



No. 14-025369- -00004-0000

Fecha: 2015-06-04 15:59:43 Dep: 2020 DIR NUEVASCR
Tra: 11 PATENTEIYII Eve: 378 FASENACIONALI
Act: 523 RESPUESTA Folios: 30

Señor Director
DIRECCIÓN DE NUEVAS CREACIONES
Superintendencia de Industria y Comercio
E. S. D.

Trámite : 011
Evento : 378
Actuación : 523

04

Referencia : Expediente No. **14.025.369**
Asunto : Solicitud de patente para: **"PESTICIDAS DERIVADOS DE
FENIL-TRIAZOLIL-FENIL"**
Solicitante : DOW AGROSCIENCES LLC
Fecha de solicitud : 06 de Febrero de 2014

1. Yo, Luz Clemencia DE PÁEZ, en mi condición de apoderada de la sociedad solicitante de la patente de la referencia, me refiero al oficio No. **2408** notificado por fijación en lista el 10 de Marzo de 2015.

2. Mediante oficio No. **2408** se puso en conocimiento del solicitante el informe que contiene las observaciones hechas por su Despacho sobre la presente solicitud de patente. Las observaciones realizadas en el oficio se refieren a la claridad del título, a la presencia de reivindicaciones referentes a un uso o segundo uso, a la presencia de excepciones de patentabilidad, a la claridad del pliego reivindicatorio, a la falta de unidad de invención, y al nivel inventivo de las reivindicaciones 1 a 9 en vista de los documentos D1: WO2011017513, D2: US4017529 y D3: US2009209476.

3. Respuesta a las objeciones

De este modo y con el objeto de atender las objeciones arriba mencionadas, se desea presentar las siguientes modificaciones, aclaraciones y argumentos:

"EXCELENCIA LEGAL EN UN MUNDO SIN FRONTERAS"

BOGOTÁ - Edificio Siski - Carrera 4 No. 72 - 35 - Tel: (57-1) 347 3611 - Fax: (57-1) 211 8650 - Fax in U.S.A. (1-305) 675 7743 - COLOMBIA
MEDELLÍN - San Fernando Plaza - Carrera 43 A No. 1 - 50 - Torre Protección Oficina 852 - Tel: (57-4) 520 4760 / (57-4) 326 0769 / (57) 317 365 8663 / (57) 318 532 2365- COLOMBIA

3.1. Modificación de una solicitud en trámite

De acuerdo con el artículo 34 de la Decisión 486 de la Comisión de la Comunidad Andina, donde se establece que la modificación de una solicitud es posible en cualquier momento del trámite mientras no implique una ampliación de la solicitud inicial, me permito presentar un nuevo pliego reivindicatorio conformado por 2 reivindicaciones.

Con el fin de atender a la objeción hecha por el señor examinador, se solicita el cambio de título de la presente solicitud de patente a:

“PESTICIDAS DERIVADOS DE FENIL-TRIAZOLIL-FENIL”

Se han introducido las siguientes modificaciones al capítulo reivindicatorio:

- Las reivindicaciones 2 a 90 han sido eliminadas. En consecuencia, el nuevo pliego reivindicatorio ya no contiene materia exceptuada de patentabilidad, ni usos o segundos usos.
- El pliego reivindicatorio ha sido modificado para eliminar los términos “*such*” y “*further comprising one or more of the following*”.
- La nueva reivindicación 1 ha sido modificada para limitar la materia objeto de la invención, de manera que ahora la definición de todos los radicales pertenecientes a los compuestos de fórmula (I) ha sido restringida a una generalización razonable del tipo de compuestos que han sido sintetizados y probados, además que se encuentra debida y totalmente soportado en la descripción.
- Adicionalmente, se desea aclarar, con respecto a la objeción hecha a la expresión “*al lo menos*”, que dicha expresión se refiere al número de radicales R^1 , R^2 , o R^4 , en (K).

Con respecto a lo anterior se desea aclarar que el número específico de los elementos mencionados no está relacionado en lo absoluto con el concepto inventivo de la solicitud. Es requerida la presencia de uno o más de cada uno de estos elementos para que la invención cumpla con las características técnicas requeridas. Por lo tanto una limitación en cuanto al número de estos elementos dentro de la invención implicaría una limitación indebida con respecto a su alcance y facilitaría a terceros la no comisión de una infracción literal de la misma.

A manera de ejemplo, se podría definir un número entre 1 y **N** de radicales; un tercero podría simplemente emplear **N+1** radicales, lo que no representaría una infracción literal de las reivindicaciones con un artículo esencialmente igual. Así, resulta necesario incluir la expresión “al menos” como una manera de definir el número **N** tal que proteja el alcance de la invención y la proteja de posibles infractores.

Por último, una persona medianamente versada en la materia estará en capacidad de entender la expresión en cuestión y podrá reproducir inequívocamente la invención con la información proporcionada.

- Se ha introducido la nueva reivindicación 2 que se encuentra dirigida a un número de compuestos específicos tal como se encuentra definida.
- Finalmente, se ha modificado el pliego reivindicatorio de tal forma que ahora se encuentra dirigido al grupo inventivo 1.

Así, las reivindicaciones son ahora muy claras y precisas en el momento de señalar la naturaleza de la invención reivindicada y es evidente que las características técnicas esenciales de la invención están establecidas.

Por lo anterior, las objeciones formuladas por el Examinador en cuanto a la claridad y concisión del pliego reivindicatorio se ven superadas. En consecuencia, se solicita atentamente tener en cuenta este nuevo pliego reivindicatorio para el nuevo análisis de patentabilidad.

3.2 Argumentos a favor de la falta de Nivel Inventivo

El artículo 18 de la Decisión 486 señala lo siguiente:

“Artículo 18.- Se considerará que una invención tiene nivel inventivo, si para una persona del oficio normalmente versada en la materia técnica correspondiente, esa invención no hubiese resultado obvia ni se hubiese derivado de manera evidente del estado de la técnica.”

Una invención se considerará inventiva si la misma no hubiese resultado obvia a partir del arte previo para una persona del oficio normalmente versada en la materia. Esto

a su vez implica que el Examinador a cargo de esta evaluación debe colocarse en el momento exacto en el que el inventor buscaba resolver el problema técnico contemplado en la solicitud y para el cual poseía la literatura y el conocimiento disponibles en dicho momento.

Siguiendo entonces los lineamientos de la Jurisprudencia aplicable al análisis del nivel inventivo de una solicitud de patente, se considera que una anterioridad o una combinación de dos anterioridades puede afectar dicho requerimiento de patentabilidad, si una persona medianamente versada en la materia encuentra enseñanzas o sugerencias claras que, a partir de un conocimiento medio en dicho campo, le permitan llegar de manera obvia a la invención reivindicada. Adicionalmente se considera que una mejora sustancial, un efecto inesperado o una ventaja con respecto al arte previo es un indicio de nivel inventivo, por cuanto aporta un salto tecnológico en la regla técnica.

Con base en lo anterior, respetuosamente manifestamos nuestro desacuerdo con el análisis llevado a cabo por el Examinador y con las conclusiones del mismo y solicitamos atentamente la reconsideración de las razones por las cuales se objeta el nivel inventivo, con base en la siguiente información:

A este respecto, el señor Examinador elevó una objeción hacia la reivindicación 1 en vista de la combinación de los documentos D1 y D2, y D2 y D3. Dada esta objeción, amablemente mi representada desea poner de presente la presente Acción Oficial No. 2408 comprende análisis retrospectivo. De acuerdo al artículo adjuntado a este documento, el Dr. Ertl señala lo siguiente: "Se encontraron 849 574 sustituyentes únicos. Extrapolado al número de moléculas orgánicas conocidas, este resultado sugiere que alrededor de 3,1 millones sustituyentes son conocidos".

Así las cosas, existen varios lugares en el documento D1 para modificar las moléculas de la misma, y por tanto de acuerdo con el Dr. Ertl, existen alrededor de 3,1 millones de sustituyentes, entonces ¿cómo una persona del oficio normalmente versada en la materia, sabe qué sustituyente debería recoger y saber dónde ponerlo sin la contribución inventiva realizada por los inventores de la presente solicitud? En consecuencia, la combinación de los documentos D1 con D2 o D2 con D3, no son más que la reconstrucción retrospectiva de las moléculas inventivas.

Por tanto, respecto a la afirmación del Examinador, en cuanto a la obviedad de la presente invención en vista de lo divulgado por la combinación de los documentos D1 y D2 o D2 y D3, manifestamos nuevamente nuestro desacuerdo y concluimos que el análisis realizado en la acción oficial No. 2408, se basó en un análisis retrospectivo o también conocido como "hindsight", el cual lo define el INSTRUCTIVO DE EXÁMENES DE SOLICITUD DE PATENTES DE INVENCIÓN Y MODELOS DE UTILIDAD, de la siguiente forma:

Análisis ex post facto o retrospectivo o hindsight

Es importante tener en cuenta que una invención reivindicada que, a primera vista, parece evidente, puede ser inventiva. Una vez formulada una nueva idea, a menudo se puede demostrar teóricamente el modo en el que se podría llegar a ella, a partir de algo conocido, a través de una serie de etapas aparentemente fáciles. El examinador de patentes debe evitar el uso de un análisis ex post facto de este tipo.

Es decir, el estado de la técnica se debe examinar sin el beneficio del conocimiento ofrecido retrospectivamente por la invención reivindicada. La indicación o la sugerencia que permita llegar a la invención reivindicada deben proceder del estado de la técnica o de los conocimientos generales de la persona del oficio normalmente versada en la materia, y no basarse en la divulgación del solicitante.

Por lo tanto, un técnico con conocimientos medios en la materia no estaba en condiciones de plantearse el problema, tampoco de resolverlo en la forma en que se reivindica y tampoco de prever el resultado, por lo que la presente solicitud posee la altura inventiva para que se le conceda el privilegio solicitado.

En consecuencia de lo anterior, la invención de la presente solicitud tiene nivel inventivo en vista de los documentos D1, D2 y D3 del estado de la técnica citados por el Examinador.

En el evento que su Despacho no esté de acuerdo con los anteriores planteamientos, respetuosamente solicitamos que se indique por qué los documentos citados hubieran sido relevantes para prever la solución al problema técnico acá sugerida, cuál es el documento más cercano del estado del arte y cuál era, a la fecha de la prioridad, las enseñanzas globales del estado del arte frente a las cuales se debe evaluar la presente invención, para efecto de evitar un análisis en retrospectiva ("hindsight" ó "ex post facto") como el que se ha hecho en el presente caso.

