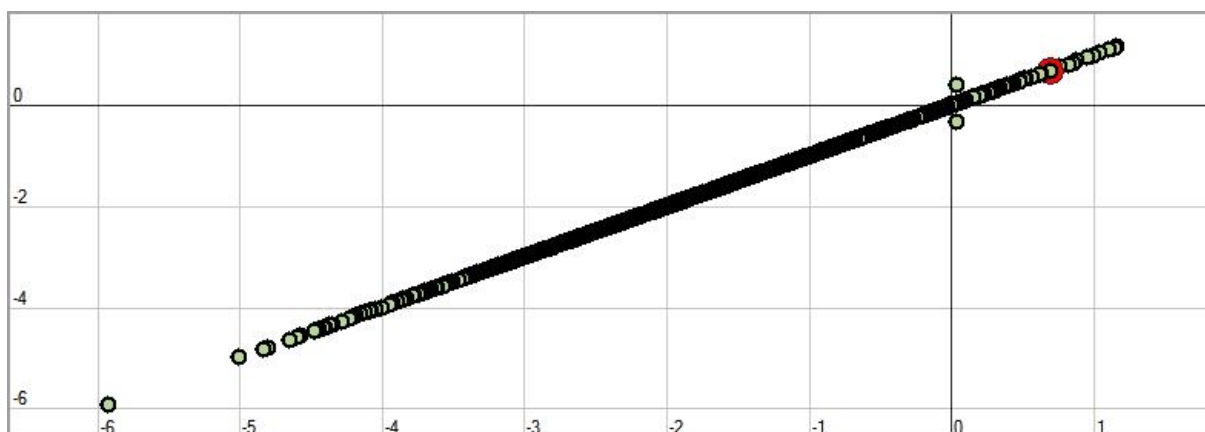
### I.3. Application de l'algorithme d'apprentissage Boosting de Gradient

#### I.3.1. XG-BOOST Régression

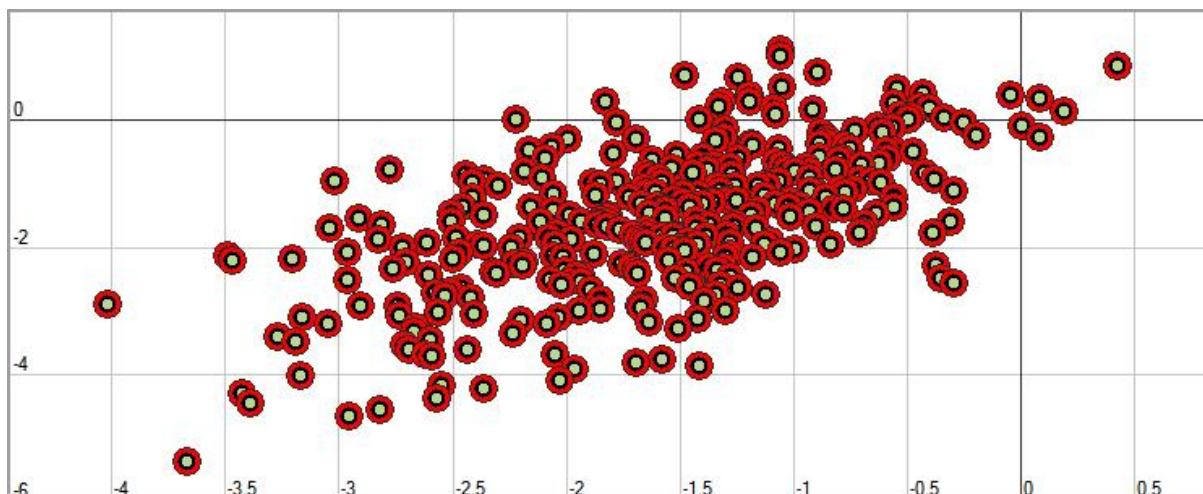
Nous avons appliqué un algorithme d'apprentissage supervisé porte le nom de Boosting de Gradient (XGBOOST). Le principe de son fonctionnement est basé sur la combinaison des résultats d'un ensemble de modèles plus simple afin de fournir une meilleur prédiction.

Nous avons appliqué cet algorithme sur l'ensemble d'apprentissage composé de 935 molécules et l'ensemble de prédiction externe pour valider le modèle obtenu, et nous avons obtenu les résultats suivant :

- Le coefficient de détermination issus de l'emble d'apprentissage et qui mesure la corrélation de la variable dépendante (pIC<sub>50</sub>) avec les descripteurs moléculaires était :  $R^2$  in train=0.99. c'est-à-dire que le modèle explique plus de 99 pour cent de la variation des valeurs de pIC<sub>50</sub> expérimentales.
- Lors de la validation externe du modèle de XGBOOST de régression obtenus, nous avons l'appliqué sur l'ensemble de prédiction et la valeur du coefficient de validation externe était =0,36.
- La valeur de  $Q^2$  présente une faible valeur et loin d'être applicable dans le criblage virtuel des bases de données.
- Nous avons tracé la projection des valeurs expérimentales en fonction des valeurs calculées de pIC<sub>50</sub> pour les deux sous-ensembles d'apprentissage et de validation externe.



**Figure18** : Projection de la variation des valeurs d pIC<sub>50</sub> expérimentales en fonction de pIC<sub>50</sub> calculés pour l'ensemble d'apprentissage.



**Figure19** : Projection de la variation des valeurs d  $pIC_{50}$  expérimentales en fonction de  $pIC_{50}$  calculés pour l'ensemble de validation externe.

Afin de construire un modèle de QSAR possède un bon pouvoir prédictif, nous avons exploité l'algorithme de XGBOOST en tant que modèle de classification. Ceci est appliqué par la conversion des valeurs quantitatives de  $IC_{50}$  expérimentales en valeurs logiques (actives et inactives), sachant que les valeurs de  $IC_{50}$  supérieure seuil  $IC_{50} = 200$  nM sont classées molécules inactives.

L'application de cet algorithme de classification nous a fourni une matrice de confusion mentionnée en – dessous :

<b>precision</b>	<b>recall</b>	<b>f1-score</b>	<b>support</b>	
<b>Active</b>	0.73	0.80	0.76	186
<b>Nonactive</b>	0.70	0.62	0.66	143
<b>accuracy</b>			0.72	329
<b>macroavg</b>	0.72	0.71	0.71	329
<b>weightedavg</b>	0.72	0.72	0.72	329

D'après cette matrice de confusion, le modèle possède un bon pouvoir prédictif avec un taux d'exactitude égale à 73%.

Cette valeur de prédiction nous a encouragés à exploiter le modèle obtenu dans un criblage virtuel.

Nous avons téléchargé une chimiothèque composé de 10460 molécules obtenues à partir de la base ZINC (produits naturels).

L'application du modèle nous a donné une série de 667 molécules actives contre la BTK et 9787 molécules inactives.

Les molécules sont listées dans l'**annexe 01**.

## **Conclusion**

L'objectif de ce travail est d'identifier de nouvelles structures chimiques capables à inhiber la cible anticancéreuse Bruton Tyrosine Kinase. Nous avons commencé par une étude structurale et mécanisme d'action de l'enzyme BTK, suivie par une introduction à l'approche Quantitative Structure Activity Relationship. Dans la partie expérimentale, nous avons construit plusieurs modèles quantitatifs en utilisant une chimio thèque constituée de 1317 molécules ayant une valeur d'IC50. Les algorithmes d'apprentissage linéaires et non linéaires. Les faibles valeurs de prédiction obtenues par les méthodes MLR, PLS et XGBOOST, nous ont poussées à l'analyse qualitative en utilisant l'algorithme de classification XG-BOOST. D'après la matrice de confusion et la bonne valeur de précision, nous a encouragés à l'utiliser dans le criblage virtuel d'une chimio thèque constituée de 10460 molécules de la base ZINC. La phase de criblage virtuel a donnée un ensemble de 667 molécules actives contre la Bruton Tyrosine Kinase. Ces molécules peuvent être le projet d'une étude complémentaire par le docking moléculaire afin de sélectionner les meilleures hits et les tester in-vitro et in-vivo.

## Référence

- [1] <https://www.aps.dz/ar/>
- [2] <https://Sante.journaldesfemmes.fr/>, Dr Jean Ménard ,15 avril 2020 à 9 :39 pm
- [3] [http://www.cos-eu.com/sinformer-sur le -cancer/mieux-comprendre-les différents-types -de -cancer/](http://www.cos-eu.com/sinformer-sur-le-cancer/mieux-comprendre-les-differents-types-de-cancer/).15 avril à 9:39 pm
- [4] (Guides des cancers): **CARINE MAILLARD, DR DIDIER VANDER STEINHEIL, PATRICIA SERVAIS PHD**, Année 2017, page 27,28,41,42,43,71,72,72,73 de 124 pages, le 17 avril 2022
- [5] **CHARBONNEAU DAVID**. Elément structuraux essentiels à la thermo stabilité de nouvelles enzymes lipolytique. Biophysique et biologie cellulaires. Université du gué bec : université du Québec à trois –rivières ,2013,266
- [6] **J.W. KEILLOR**. Inhibition enzymatique. Université de Montréal. Faculté des arts et des sciences. Département de chimie .2019.103 page
- [7] **DELA.AHMED**.la cinétique enzymatique. Université 7 novembre à Carthage. Faculté des sciences de Bizerte. Département de biologie .2010.23 page
- [8] **JINGJING WU, CHRISTINA LIU, STELLA T TSUI AND DELONG LUI**,2016, Second – génération inhibitors of Burton Tyrosine Kinase, Journal of hematology & oncology, DOI 10.1186/s13045-016-0313-y, page 1 de 7 pages
- [9] **CHIARA BRULLO, CARLA VILLA, ANDREA SPALLAROSSA**,2021, BTK inhibitors: A medicinal chemistry and drug delivery perspective, International journal of molecular sciences, DOI.org/10.3390/ijms22147641, page 2 de 16 pages
- [10] **YIFAN FENG, WEINING DUAN, XIAOCHUAN CU, CHENGYUAN LIANG & XIN MINHANG**,(2010-2018), Journal Escpert opinion on thérapeutique patents, DOI:10,1080/13543776.2019.1594777, page 3,5,6,7 de 61 pages
- [11] **JIAKUO LIU, CHENGJUAN CHEN, DONGNEI, JIE ZHANG, TIANAI ZHANG**, 2021, Emerging small-molecule inhibitors of the Burton’s Tyrosine Kinase(BTK):Current development, European journal of medicinal chemistry 217, 023-5234, page 1et2 de 24 pages
- [12] **RUDI.W.HENDRIKS, SARAVAUAN YUVARJ ET LAUNEUS P. KIL**. 2014. Targeting Burton’s tyrosine Kinas in B cell malignancies. Nature Reviews Cancer, page 2,3,7 de14 pages

[13] **SAMIR CHTITA**, Modélisation de molécule organique hétérocyclique biologiquement actives par les méthodes QSAR/QSPR. Recherche de nouveau médicaments, Meknès au Maroc, Université Moulay Ismail, Année 2017, page 21, 45,46,47,48 de 186 pages.

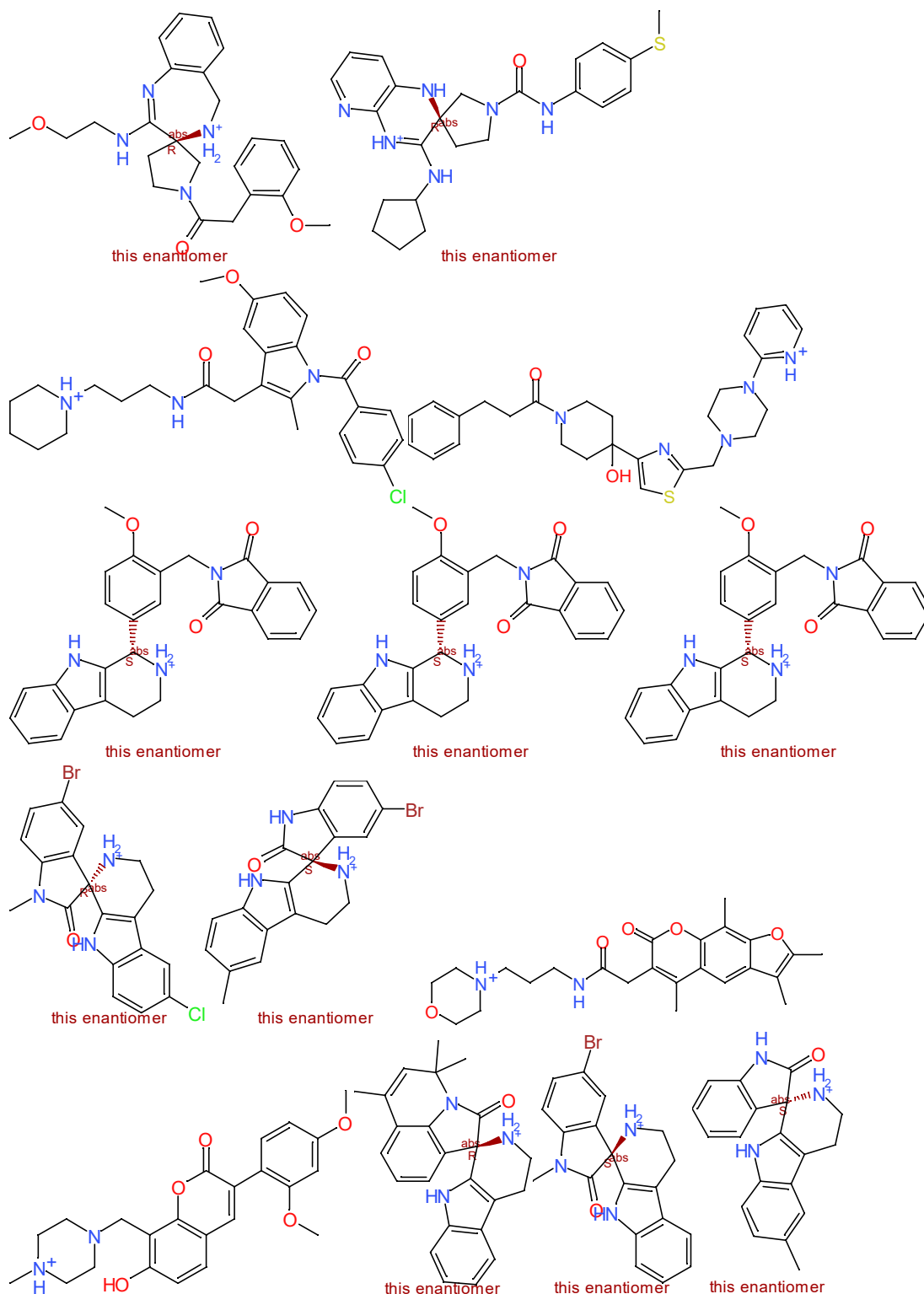
[14] **KHAINNAR MEHDI, AIMEUR OUSSAMA**, Etude des relations quantitatives structure-activité (QSAR) de plusieurs familles de composés sur des données biologique, Constantine, Université Frères Mentouri, 2018-2019, page 29 de 61 pages.

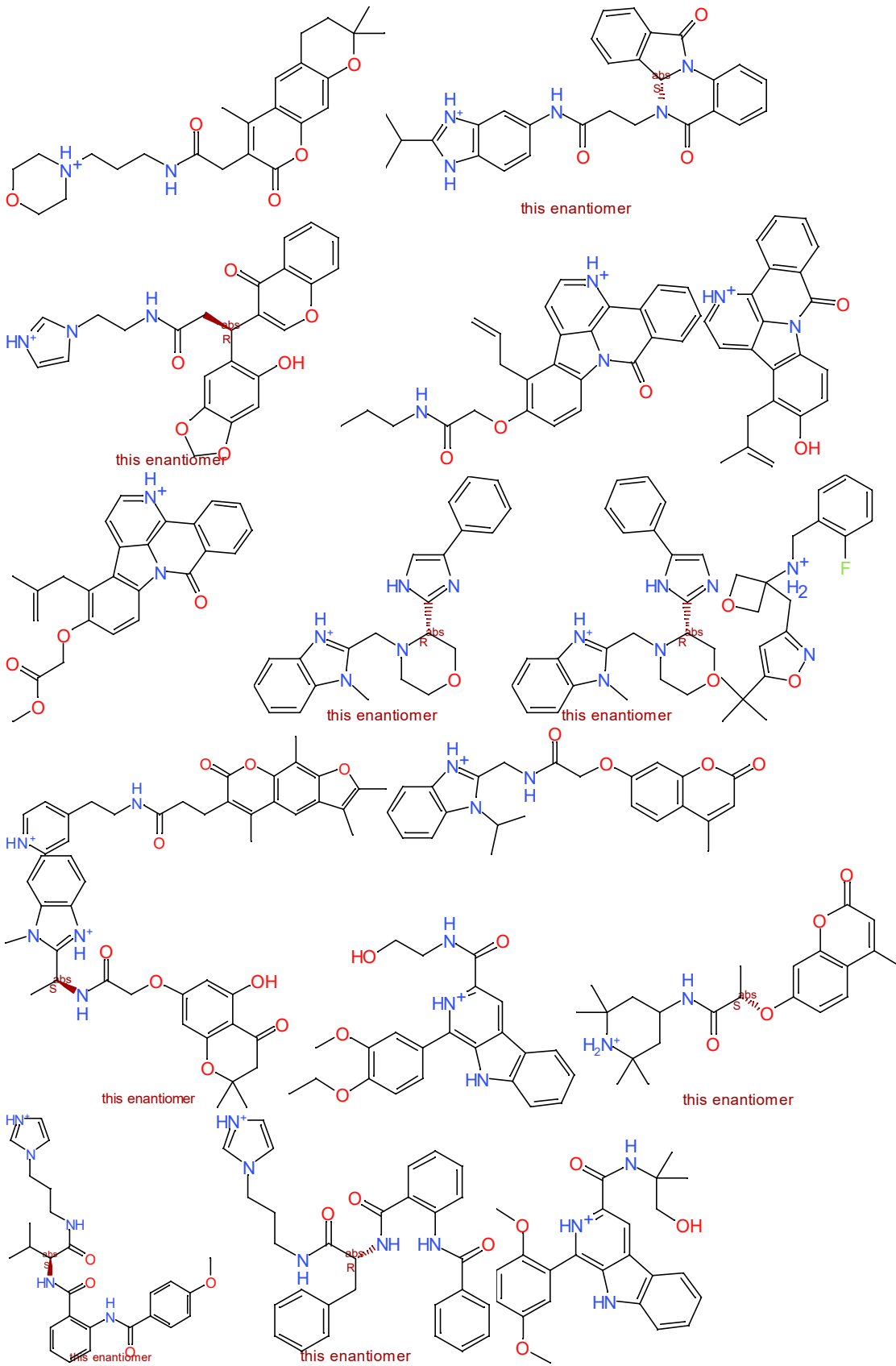
[15] **M. SAIHI YOUSEF**. Etude de la relation quantitative structure –activité inhibitrice des enzymes hydrolytique : cas des alphas galactosidases. Université Badji Mokhtar Annaba. Année 2015, page 55,57,67,69,70,75,76,77 de 109 pages.

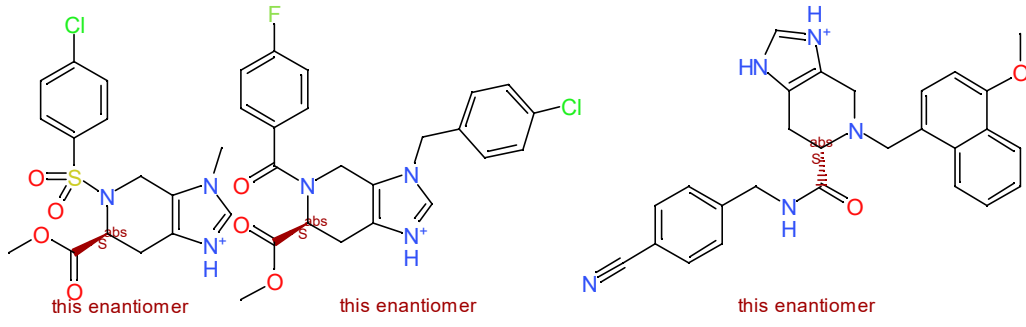
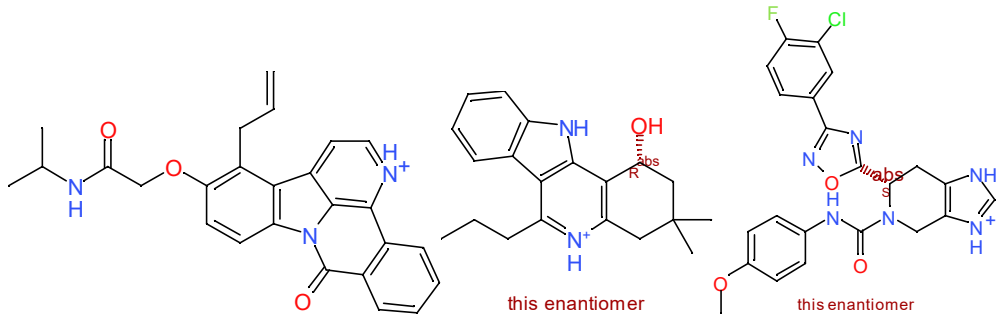
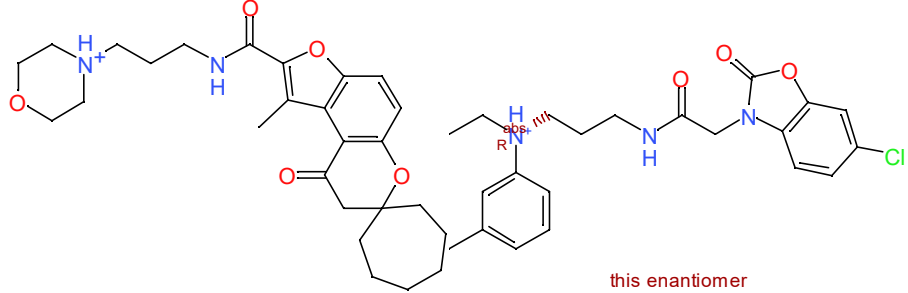
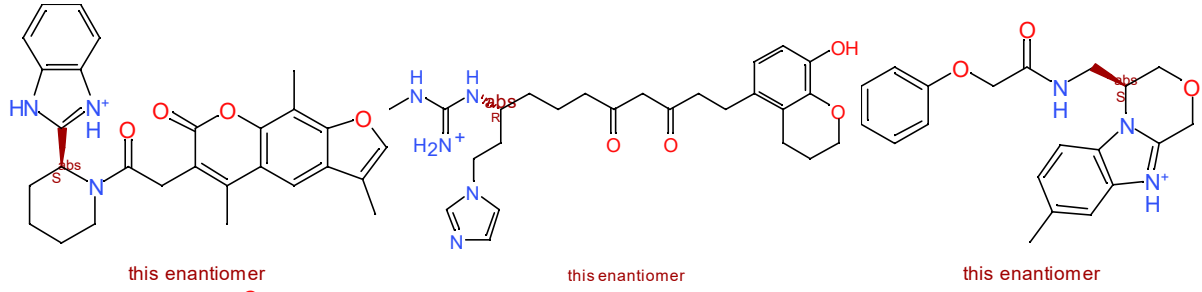
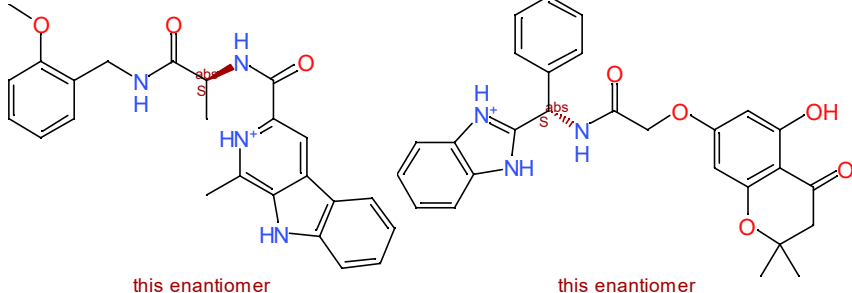
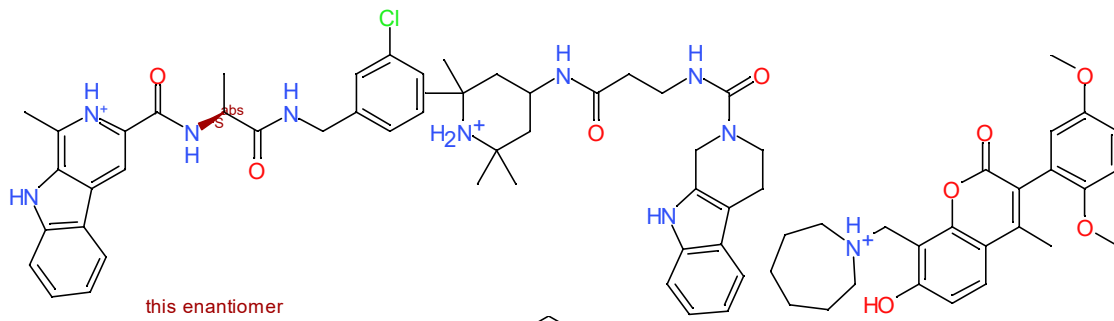
[16] **Mr YOUFI YUCEF**, Etude QSAR de l'activation anti-oxydante d'une série de composés phénolique, TLEMCEM, Université ABOU-BEKER BELAID, 2018, 37,38,39 de 94 pages.