La respuesta a cada una de las anteriores preguntas es fundamental para efecto de poder controvertir las objeciones de manera más adecuada a la práctica de patentes y para controvertir la decisión que se adopte ante la autoridad jurisdiccional, llegado el caso.

3.3 Conclusión

Por todo lo anterior, el documento D1, el documento D2, el documento D3, o su combinación, no resultan anterioridades válidas capaces de afectar el nivel inventivo de la invención de la presente solicitud. Por lo tanto, consideramos que el Examinador está excluyendo el hecho de que el inventor en el momento de solucionar el problema técnico de la solicitud tuvo que emplear habilidad inventiva, pues la presente invención no resulta de la mera extrapolación de las enseñanzas del arte previo, como se afirma. Se está desconociendo el mérito investigativo empleado y que está aportando una ventaja al arte previo.

4. Puesto que se ha dado respuesta a todas las objeciones formuladas por el Examinador, se solicita atentamente proceder al otorgamiento de la solicitud de patente. En este sentido, respetuosamente recordamos al Examinador que la patente no podrá ser negada con base en objeciones diferentes a las expuestas en la acción oficial No. 2408, toda vez que al no haber sido trasladadas a mi representada, dicha decisión violaría tanto el tenor literal del artículo 45 de la Decisión 486 como el derecho constitucional al debido proceso. En este evento, y con base en el inciso 2 del artículo 45, respetuosamente solicitamos al Examinador expedir una nueva acción oficial que permita a mi representada contra argumentar las mismas o modificar la solicitud con el fin de cumplir los requisitos de patentabilidad.

De otro lado, respetuosamente manifiesto que a pesar de eliminar materia del pliego reivindicatorio originalmente presentado, mi representada no renuncia a la protección que pueda obtener sobre dicha materia y se reserva el derecho de presentar una solicitud divisional que la contenga durante el trámite de la presente solicitud, beneficiándose de la fecha de prioridad inicialmente reivindicada.

5. Anexos

- Nuevo pliego reivindicatorio conformado por 2 reivindicaciones
- Artículo Dr. Ertl

Señor Director,


LUZ CLEMENCIA DE PÁEZ

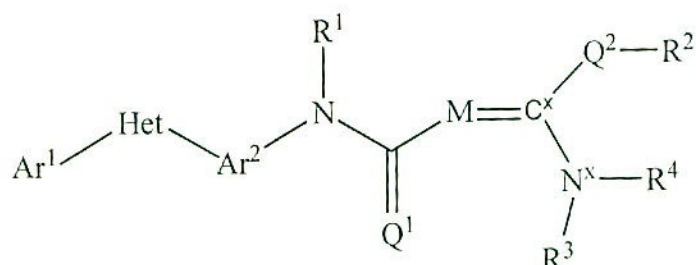
C.C. 35.456.344 de Usaquén

T.P.A. No. 23.555

man COLO-10975-0841-0204-2
SGD\SGD\ID-280498\WPC10975
No. M-2015-280498\SGD

REIVINDICACIONES

1. Una composición que comprende una molécula que posee la siguiente fórmula:



Fórmula Uno

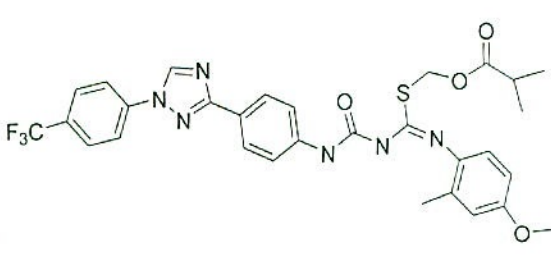
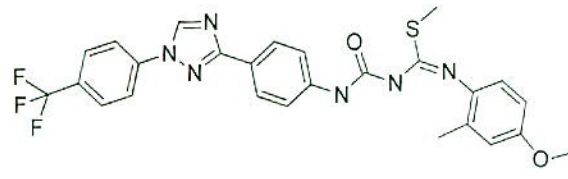
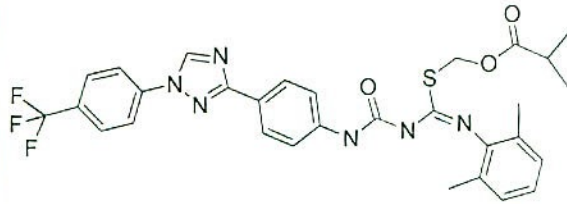
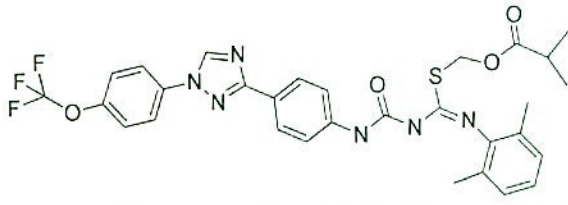
en donde:

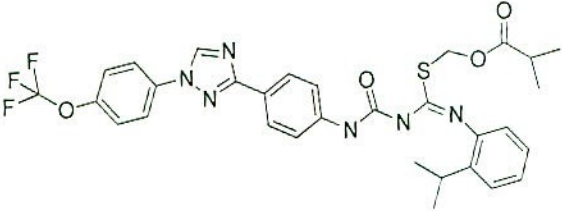
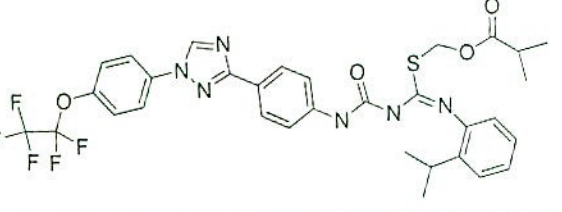
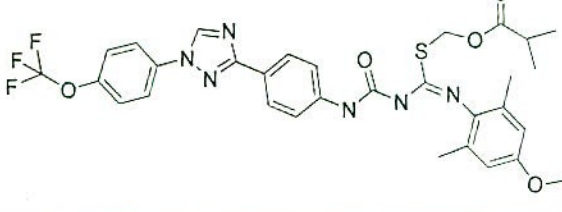
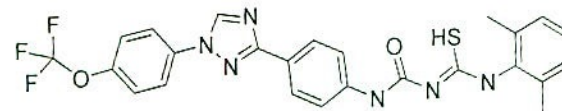
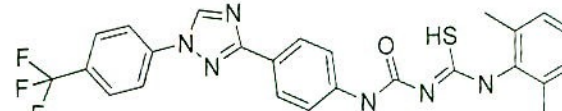
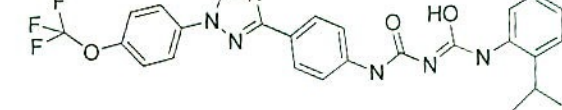
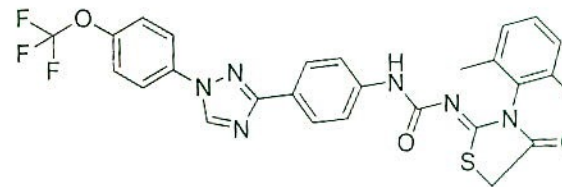
en la que:

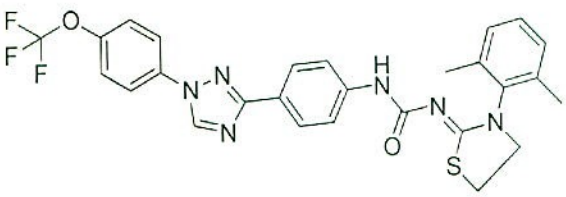
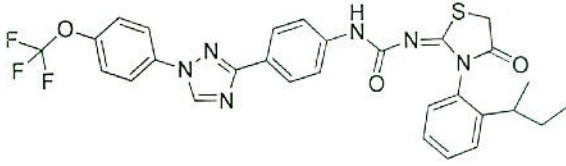
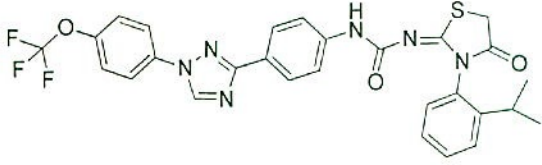
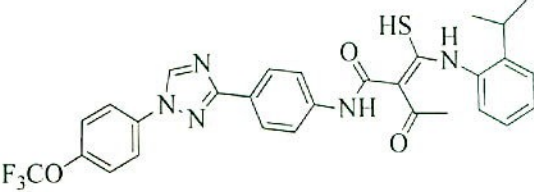
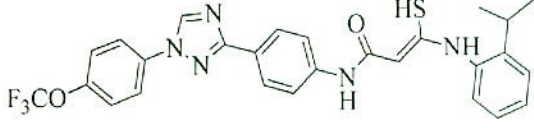
- (A) Ar^1 es un fenilo sustituido que posee uno o más sustituyentes seleccionados de haloalquilo $\text{C}_1\text{-C}_6$ y haloalcoxi $\text{C}_1\text{-C}_6$;
- (B) Het se selecciona de pirazolilo y triazolilo;
- (C) Ar^2 es fenilo o fenilo sustituido que posee uno o más sustituyentes alquilo $\text{C}_1\text{-C}_6$;
- (D) R^1 es H;
- (E) R^2 es (K), H, alquilo $\text{C}_1\text{-C}_6$, o alquilo $\text{C}_1\text{-C}_6\text{-O-C(=O)(alquilo C}_1\text{-C}_6)$;
- (F) R^3 es fenilo sustituido en donde dicho fenilo sustituido posee uno o más sustituyentes seleccionados de F, Cl, alquilo $\text{C}_1\text{-C}_6$, alcoxi $\text{C}_1\text{-C}_6$, cicloalquilo $\text{C}_3\text{-C}_6$, y fenilo;
- (G) R^4 es H;
- (H) M es N o C-R^5 , en donde R^5 se selecciona de H, CN y $\text{C(=O)(alquilo C}_1\text{-C}_6)$;
- (I) (1) Q^1 es O,
(2) Q^2 es O o S;