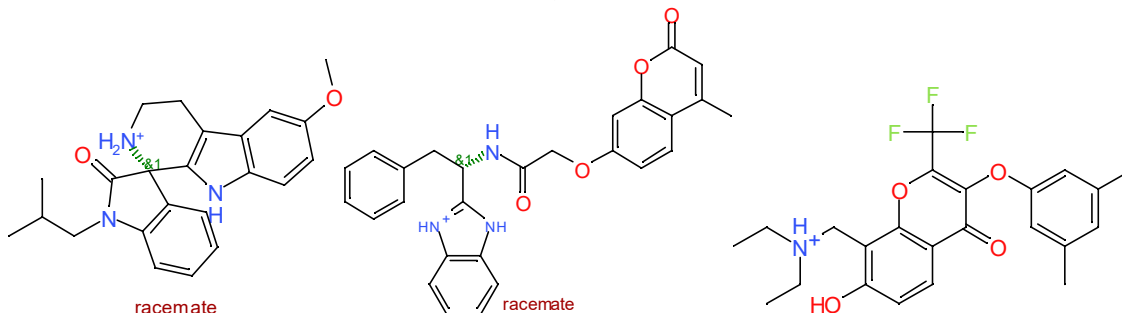
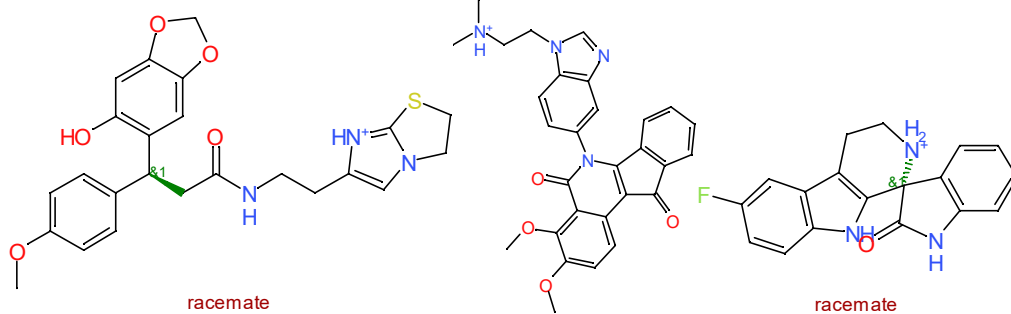
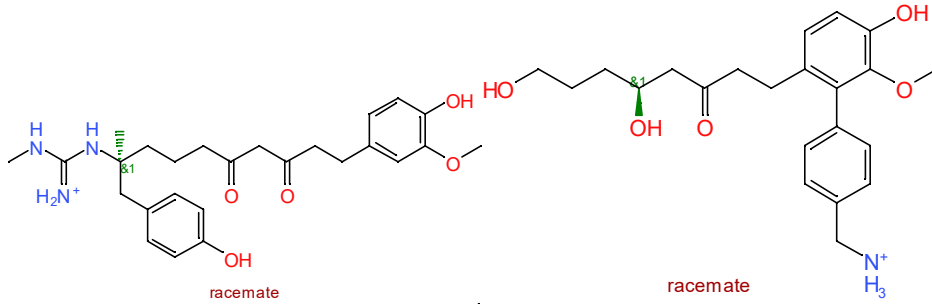
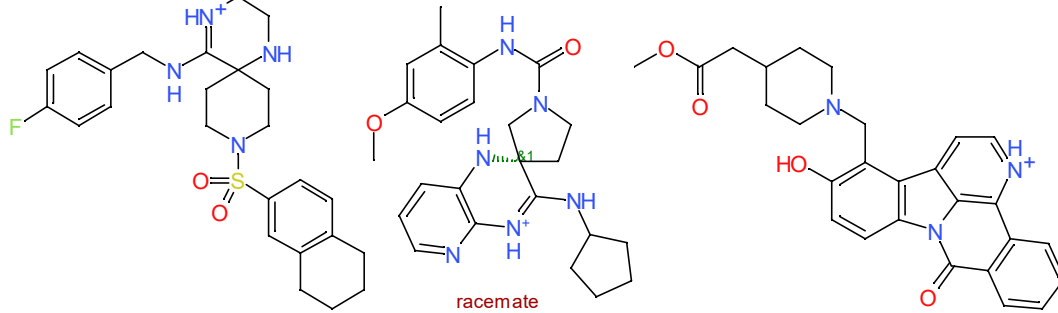
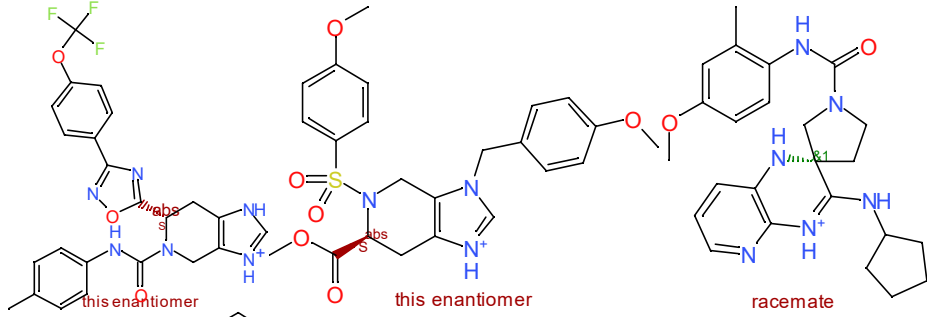
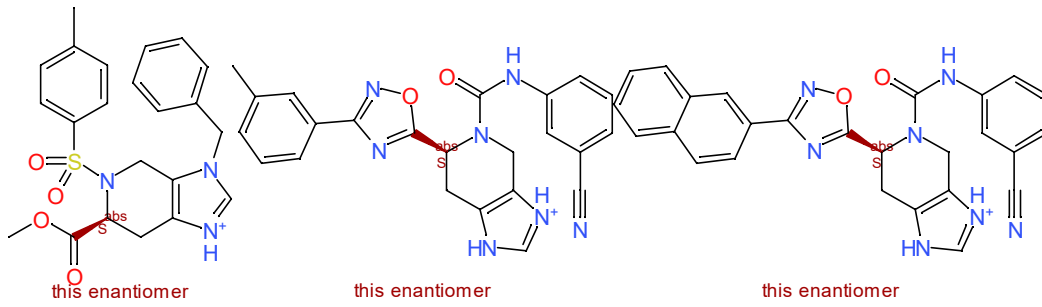
[17] **KEBAIRI MAROUA, OMRI MAROUA**, Modélisation par les réseaux de neurones artificiels : application QSAR, BISKRA, Université MOHAMED KHIDER, 2020-2021, page 30,31,33 de 109 page.

**ANNEXE 01.** Les structures chimiques tirées de la base de ZINC et identifiées par le classifieur XG-BOOST.

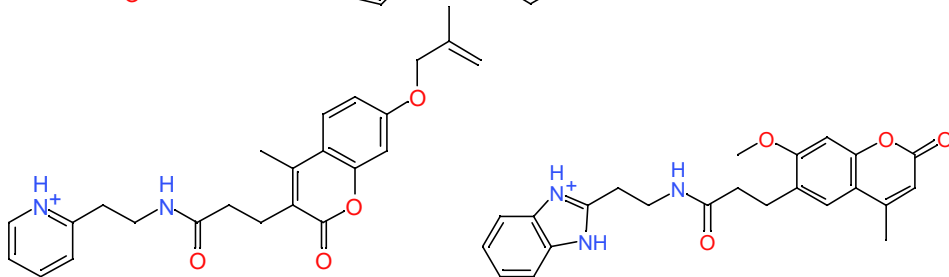
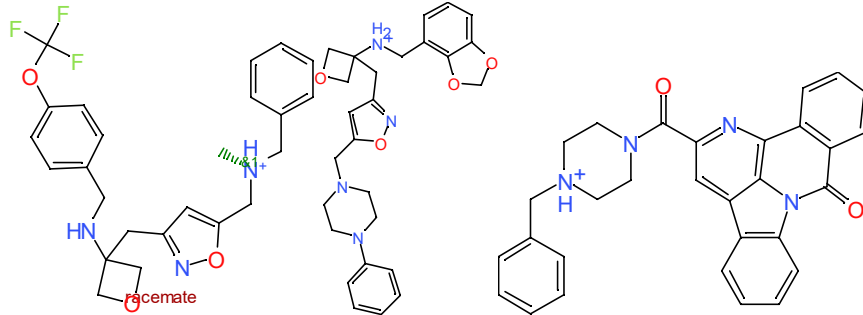
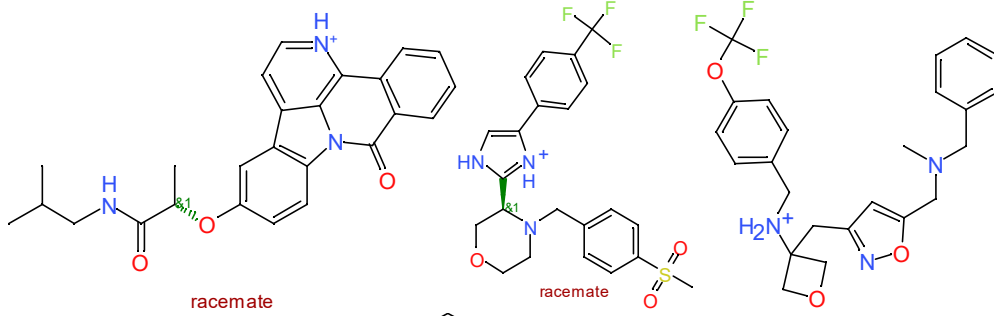
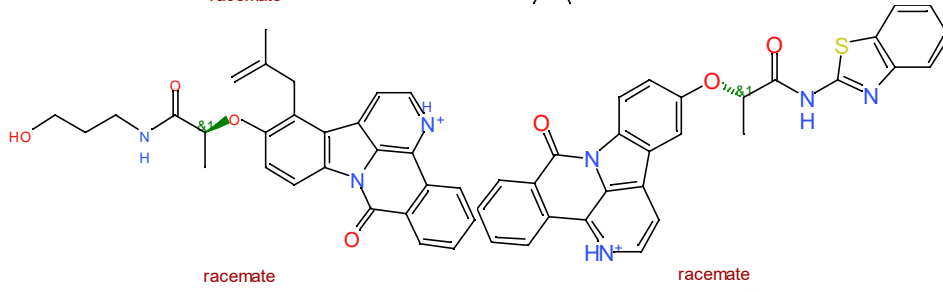
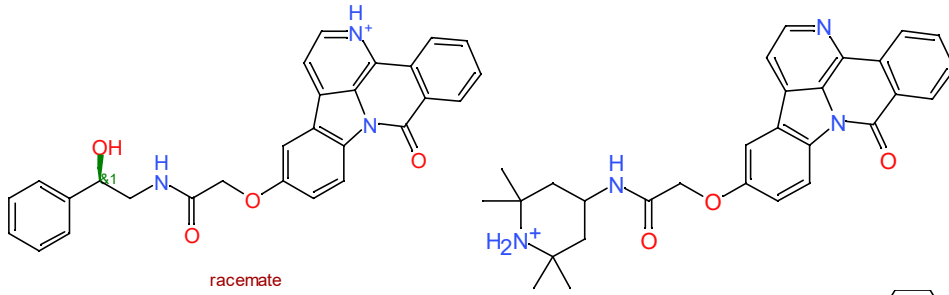
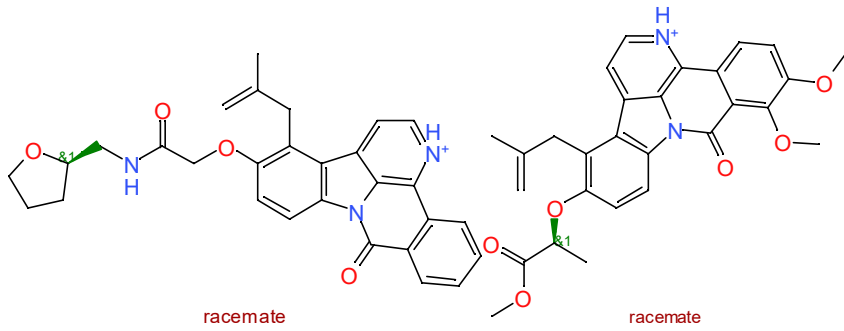


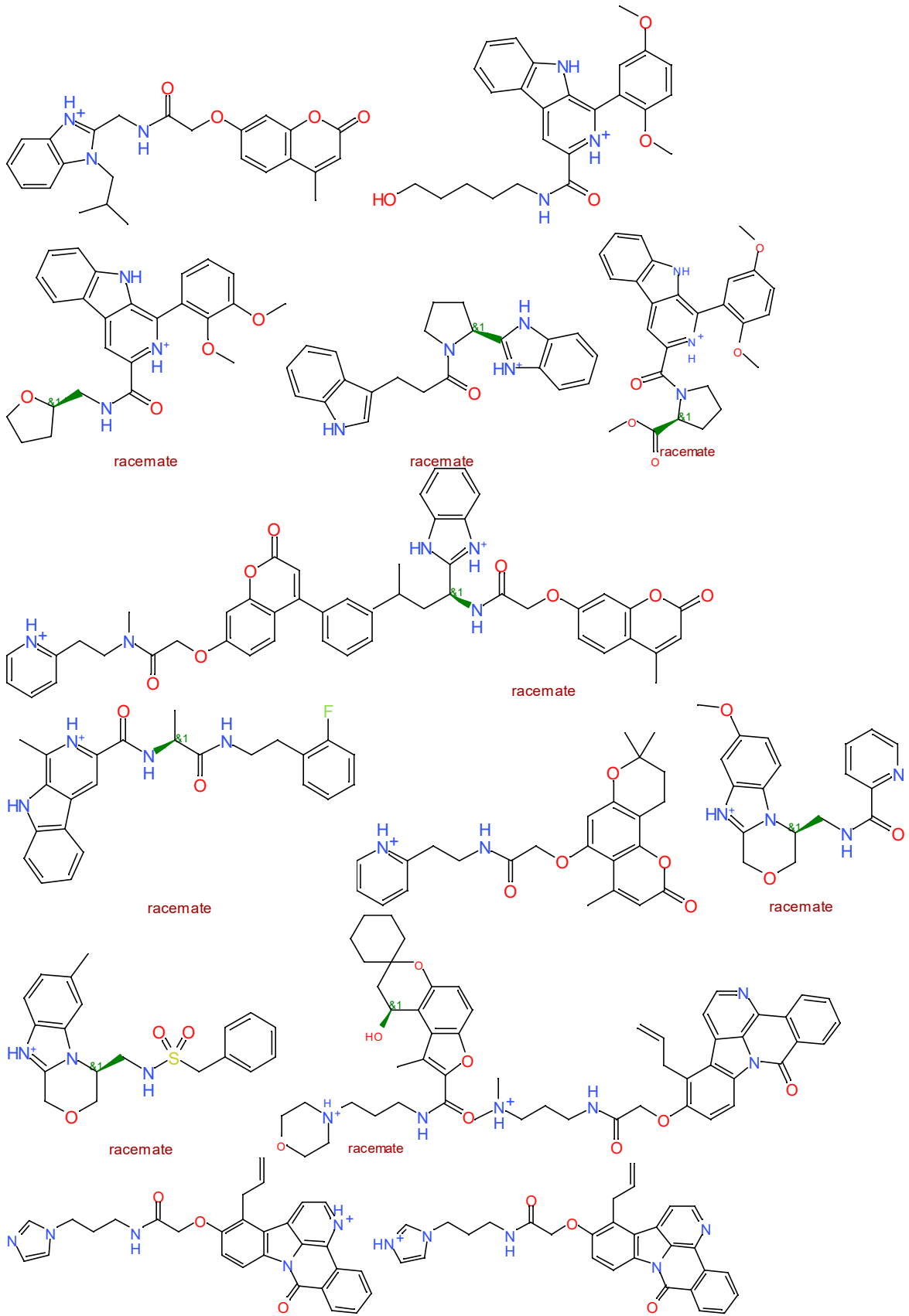




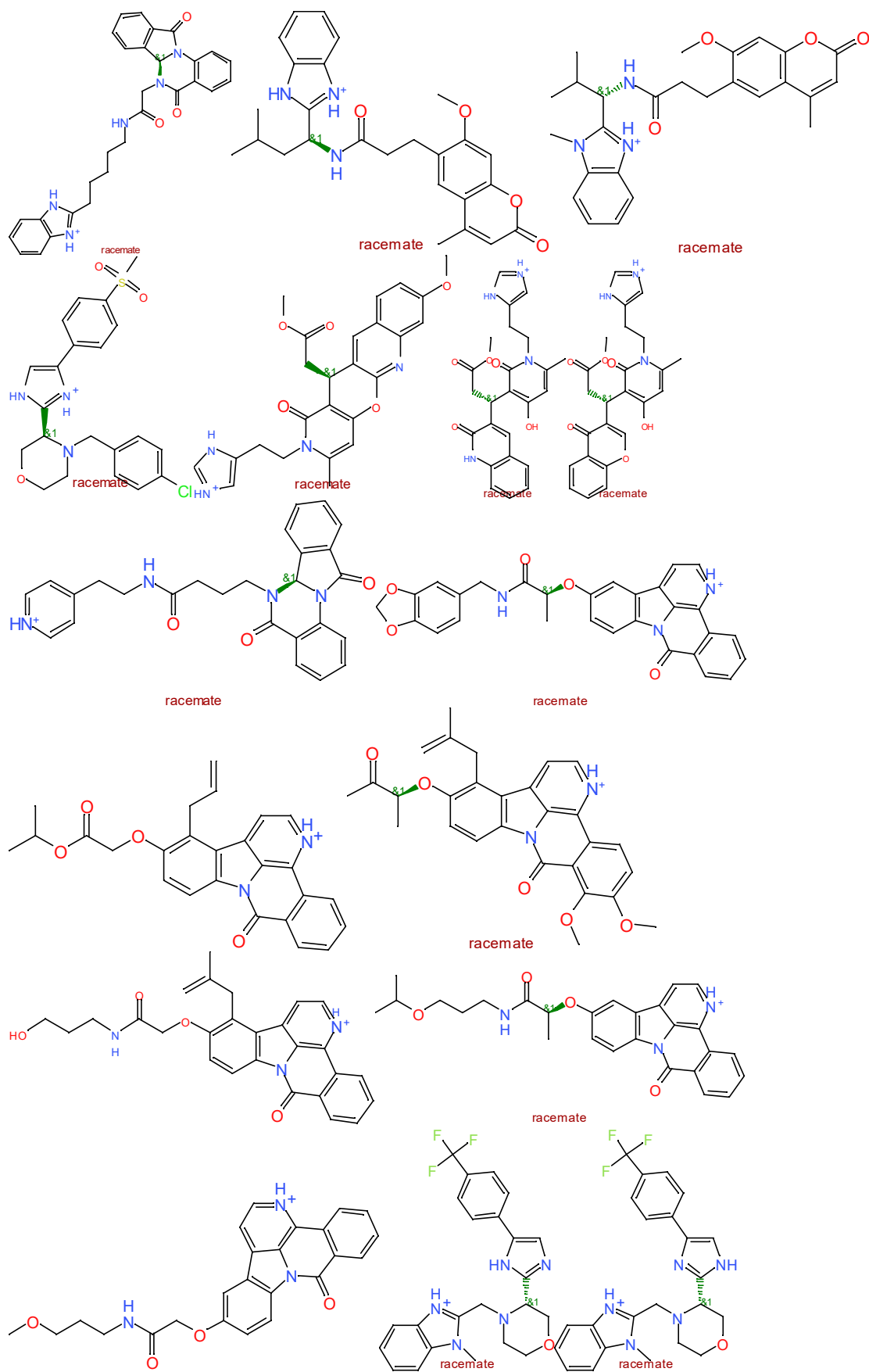


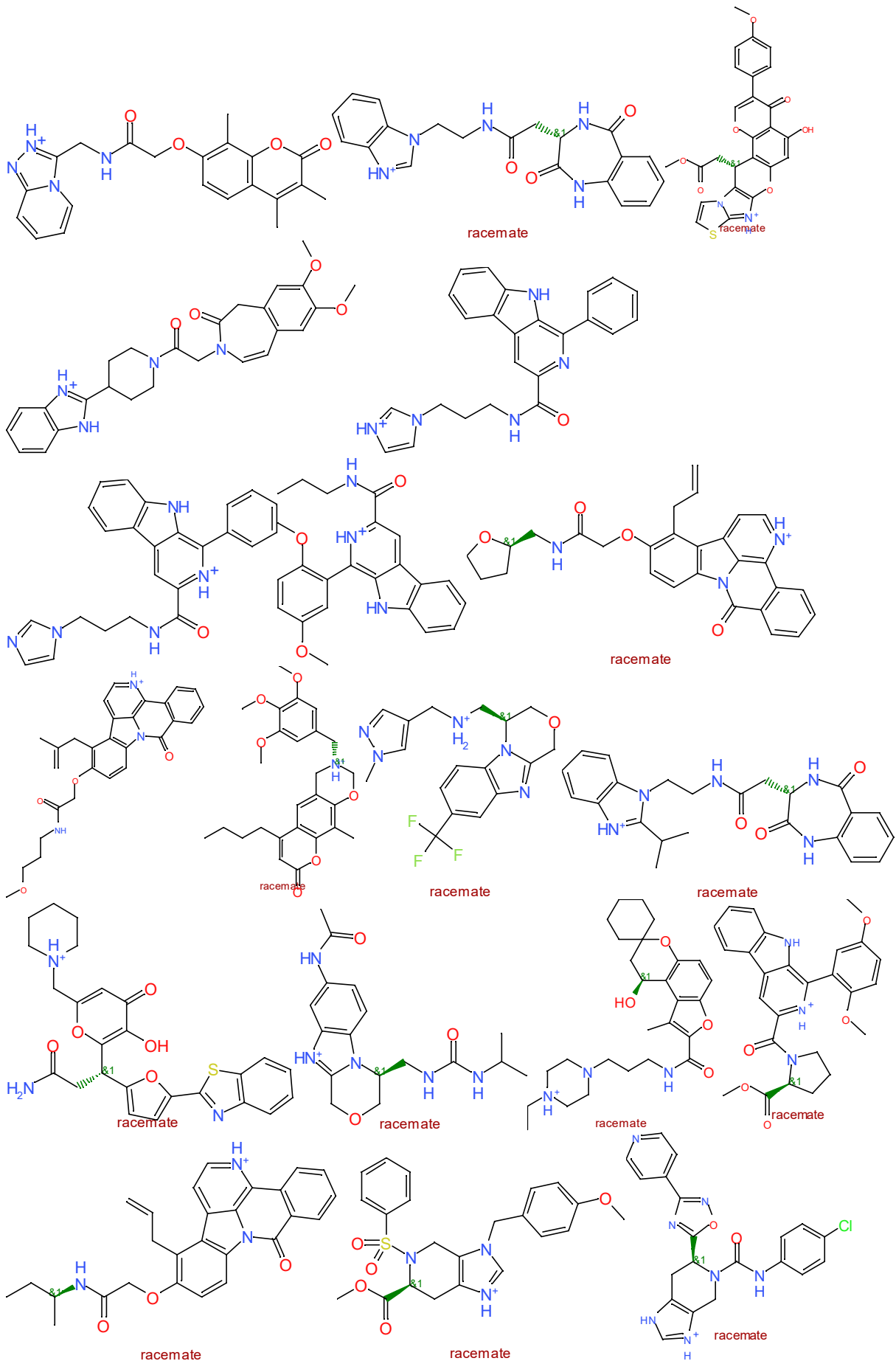


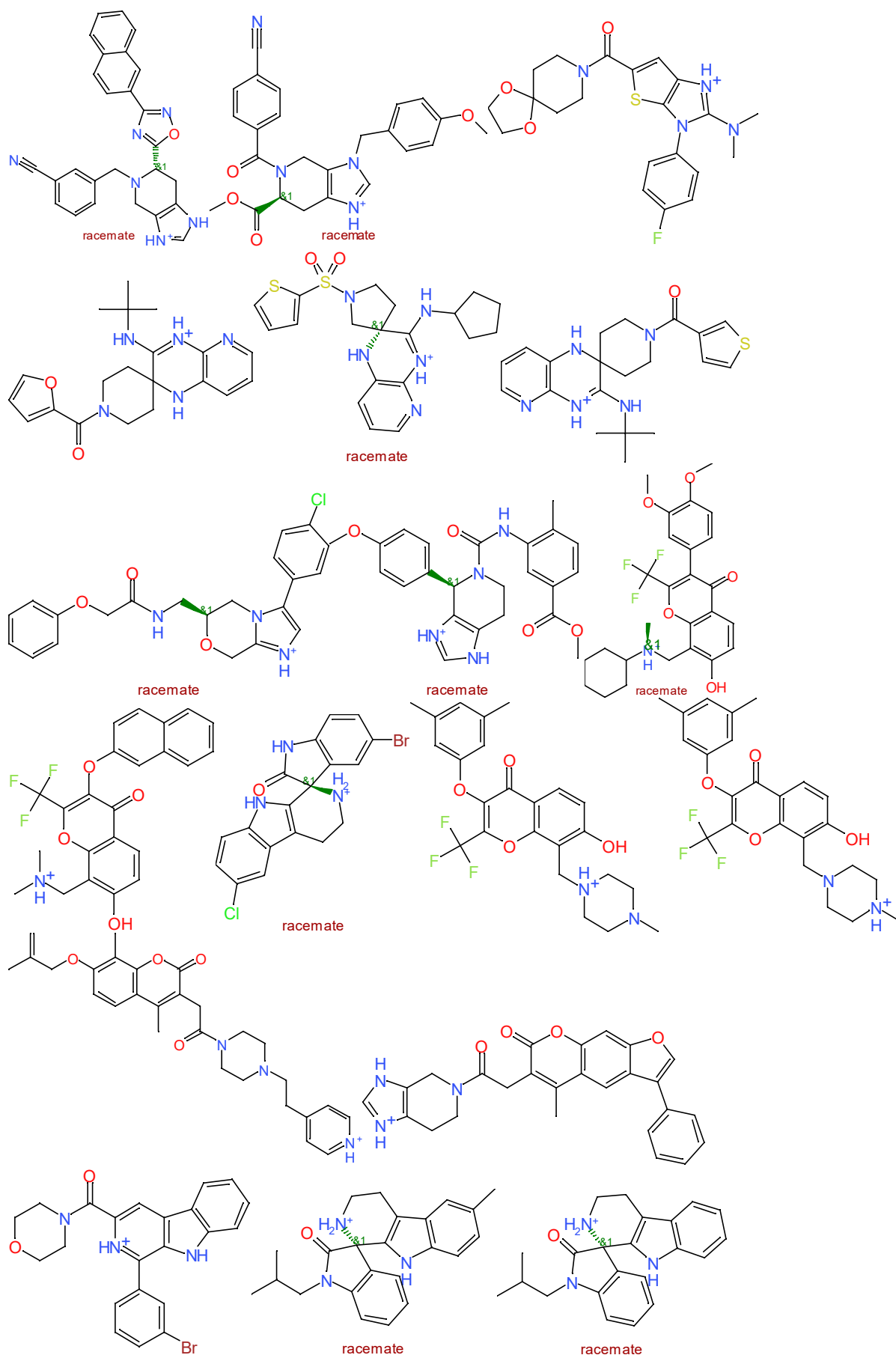