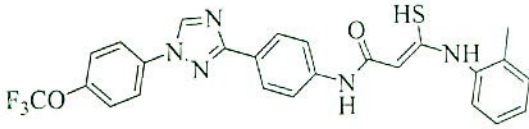
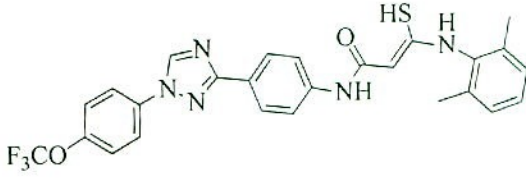
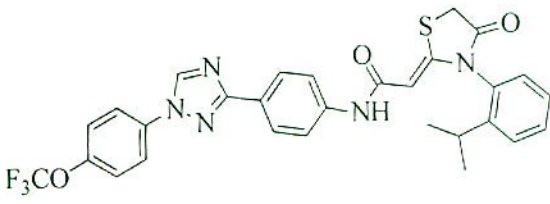
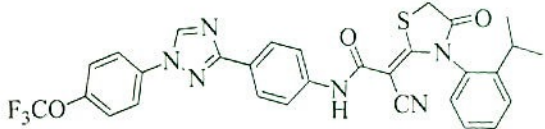
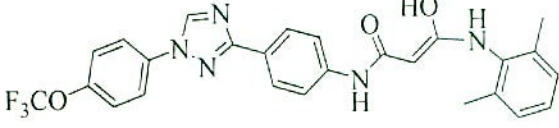
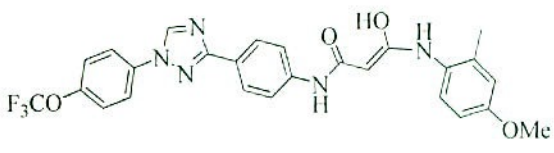
(K) R^2 y R^4 son (K) en donde R^2 y R^4 junto con $C^x(Q^2)(N^x)$, forman un grupo ciclilo hidrocarbilo, saturado o insaturado, de 4 a 7 miembros, en donde dicho grupo ciclilo hidrocarbilo puede estar sustituido con oxo o alquilo C_1-C_6 o en donde la unión entre Q^2 y N^x es $CH_2C(=O)$, CH_2CH_2 , $CH_2CH_2CH_2$, o $CH_2CH(CH_3)$; o en donde al menos uno de R^1 , R^2 , o R^4 , es H y dicha molécula existe en forma de uno o más tautómeros.

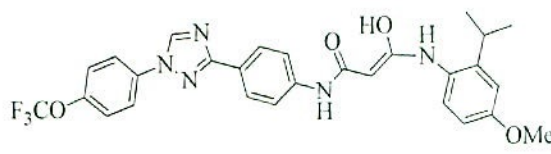
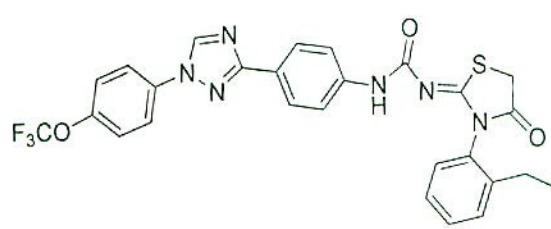
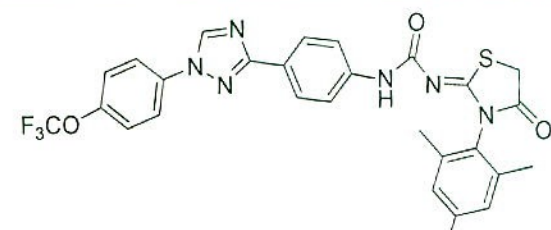
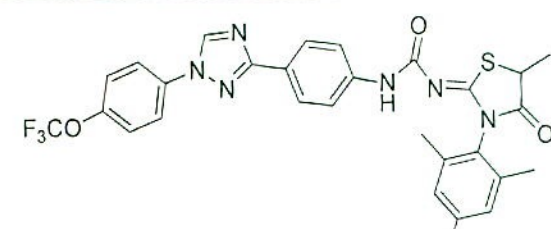
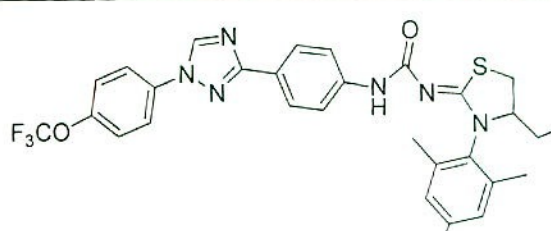
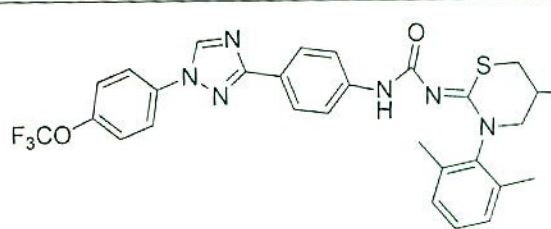
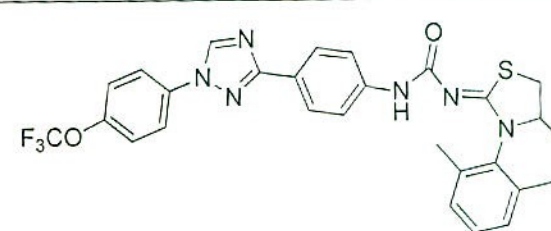
2. Una composición de acuerdo con la reivindicación 1, en donde dicha molécula se selecciona de uno de los siguientes:

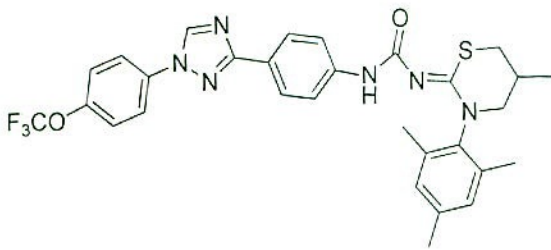
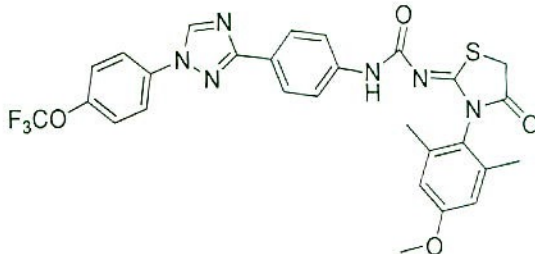
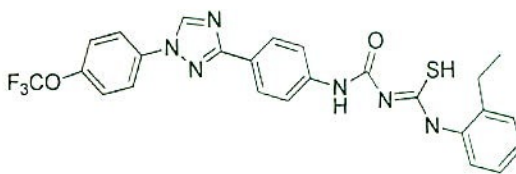
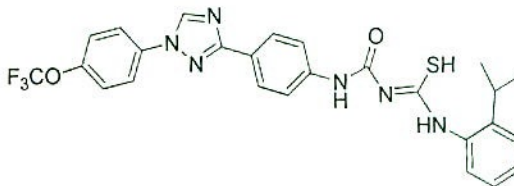
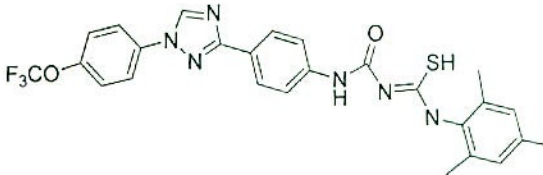
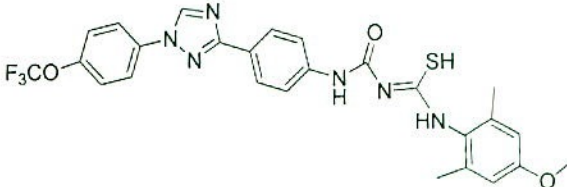
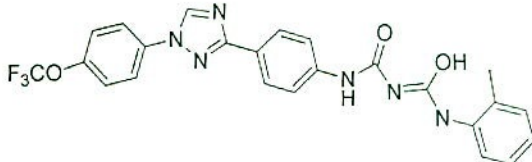
No.	Estructura
A1	
A2	
A3	
A4	

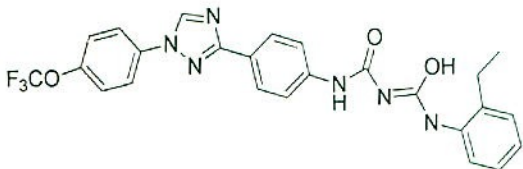
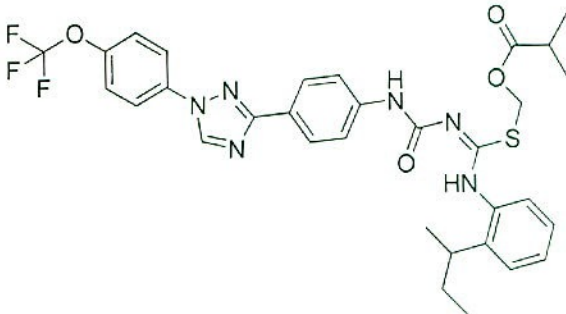
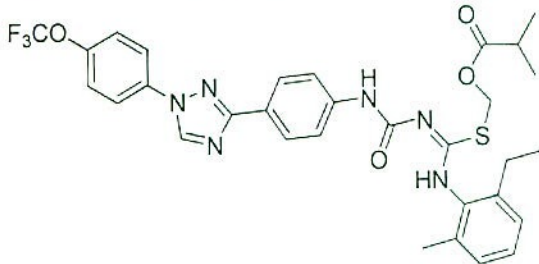
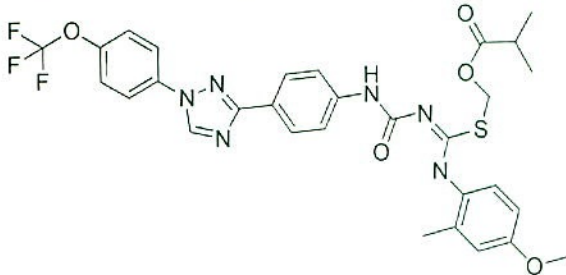
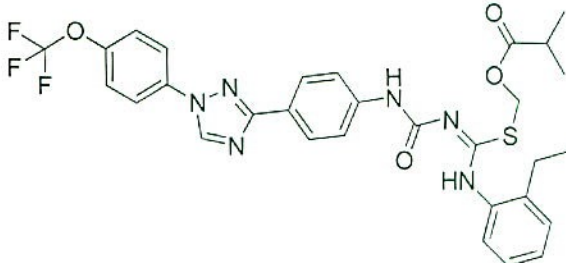
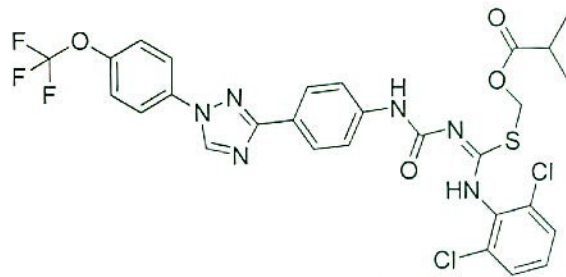
A5	
A6	
A7	
A11	
A12	
A13	
A14	