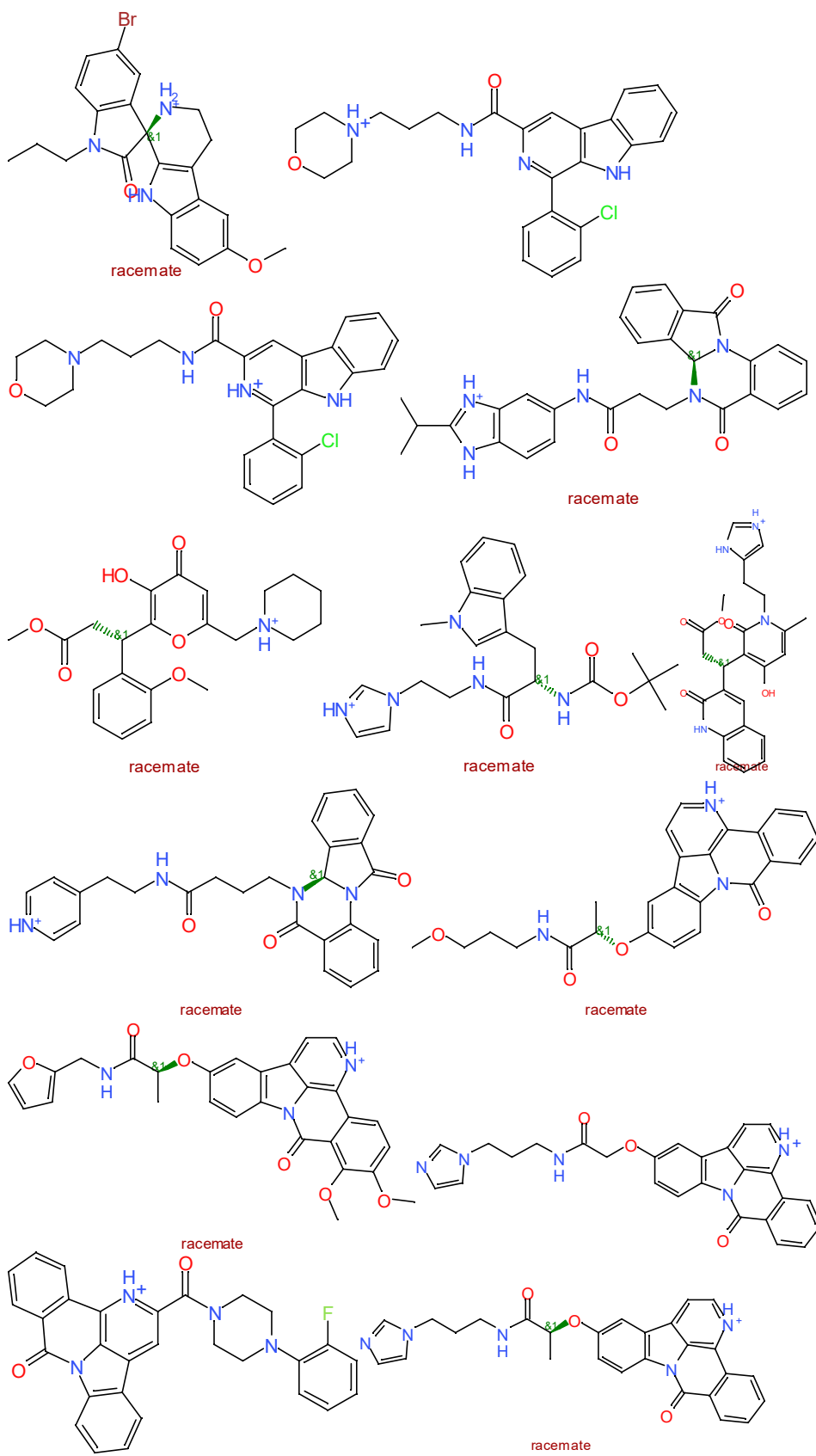


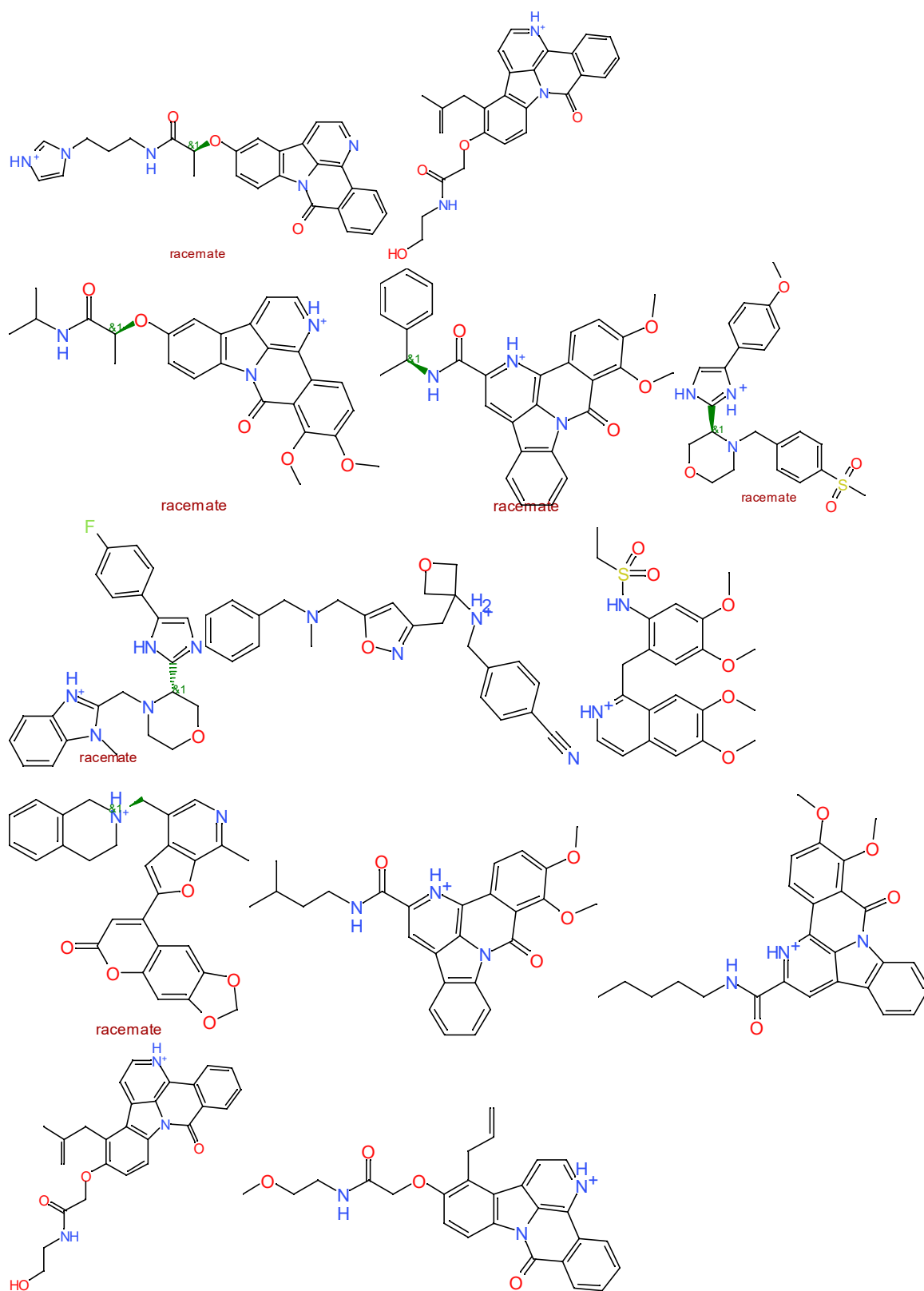


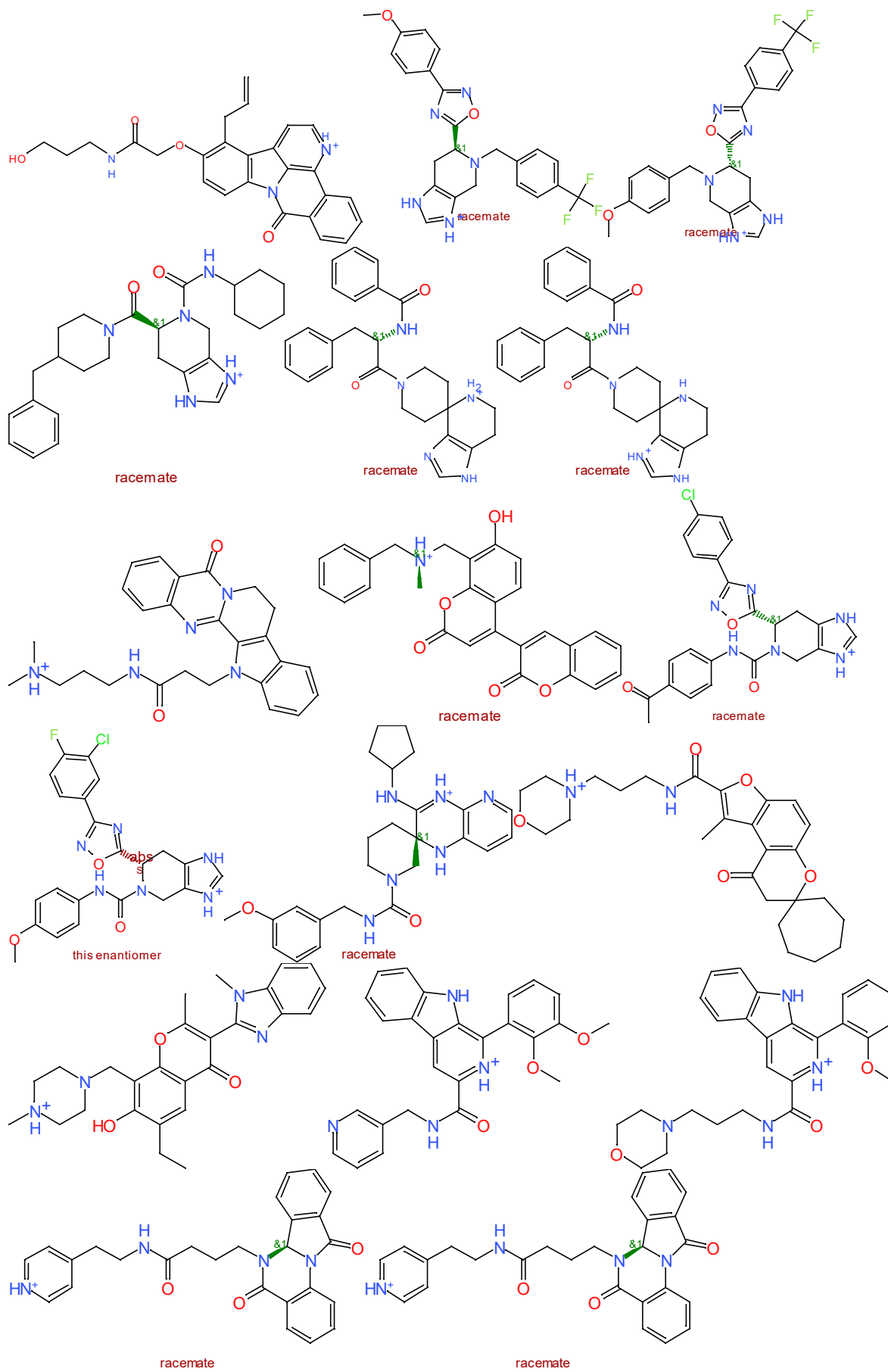




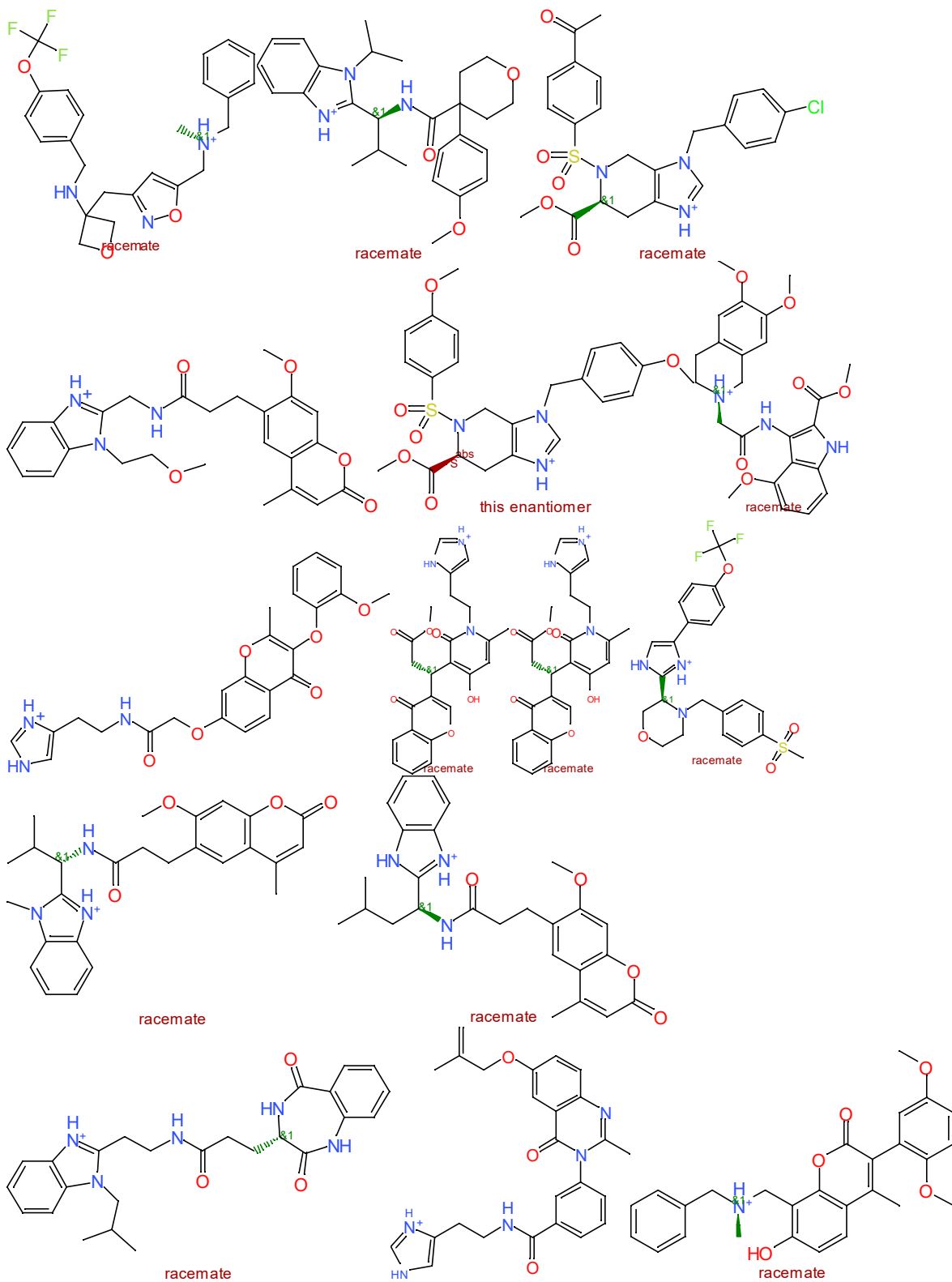


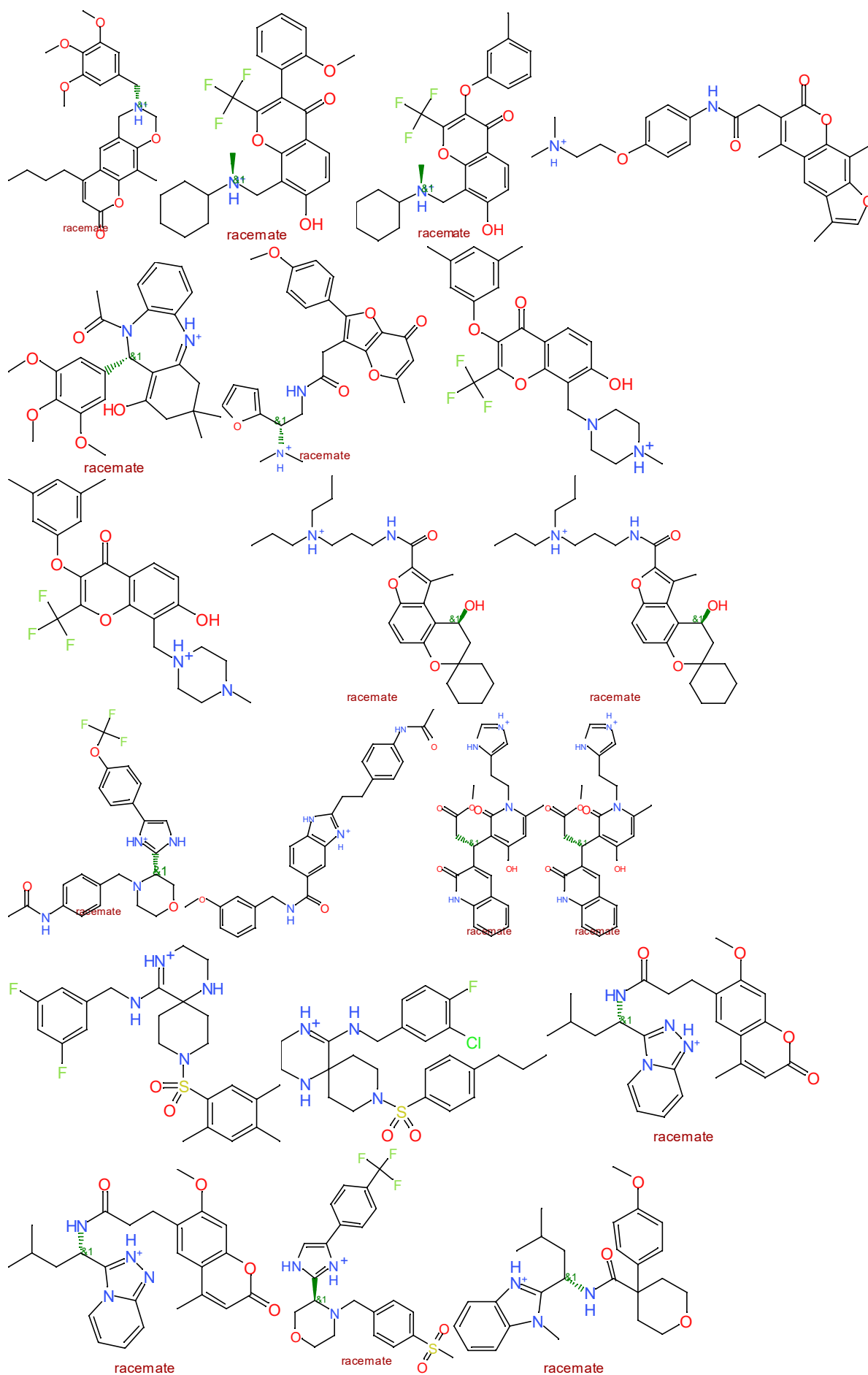


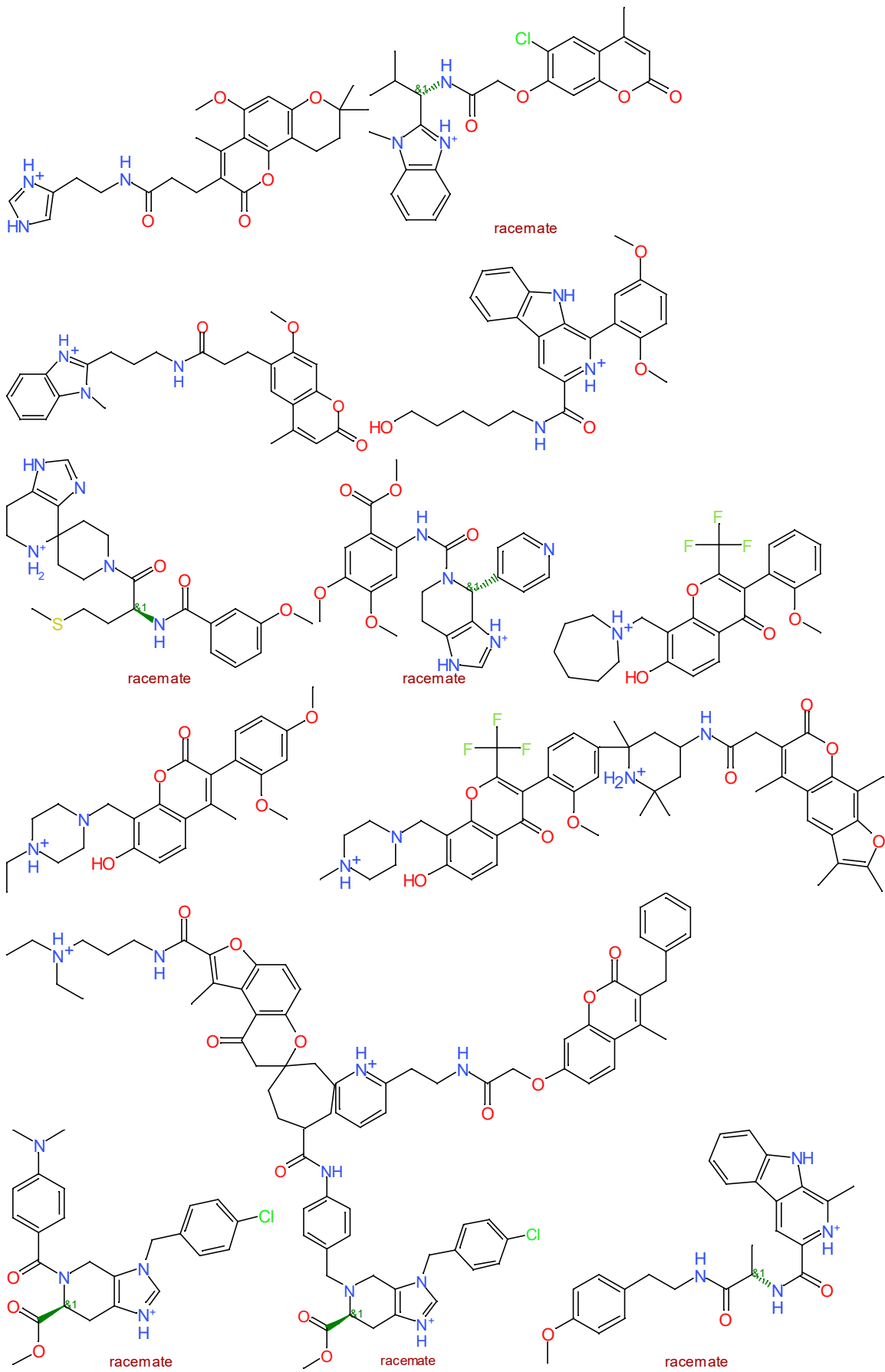


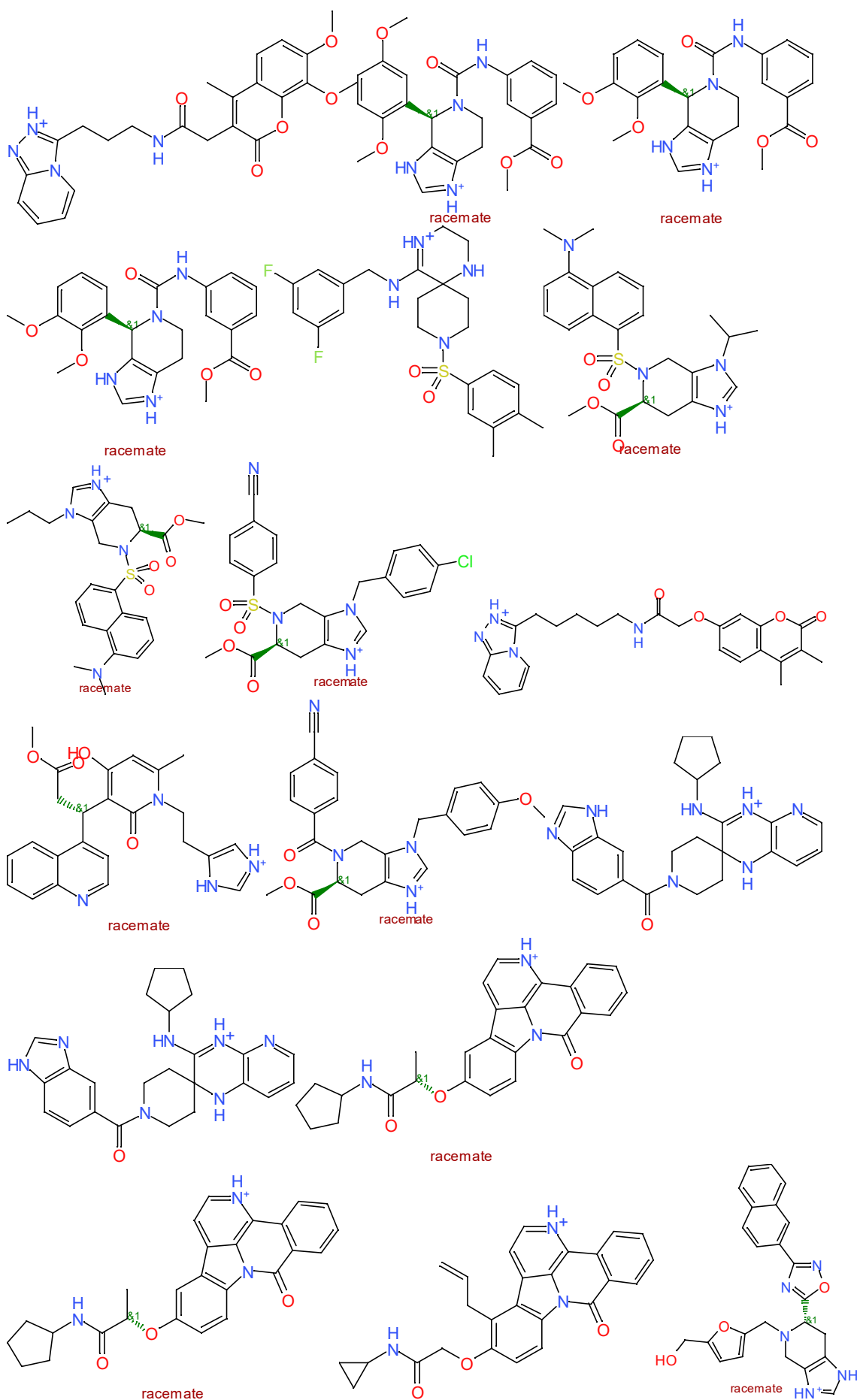


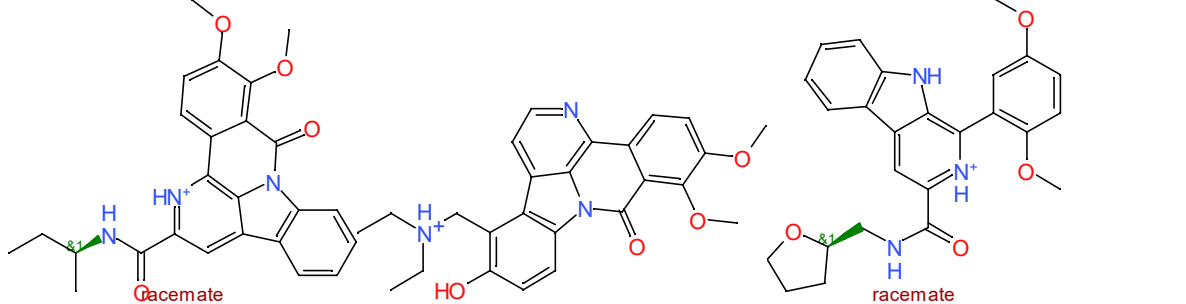
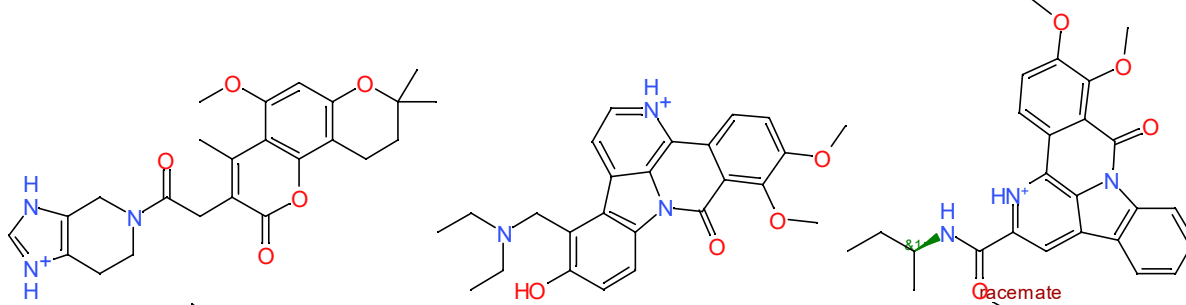