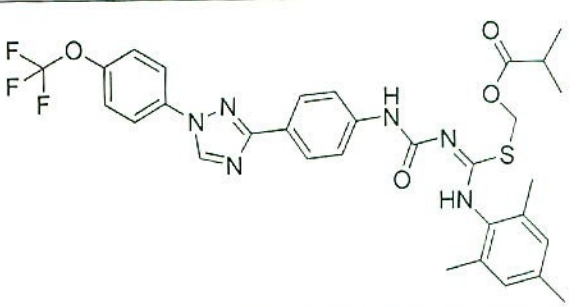
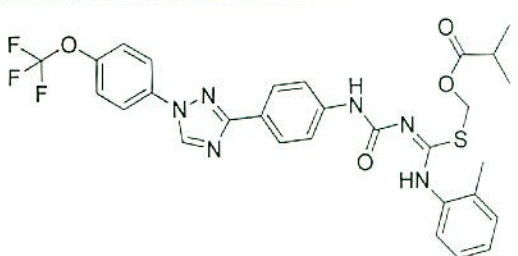
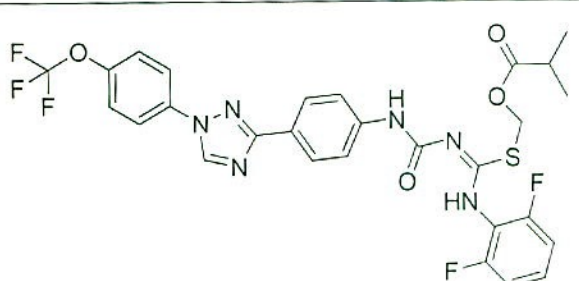
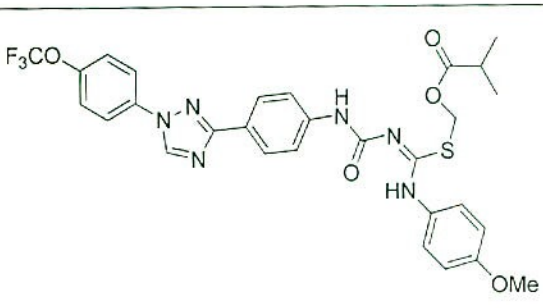
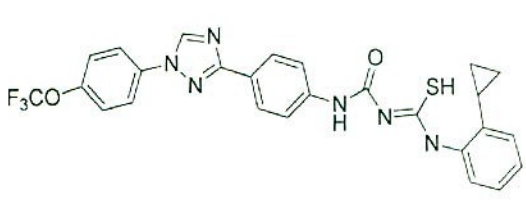
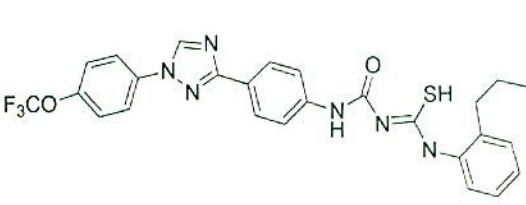
A15	
A18	
A19	
A20	
A21	

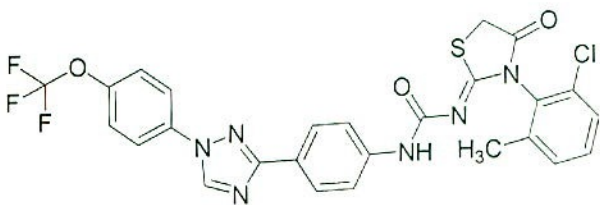
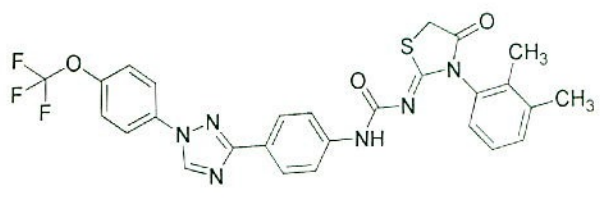
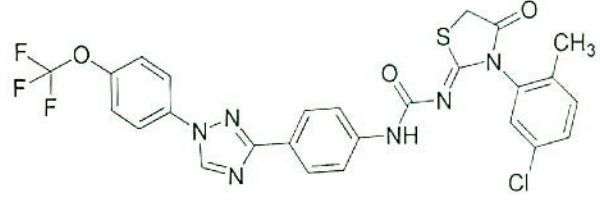
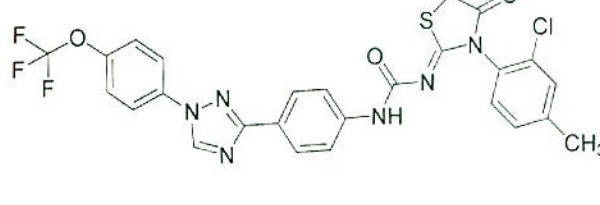
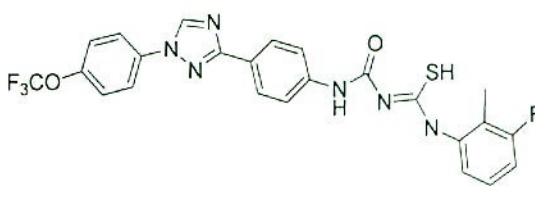
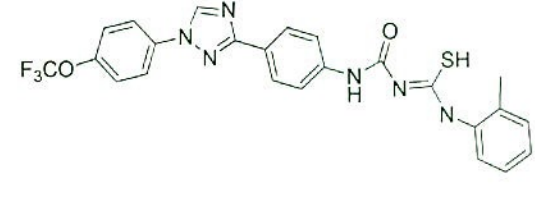
A22	
A23	
A24	
A31	
A32	
A33	

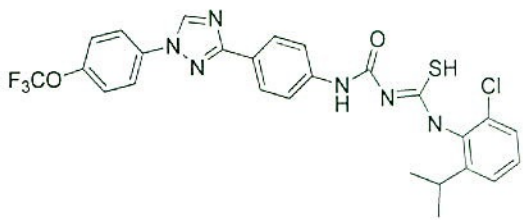
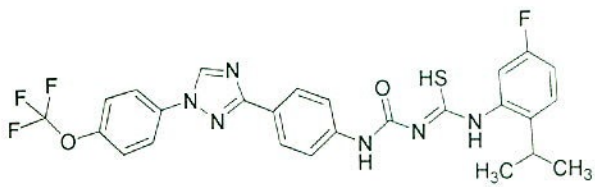
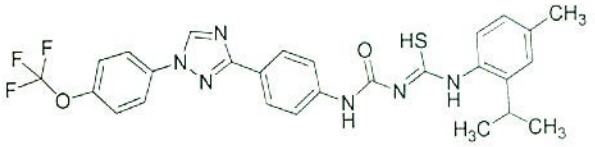
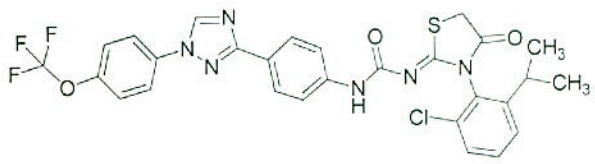
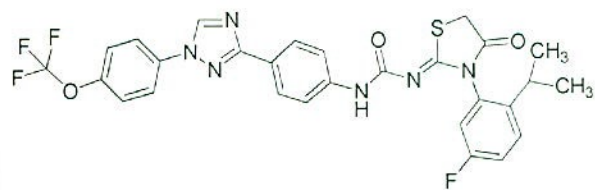
A34	 <chem>COc1ccc(cc1)N/C=C/C(=O)Nc2ccc(cc2)n3nn(nc3)c4ccc(cc4)Oc5ccc(cc5)F(F)F</chem>
A35	 <chem>CCc1ccccc1N2C(=O)CCSC2=N/C=C/C(=O)Nc3ccc(cc3)n4nn(nc4)c5ccc(cc5)Oc6ccc(cc6)F(F)F</chem>
A36	 <chem>Cc1cc(C)cc(C)c1N2C(=O)CCSC2=N/C=C/C(=O)Nc3ccc(cc3)n4nn(nc4)c5ccc(cc5)Oc6ccc(cc6)F(F)F</chem>
A37	 <chem>Cc1cc(C)cc(C)c1N2C(=O)CCSC2=N/C=C/C(=O)Nc3ccc(cc3)n4nn(nc4)c5ccc(cc5)Oc6ccc(cc6)F(F)F</chem>
A38	 <chem>CCc1ccccc1N2C(=O)CCSC2=N/C=C/C(=O)Nc3ccc(cc3)n4nn(nc4)c5ccc(cc5)Oc6ccc(cc6)F(F)F</chem>
A39	 <chem>Cc1cc(C)cc(C)c1N2C(=O)CCSC2=N/C=C/C(=O)Nc3ccc(cc3)n4nn(nc4)c5ccc(cc5)Oc6ccc(cc6)F(F)F</chem>
A40	 <chem>Cc1cc(C)cc(C)c1N2C(=O)CCSC2=N/C=C/C(=O)Nc3ccc(cc3)n4nn(nc4)c5ccc(cc5)Oc6ccc(cc6)F(F)F</chem>

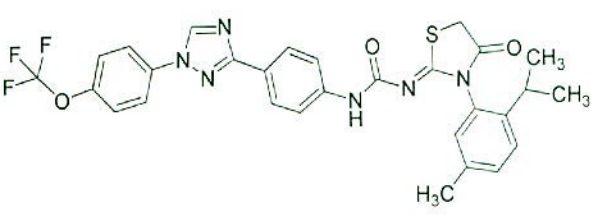
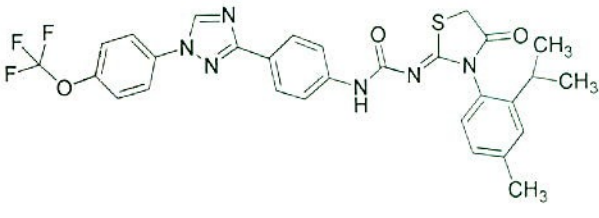
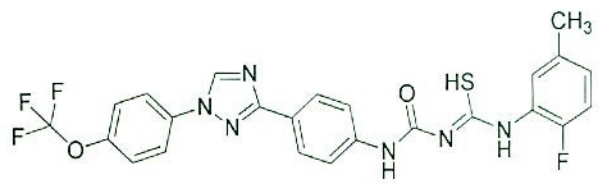
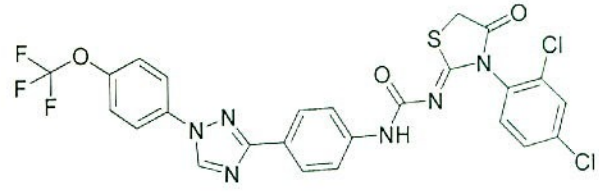
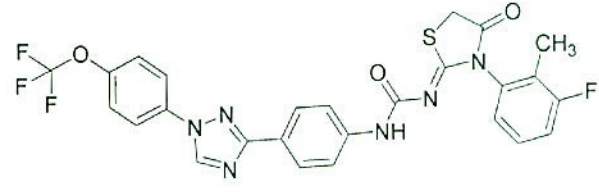
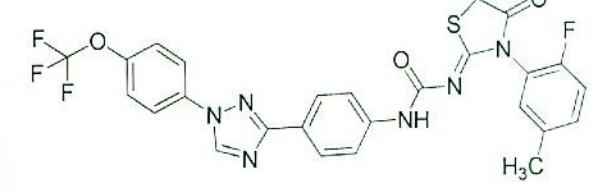
A41	
A44	
A46	
A48	
A49	
A50	
A51	

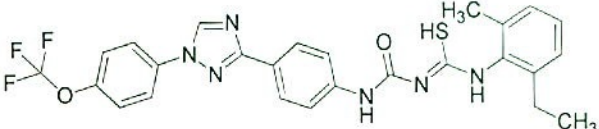
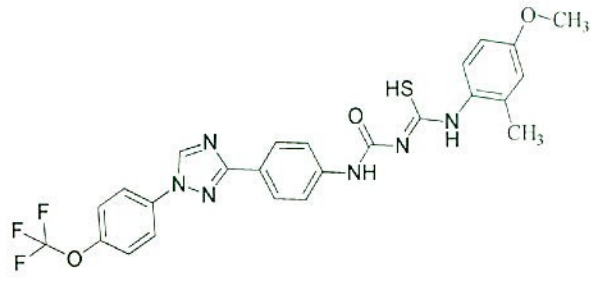
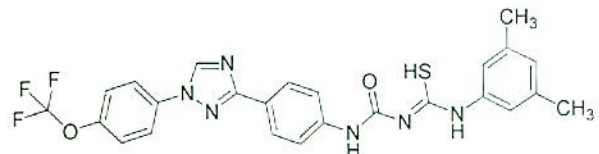
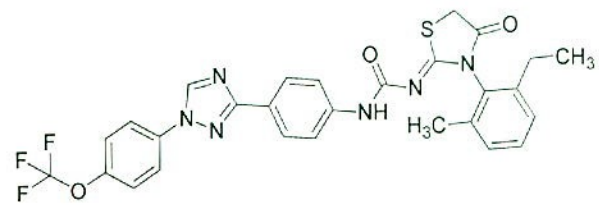
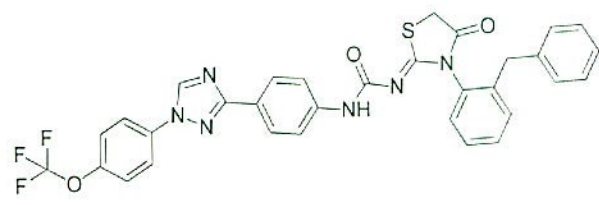
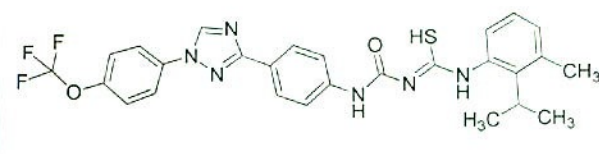
A52	
A54	
A55	
A56	
A57	
A58	

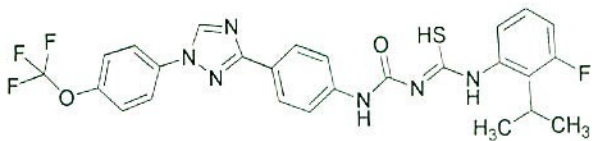
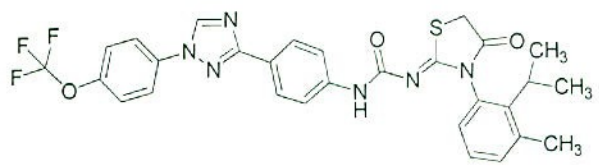
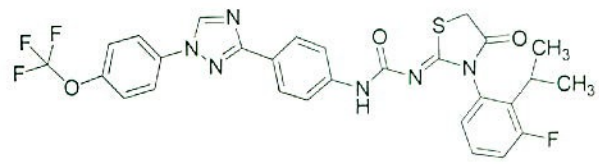
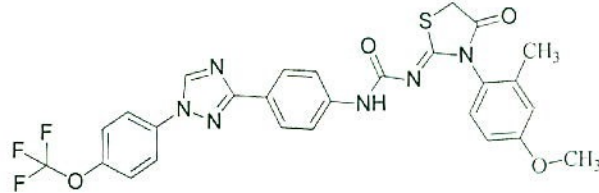
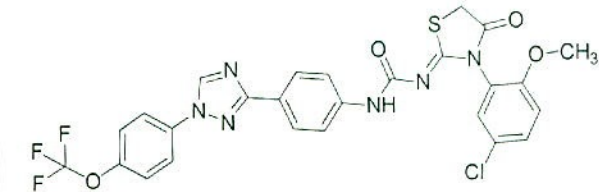
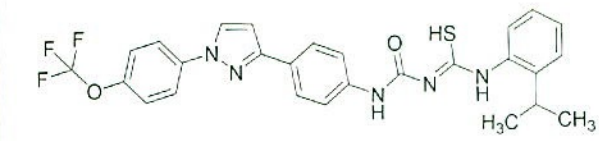
A59	
A60	
A61	
A62	
A63	
A64	

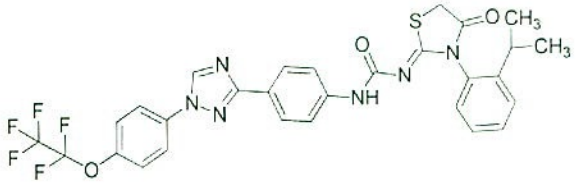
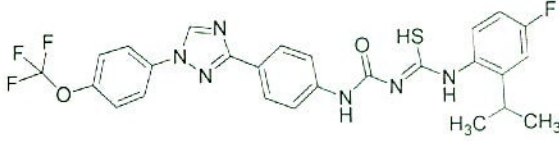
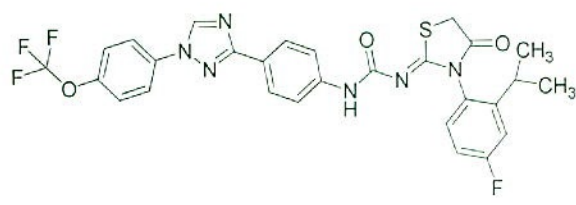
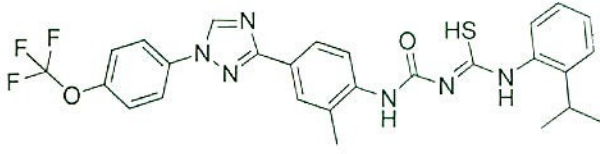
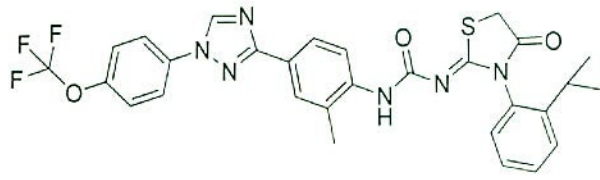
A74	
A75	
A76	
A77	
A79	
A80	

A81	
A82	
A84	
A85	
A86	

A87	 <chem>COC1C(=O)N(C1Sc2ccc(C)cc2)C(=O)Nc3ccc(cc3c4nnn(c4)c5ccc(OC(F)(F)F)cc5)</chem>
A88	 <chem>CC1C(=O)N(C1Sc2ccc(C)cc2)C(=O)Nc3ccc(cc3c4nnn(c4)c5ccc(OC(F)(F)F)cc5)</chem>
A89	 <chem>CC1=CN(C(=N1)S)C(=O)Nc2ccc(cc2c3nnn(c3)c4ccc(OC(F)(F)F)cc4)</chem>
A92	 <chem>Clc1ccc(Cl)c(c1)N2C(=O)N(C2Sc3ccccc3)C(=O)Nc4ccc(cc4c5nnn(c5)c6ccc(OC(F)(F)F)cc6)</chem>
A93	 <chem>CC1C(=O)N(C1Sc2cc(F)ccc2)C(=O)Nc3ccc(cc3c4nnn(c4)c5ccc(OC(F)(F)F)cc5)</chem>
A94	 <chem>CC1C(=O)N(C1Sc2ccc(F)cc2)C(=O)Nc3ccc(cc3c4nnn(c4)c5ccc(OC(F)(F)F)cc5)</chem>

A97	
A98	
A100	
A103	
A105	
A106	

A107	 <chem>CC(C)c1cc(F)ccc1NC(=S)NC(=O)Nc2ccc(cc2)c3nnn(c3)c4ccc(OC(F)(F)F)cc4</chem>
A108	 <chem>CC1(C)C(=O)N(C1S)C(=S)NC(=O)Nc2ccc(cc2)c3nnn(c3)c4ccc(OC(F)(F)F)cc4</chem>
A109	 <chem>CC1(C)C(=O)N(C1S)C(=S)NC(=O)Nc2ccc(cc2)c3nnn(c3)c4ccc(OC(F)(F)F)cc4</chem>
A110	 <chem>CC1(C)C(=O)N(C1S)C(=S)NC(=O)Nc2ccc(cc2)c3nnn(c3)c4ccc(OC(F)(F)F)cc4</chem>
A111	 <chem>CC1(C)C(=O)N(C1S)C(=S)NC(=O)Nc2ccc(cc2)c3nnn(c3)c4ccc(OC(F)(F)F)cc4</chem>
A114	 <chem>CC(C)c1ccc(NC(=S)NC(=O)Nc2ccc(cc2)c3nnn(c3)c4ccc(OC(F)(F)F)cc4)cc1</chem>

A115	
A116	
A117	
A122	
A123	

Cheminformatics Analysis of Organic Substituents: Identification of the Most Common Substituents, Calculation of Substituent Properties, and Automatic Identification of Drug-like Bioisosteric Groups

Peter Ertl*

Novartis Pharma AG, Molecular Simulation Group, WKL-125.14.20, CH-4002 Basel, Switzerland

Received August 14, 2002

A large set of more than 3 million molecules was processed to find all the organic substituents contained in the set and to identify the most common ones. During the analysis, 849 574 unique substituents were found. Extrapolated to the number of known organic molecules, this result suggests that about 3.1 million substituents are known. Based on these findings the size of virtual organic chemistry space accessible using currently known synthetic methods is estimated to be between 10^{20} and 10^{24} molecules. The extracted substituents were characterized by calculated electronic, hydrophobic, steric, and hydrogen bonding properties as well as by the drug-likeness index. Various possible applications of such a large database of drug-like substituents characterized by calculated properties are discussed and illustrated by reference to a Web-based tool for automatic identification of bioisosteric groups.