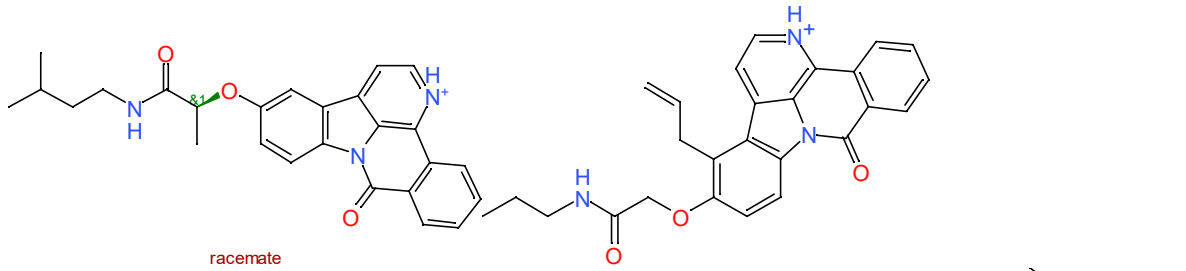
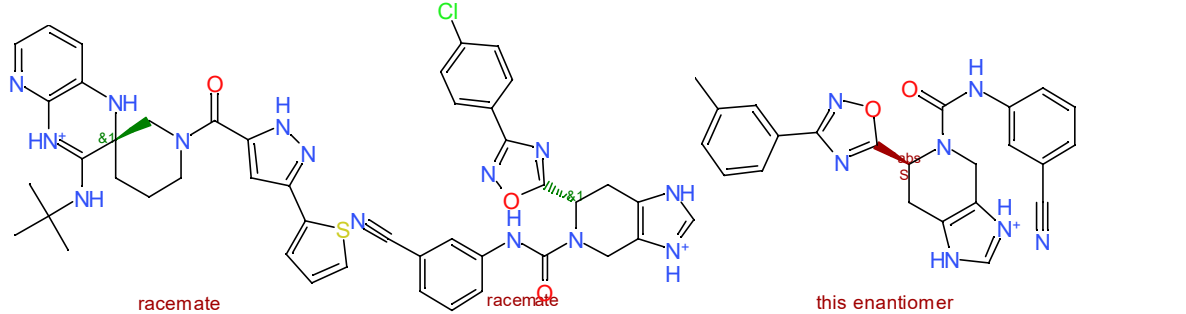
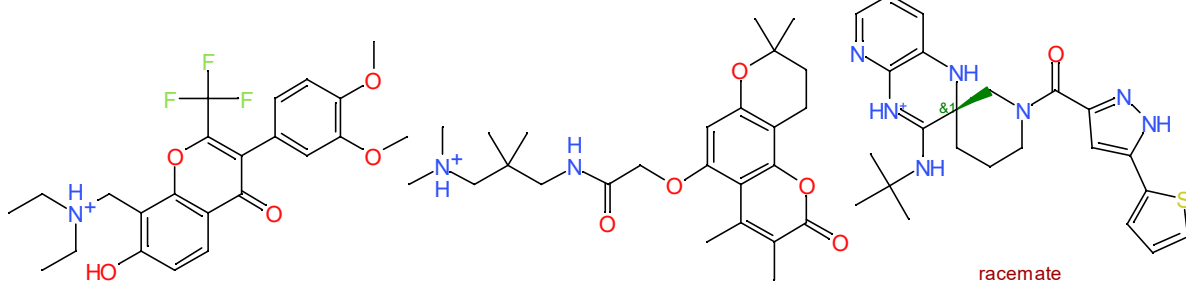
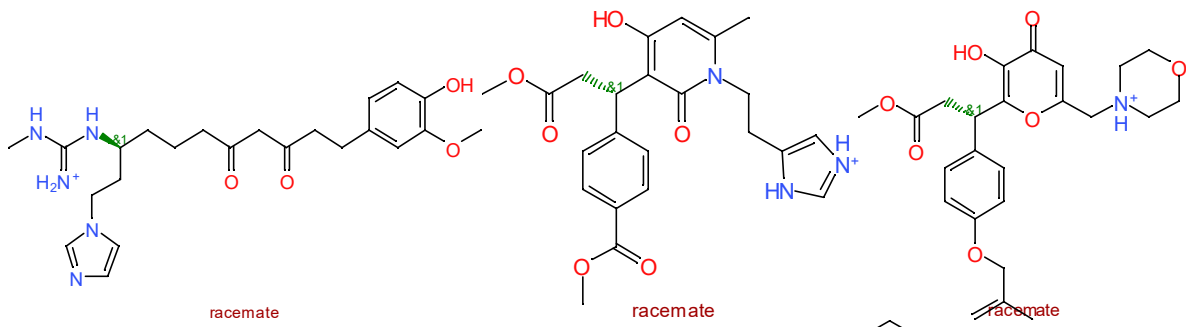