INTRODUCTION

The concept of organic substituents and their influence on molecular properties is one of the pillars of modern organic chemistry as well as a basis of QSAR analysis. Prominent examples from this area are the pioneering contribution of Hammett with regard to the effect of substituents on reactivity (including the introduction of substituent σ constants characterizing the electron donating and accepting power of substituents), the reaction mechanism theory of Ingold (including the definition of mesomeric and inductive effects of substituents), the QSAR concept of Hansch (including the definition of substituent hydrophobicity π constants and the application of substituent constants in QSAR¹), and the ideas of Craig and Topliss concerning the influence of substituent properties on biological activity applied in early drug design.^{2,3} The whole combinatorial chemistry is also based on the concept of substituents, acting in this case under the name building blocks.

The present study focuses on organic substituents from the point of view of cheminformatics and tries to answer questions about the total number of substituents in known organic chemistry space and the implications of this number for the size of virtual organic chemistry space. The characterization of substituents by calculated properties is also discussed, including a procedure for calculating substituent drug-likeness based on a comparison of the distribution of substituents in a large database of drugs versus a large database of nondrugs. Finally an example of the application of a large database of drug-like substituents with calculated properties is presented—a Web-based tool for automatic identification of bioisosteric groups.

IDENTIFICATION OF MOST COMMON SUBSTITUENTS, THE NUMBER OF KNOWN ORGANIC SUBSTITUENTS

To estimate the number of substituents and identify the most common organic substituents, we processed a large in-house database of commercially available molecules containing 3 043 941 entries. Structures stored as SMILES strings were processed, and all substituents up to 12 non-hydrogen atoms were extracted. Any group of atoms connected by a single chemically activated (breakable) nonring bond to the rest of the molecule was considered a “substituent”. Charged groups were neutralized when possible. Only substituents containing “organic” atoms (including Si, P, and Se) were collected. The whole molecular manipulation and processing was done using an in-house molecular development kit written in Java.

Analysis of the molecular database yielded 849 574 unique substituents. The substituents with the highest frequency in the database are shown in Figure 1. The size of the database processed is large enough for generalization, so we can claim that substituents displayed in this figure are the most common organic substituents.

Out of all 849 574 substituents found, however, only a very small fraction, namely 50, may be considered to be truly “common”, i.e., present in more than 1% of all molecules in the database, while 438 substituents are present in more than 0.1% of all molecules. About 64% of all substituents found are singletons (i.e., they are present only in a single molecule in the whole database processed). A similar frequency distribution of fragments in molecules from the Derwent World Drug Index was also obtained in a study by Lewell et al.⁴

Figure 2 shows the dependence of the total number of substituents identified on the number of molecules processed

* Corresponding author phone: ++41 61 6967413; fax: ++41 61 6964069; e-mail: peter.ertl@pharma.novartis.com.

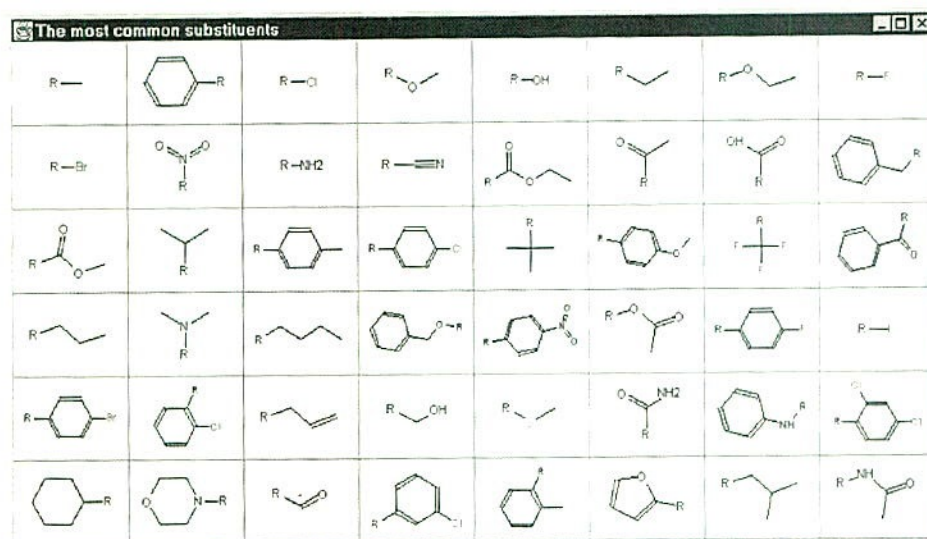


Figure 1. The most common substituents identified after processing a database of 3 043 941 molecules.

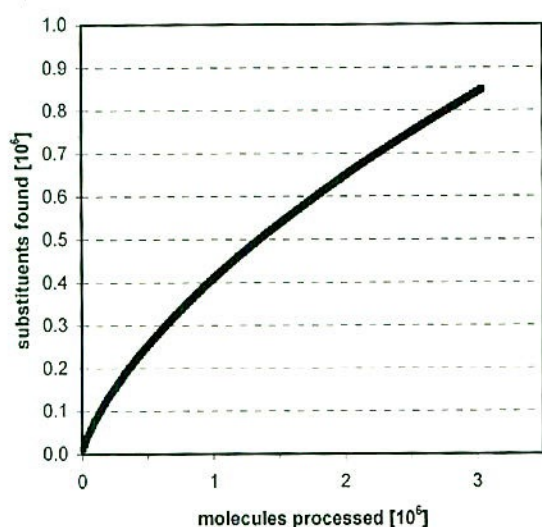


Figure 2. Dependence of the number of unique substituents found on the number of molecules processed.

(this ratio was recorded after processing of every 10 000 molecules). The form of this function was a slight surprise to us. Before the study, we expected a curve that would reach a plateau or limit after a certain number of molecules had been processed. But this was not the case: a steady growth was seen in the number of substituents found. This was even more clear when the two axes were transformed to a logarithmic scale (not shown). In this case, the dependence is clearly linear with a correlation coefficient of $r^2 = 0.998$ for the 304 points recorded. This proves that the diversity of organic chemistry is enormous, and even a relatively small unit consisting of 12 organic atoms allows so many combinations that current organic chemistry is still very far from reaching its limits. Similar conclusions about practically unlimited growth have also been made in studies on the number of types of organic reactions in a large reaction database.⁵

Using the information shown in Figure 2, one can extrapolate to the total size of known organic molecules (about 19 million). According to this analysis, the total number of organic substituents (up to 12 atoms) known in

organic chemistry is about 3.1 million. This number of course depends on the size of the substituent: if substituents of up to 15 atoms were considered, the result would be 7.8 million.

SIZE OF VIRTUAL ORGANIC CHEMISTRY SPACE

Using the information derived in the previous section, one can speculate about the total size of virtual organic chemistry space which is accessible using currently known synthetic methods. To estimate this number, we used a very simple computational experiment, namely to determine how many molecules of the general formula R_1-X-R_2 may be constructed. R_1 and R_2 in this formula are substituents, and X is a scaffold with two attachment points. We analyzed our database of 3 million molecules using the same approach as described in the previous section to identify also the number of double-bonded scaffolds. When using 12-atomic building blocks, the number of structures which may be constructed⁶ by using the above-mentioned model is 5.2×10^{19} ; when using 15-atomic units, this number increases⁶ to 8.0×10^{20} . When using a slightly more complex formula $R_1-X(-R_2)-R_3$ and 9-atomic fragments (to limit the maximum size of products to 36 atoms), then the number of molecules which may be constructed⁶ by this approach is 6.7×10^{23} .

One may argue that many substituents used in this analysis are themselves composed of smaller building blocks/fragments. We therefore decomposed our set of 849 574 substituents in a recursive manner by using the same set of fragmentation rules as described previously, until no further fragmentation was possible. This analysis provided 62 975 "elementary" fragments, the most frequent of which are shown in Figure 3. About 35% of these fragments are monovalent, 34% bivalent, 18% trivalent, and 13% have more than 3 connection points. When constructing molecules by combining these "elementary" fragments, one can easily get more than 10^{100} structures, but most of them would be nonrealistic (not stable, or synthetically inaccessible, at least by currently known synthetic methods).

We therefore estimate the number of molecules which may be prepared by currently known synthetic methods to be somewhere between 10^{20} and 10^{24} . We are aware that there are several deficiencies in this simplified approach (for

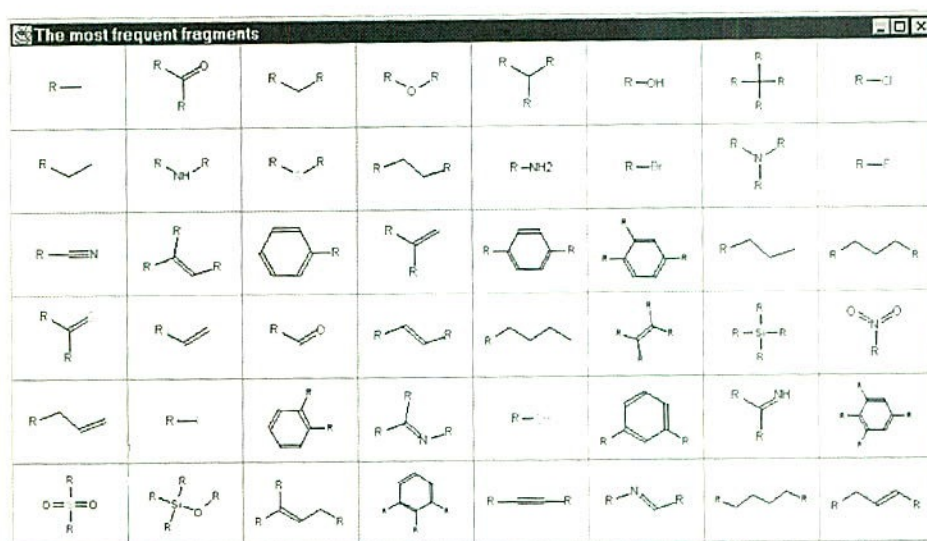


Figure 3. The most common fragments obtained by decomposition of 849 574 substituents.

example neglect of symmetry of the X scaffolds). Nevertheless, we believe the results presented here may be a useful approximation of the overall size of virtual organic chemistry space.

CALCULATION OF SUBSTITUENT PROPERTIES

An analysis of the substituents described hitherto produced a large database of about 850 000 substituents. To use this database in practical applications, the substituents need to be classified in some way. For applications in drug design, a classification based on substituent physicochemical properties is more useful than a simple classification based on structural similarity. In this study, we used substituent size, hydrophobicity, electronic properties (donating/accepting power), and hydrogen bonding characteristics. Steric substituent properties were characterized simply by topological size (number of non-hydrogen atoms) and maximum topological length. Substituent hydrophobicity (π constant) was calculated using the Ghose-Crippen atom contribution method.⁷ The interdependence between calculated and experimental π constants for 256 common substituents (Figure 4) is quite acceptable. For the calculation of substituent electronic properties (Hammett σ constant) we used methodology based on quantum chemical parameters developed in-house. Since the method has been described in detail in the original paper,⁸ the description here is confined to its basic features. The method is based on attaching a small model system (probe) to the substituent. Atomic charges on the probe are then calculated using a semiempirical quantum chemical method. The charges are of course influenced by the attracting or donating power of the substituent. In a training process these calculated charges were correlated with experimental σ constants for a set of substituents. Various probes, semiempirical Hamiltonians, and calculation protocols were tested to find a model with the best predictive power. Correlation between experimental and calculated σ constants for 368 substituents is shown in Figure 5. The correlation in this case is influenced by a low reliability of experimental σ constants for substituents that are not so common.⁸ And as a final parameter, hydrogen bonding substituent properties were generated by a simple approach taking into account the

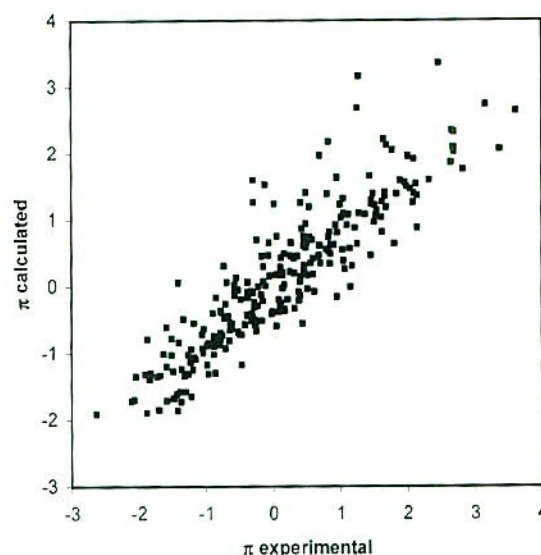


Figure 4. Correlation between calculated and experimental hydrophobicity π constants for 256 substituents.

topological distance of the hydrogen bond donor and acceptor centers from the substituent root.

A two-dimensional view (hydrophobicity versus electronic properties) of this four-dimensional substituent property space for more than 850 000 substituents is shown in Figure 6. This graph is actually nothing more than a well-known Craig plot² with the number of points, however, slightly larger than usual.

MOST COMMON SUBSTITUENTS IN DRUGS AND CALCULATION OF SUBSTITUENT DRUG-LIKENESS

To use the substituent database in drug design, not only substituent properties are important but also a characteristic known as drug-likeness or drugability. This may be defined as an ability of the substituent to generally increase the probability that its parent molecule is a drug. To identify the most common substituents in drugs, we performed an analysis of molecules in the World Drug Index database.⁹ The most common substituents identified by this procedure are shown in Figure 7. A similar analysis of substituents for

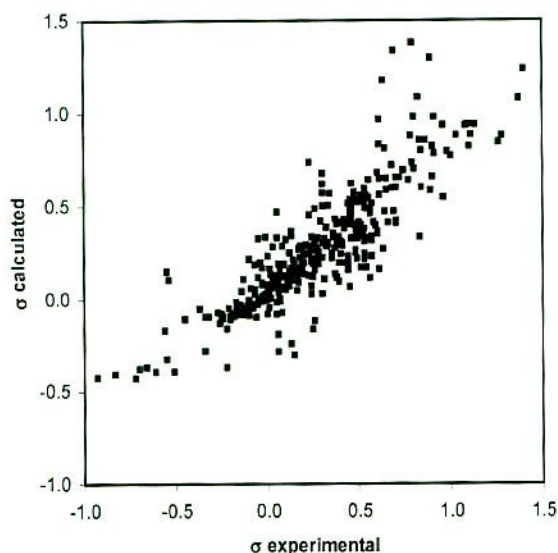


Figure 5. Correlation between calculated and experimental Hammett σ constants for 368 substituents.

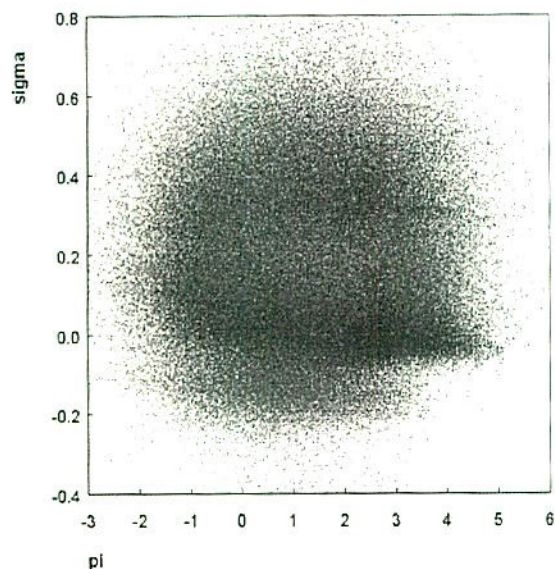


Figure 6. Distribution of calculated hydrophobic (π constant) and electronic (σ constants) properties for about 850 000 substituents (points representing aliphatic carbon substituents are blue, aromatic carbon ones violet, nitrogen green, oxygen red, and sulfur yellow).

5090 molecules from the Comprehensive Medicinal Chemistry database was published recently¹⁰ and despite a slight difference in the definition of "sidechains" used in this study and definition of substituents in our analysis, the results of these two studies are very similar.

A comparison of most common substituents found in drugs (Figure 7) and most common substituents identified in average organic molecules (Figure 1) reveals quite a marked overlap between these two sets. For example, out of 500 most common substituents in both sets 243 are common. We also analyzed a large database of molecules with toxic effects, and the most common substituents identified here (not shown) were again similar to those presented in Figures 1 and 7. These functionalities are simply the most common organic substituents and will be found at the top after analyzing any large collection of molecules. The list of substituents presented in Figure 7 is therefore not the

information needed when looking for substituents with drug-like properties. To identify substituents which beneficially influence the drug-likeness of their parent molecules one needs to identify substituents with *relatively* higher distribution in drugs than in nondrugs. A similar approach of identifying the substructure motifs present more frequently within a specific therapeutic class than in the whole World Drug Index has been described by Lewell et al.⁴

To find substituents with the highest drug likeness, we compared the distribution of substituents in two large databases of molecules: the database of nondrugs (or rather the reference set of average organic molecules) used in our earlier studies (about 3 000 000 molecules) and a subset of the World Drug Index.⁹ The World Drug Index database was preprocessed by removing various "helper" molecules (imaging, radioprotecting, or dental agents, etc) as well as organometallic structures to leave 36 482 molecules which may be termed "true drugs".

Substituent drug-likeness index was calculated according to the following formula

$$f_i = \log(nact_i / ninact_i * ninact_{total} / nact_{total})$$

where $nact_i$ is the number of drugs which contain substituent i , $ninact_i$ is the number of nondrugs which contain substituent i , $nact_{total}$ is the total number of drugs, and $ninact_{total}$ is the total number of nondrugs processed.

In several in-house studies we compared the performance of various indices for separating drugs and nondrugs, and the simple index described above provided the best results. This index ranked very well also in the analysis of Ormerod et al.¹¹ Statistical significance of results was checked by contingency analysis at the 95% level. This ensured that rare substituents with possibly random distribution were not considered in the analysis.

Examples of substituents with high drug-likeness are shown in Figure 8. These substituents (or their derivatives in the proper reactive state) may be used as reagents in drug optimization processes or as building blocks in the design of drug-like combinatorial libraries. At the same time, our analysis also provided a list of substituents with a high negative drug-likeness index, examples of which are shown in Figure 9. The statistical analysis, of course, does not provide any reasons for these groups to be in the "bad" list. They have been identified solely based on the fact that they are present with significantly lower frequency in drugs than in average organic molecules. Therefore groups in this list should be checked manually to identify substructures with potential toxic effects, too reactive groups or metabolic weakpoints. Such groups should be avoided in synthesis planning, because products containing them would very probably be molecules with poor drugability. Such "negative lists" may also be used as exclusion criteria in virtual screening.