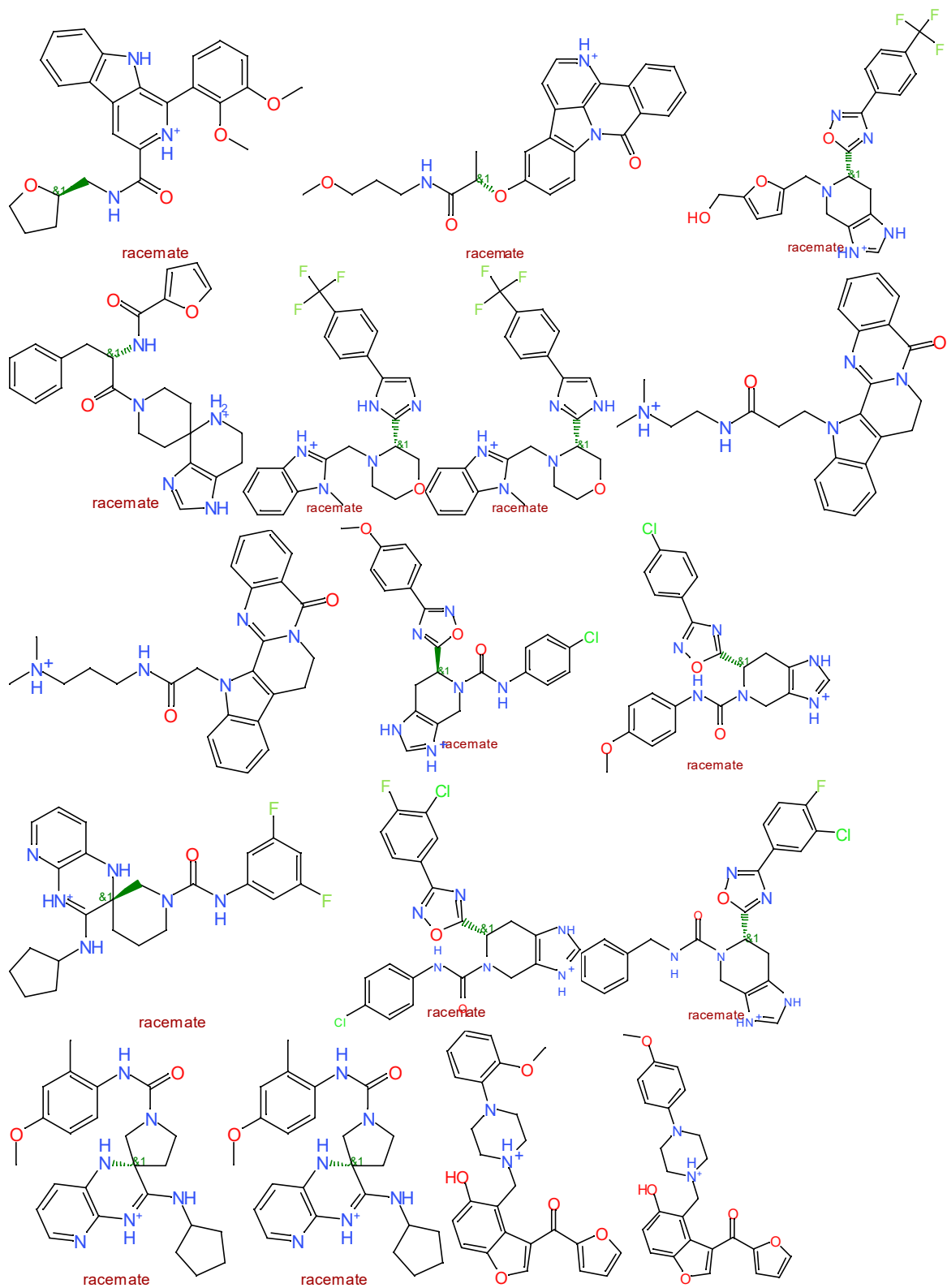




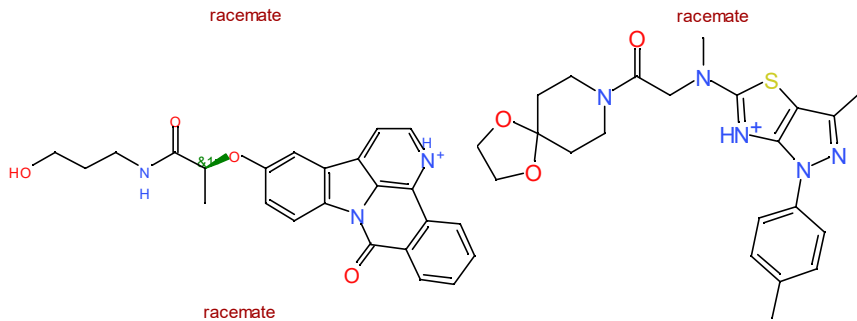
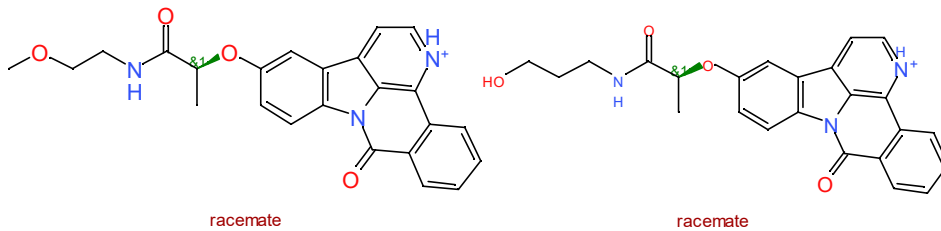
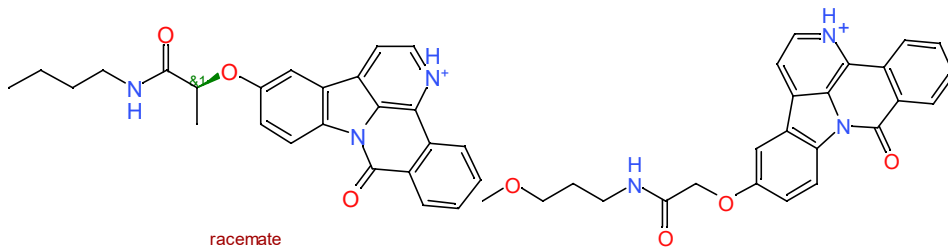
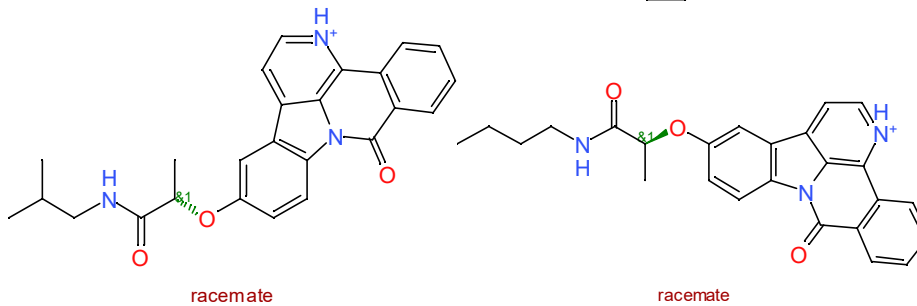
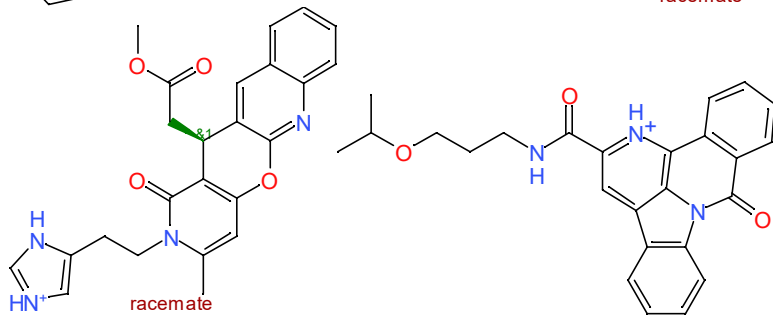
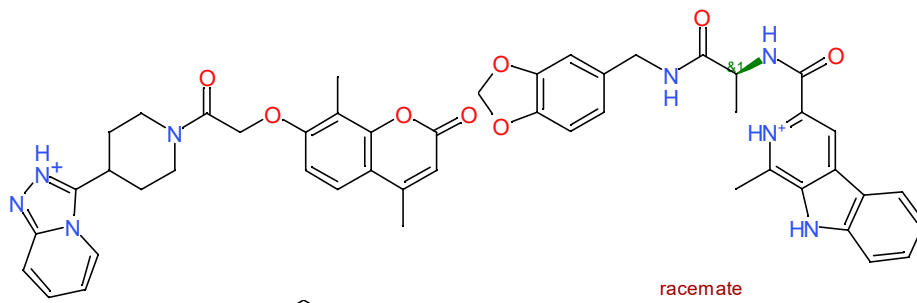


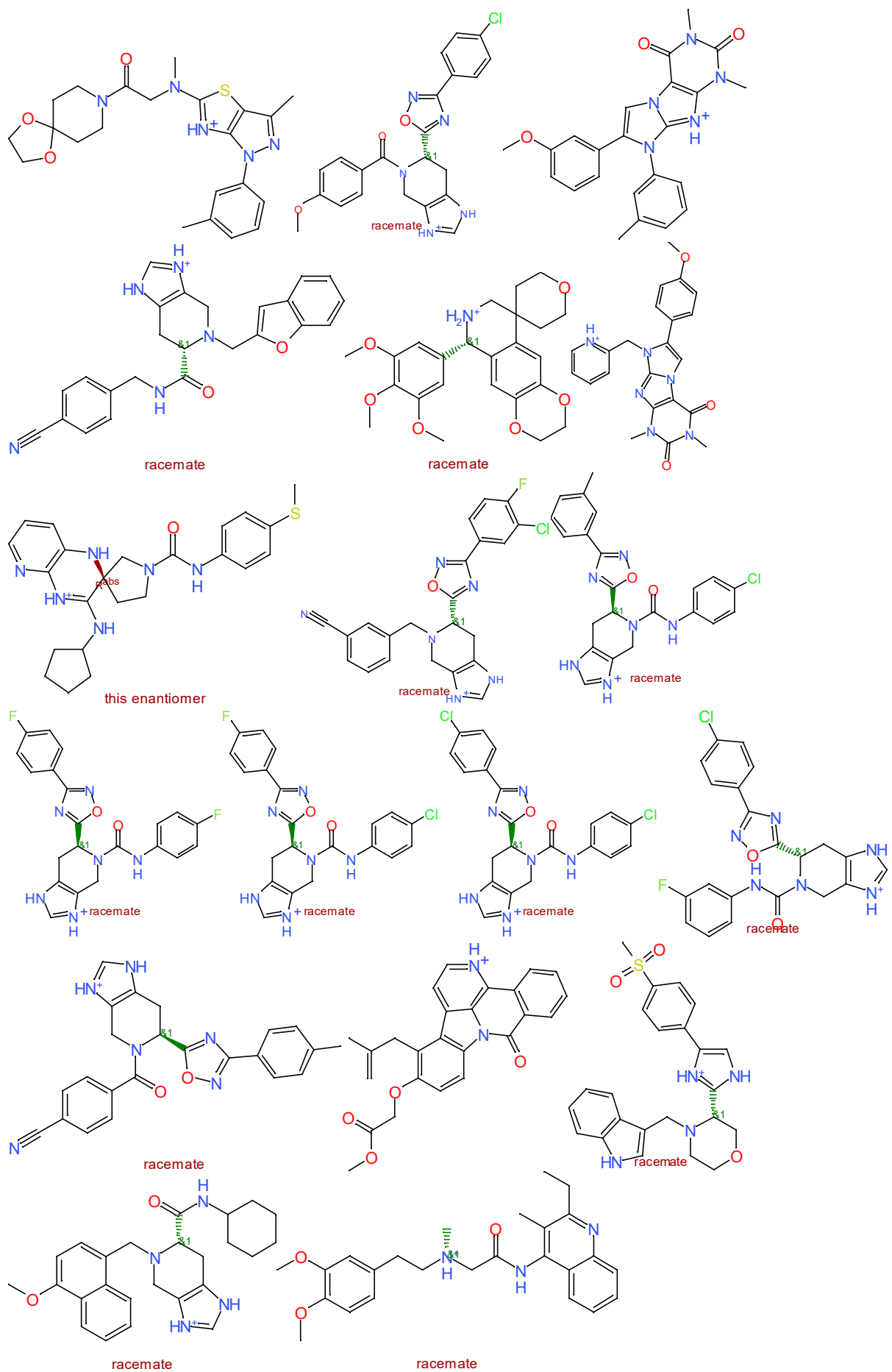




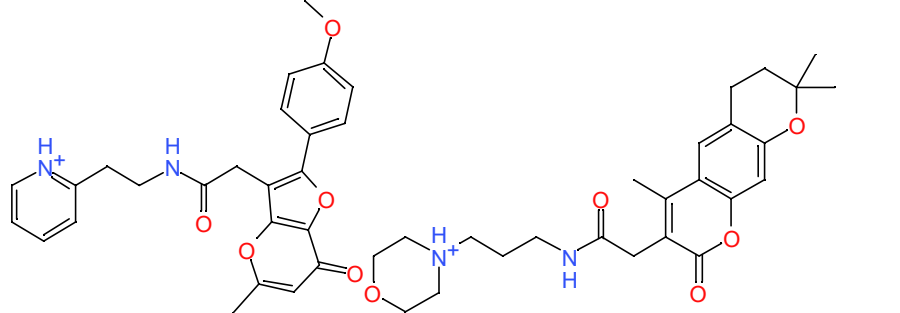
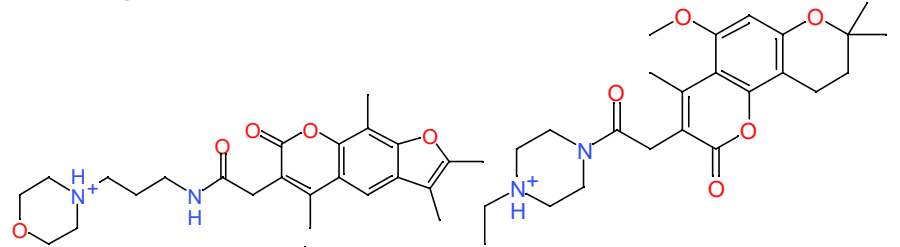
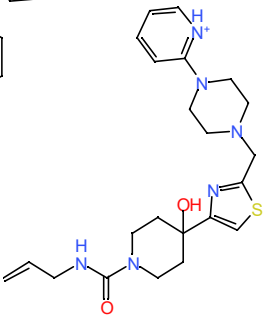
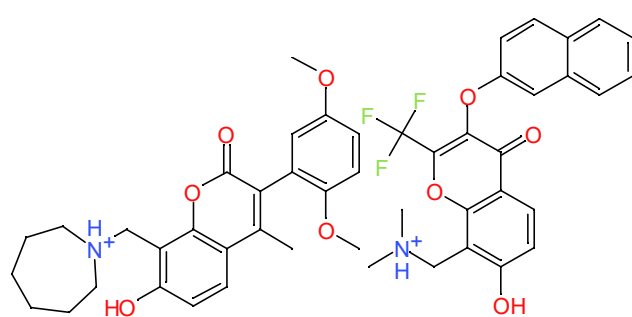
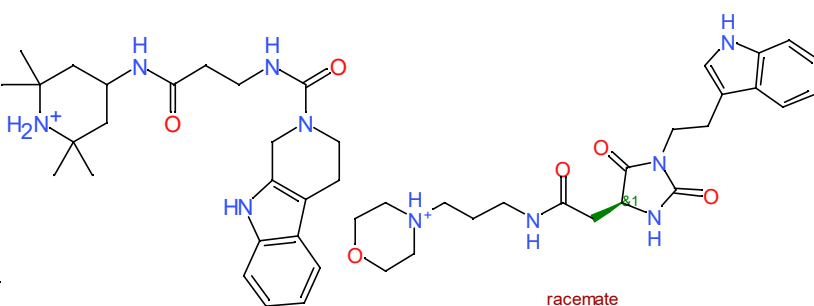
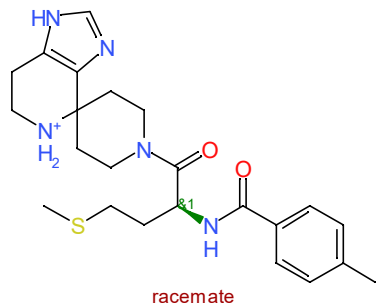
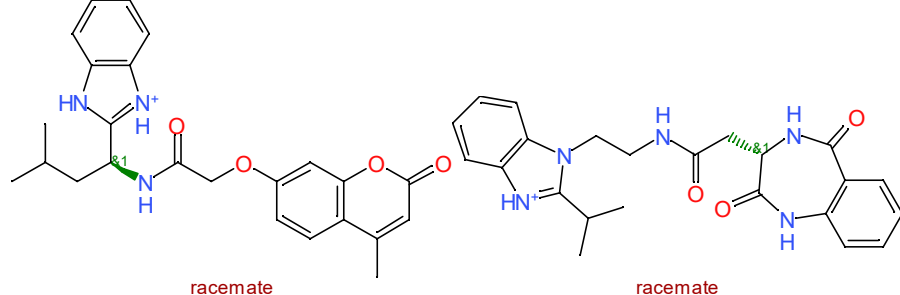
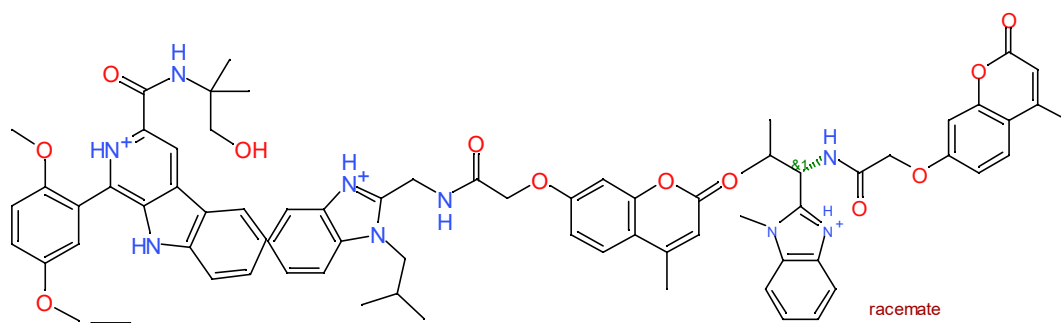






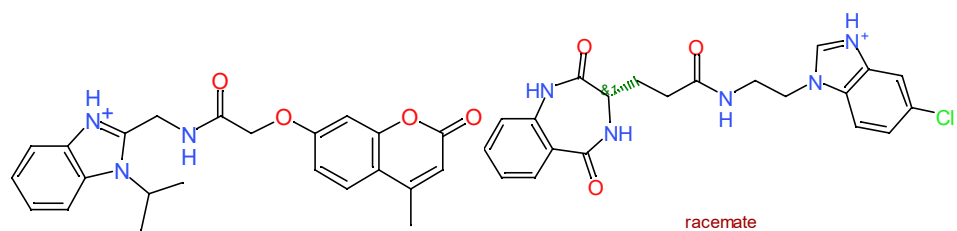












racemate