A WEB-TOOL FOR AUTOMATIC IDENTIFICATION OF BIOISOSTERIC SUBSTITUENTS

A large database of drug-like substituents/building blocks characterized by calculated properties may be used with advantage in various drug design applications, ranging from "classical" use in QSAR studies or molecular design by hand, through selection of building blocks for synthesis of com-

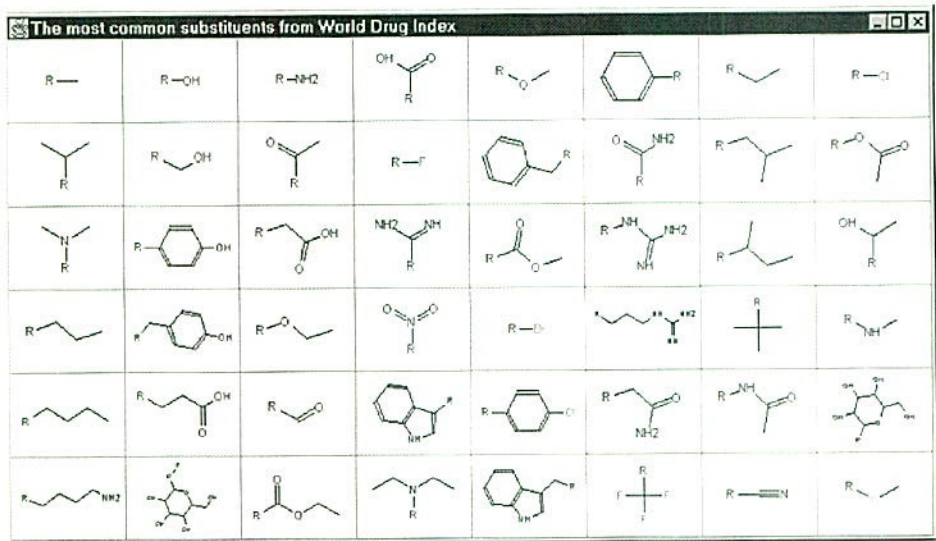


Figure 7. The most common substituents from the World Drug Index.

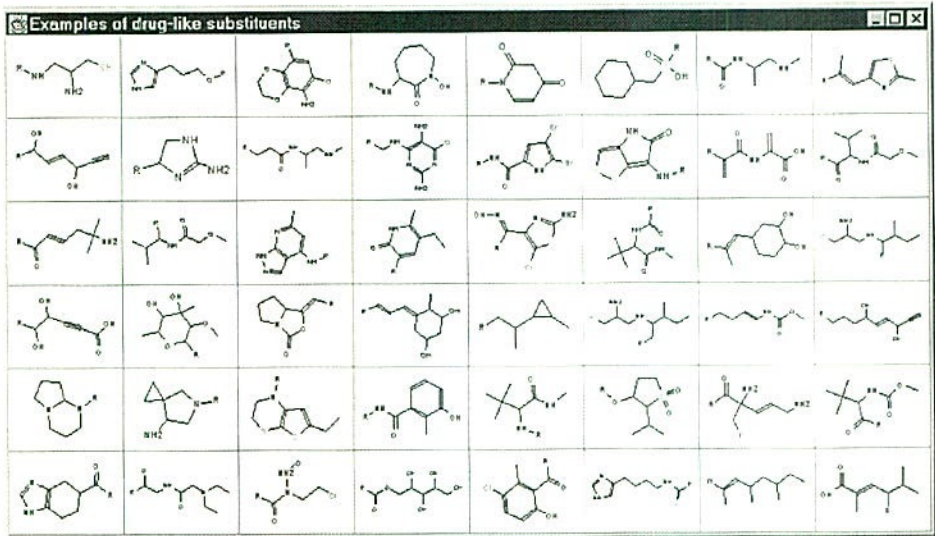


Figure 8. Examples of substituents with the highest drug-likeness index.

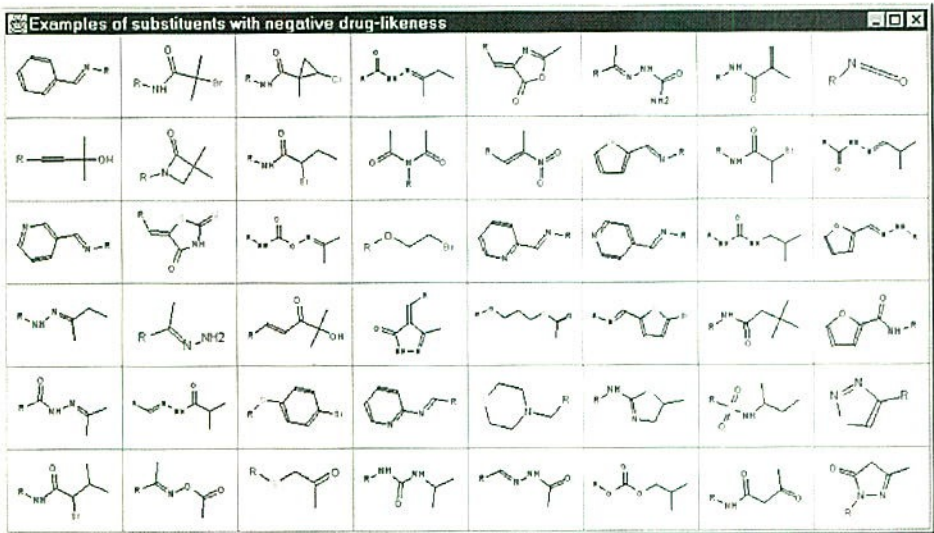


Figure 9. Examples of substituents with negative drug-likeness index.

binatorial libraries with desired characteristics, to automatic design of molecules with optimal properties by evolutionary algorithm. As an example of such applications, a Web tool for automatic identification of bioisosteric functional groups,

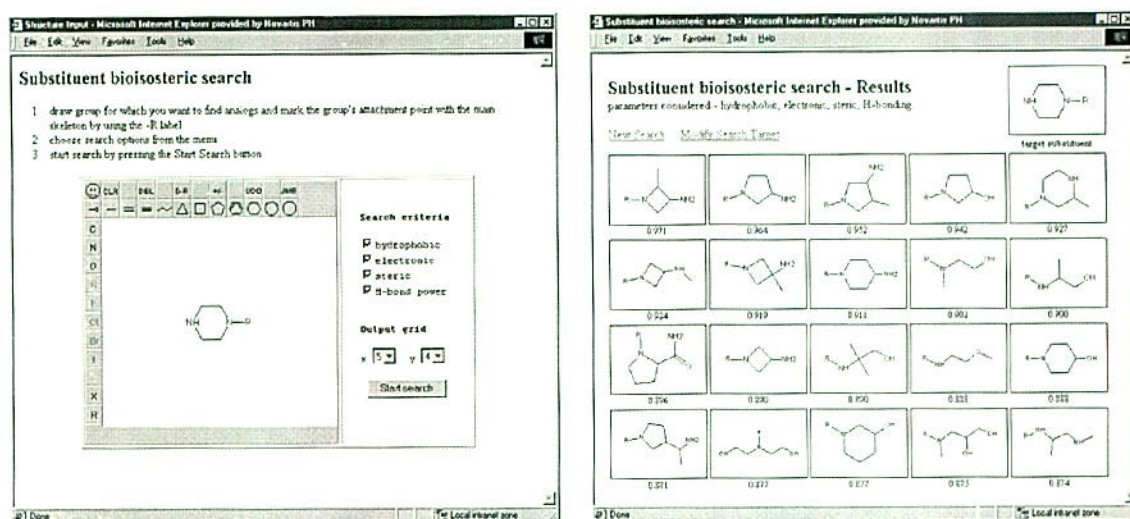


Figure 10. Web-based system for identification of bioisosteric substituents.

which is installed on the Novartis cheminformatics intranet,^{12–14} is presented here.

Bioisosteric replacement is one of the standard techniques used in medicinal chemistry.¹⁵ It can be defined as the replacement of a functional group in a bioactive molecule by another functionality of similar size and physicochemical properties. Bioisosteric transformations are used in the process of drug optimization to improve the properties of drug candidates, such as bioactivity, selectivity, and transport characteristics, and to remove unwanted side effects such as toxicity or an excessively rapid metabolism. Bioisosteric replacement is also used to design molecules more easily to synthesize or avoid patented structural features. Classical textbook examples of bioisosterically equivalent groups are phenyl and thiophenyl or carboxylic group and tetrazole. To find bioisosteric analogues of more complex groups, however, is not so easy. It requires a lot of experience, and even then the identification of a group with an optimal balance of steric, hydrophobic, electronic, and hydrogen-bonding properties, all of which influence ligand–receptor interactions, usually calls for a demanding procedure of trial and error.

That is the reason that a tool for automatic identification of substituent and linkers bioisosteric with the given target was developed. The tool presented here is an updated and modernized version of a Web application briefly presented earlier.¹⁶ The heart of the system is a database of about 10 000 drug-like substituents characterized by calculated hydrophobic, electronic, steric, and hydrogen-bonding properties. When performing a search, the substituent for which analogues are to be identified must be entered into the system. This may be done easily with help of a Java molecular editor applet¹⁷ incorporated directly into the Web page (Figure 10). After search options are chosen, the search is started. Substituents having properties compatible with those of the target are identified in the substituent database and displayed. Such lists of bioisosteric groups may serve as an “ideas generator” for bench chemists, helping them to design new bioactive molecules. The advantage of this approach is that it facilitates the identification not only of traditional analogues but also of nonclassical bioisosteric

analogues, i.e., groups which are structurally different but are compatible in terms of their physicochemical properties.

ACKNOWLEDGMENT

I thank my colleagues Paul Selzer and Bernhard Rohde for many stimulating scientific discussions concerning all areas of cheminformatics and Sigmar Dressler for making the database of 3 million molecules available.

REFERENCES AND NOTES

- (1) Hansch, C.; Leo, A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*; American Chemical Society: Washington, DC, 1995.
- (2) Craig, P. N. Interdependence between Physical Parameters and Selection of Substituent Groups for Correlation Studies. *J. Med. Chem.* **1971**, *14*, 680–684.
- (3) Topliss, J. G. Utilization of Operational Schemes for Analogue Synthesis in Drug Design. *J. Med. Chem.* **1972**, *15*, 1006–1011.
- (4) Lewell, X. Q.; Judd, D. B.; Watson, S. P.; Hann, M. N. RECAP – Retrosynthetic Combinatorial Analysis Procedure: A Powerful New Technique for Identifying Privileged Molecular Fragments with Useful Applications in Combinatorial Chemistry. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 511–522.
- (5) Rohde, B. Reaction type informetrics of chemical reaction databases: How “large” is chemistry? In *Further Advances in Chemical Information*; Royal Soc. Chem.: 1994; pp 109–127.
- (6) The result 5.2×10^{19} was obtained simply by multiplying 5.4×10^6 (number of bivalent scaffolds up to 12 atoms) and two times 3.1×10^6 (number of substituent up to 12 atoms), similarly also for 15 atomic groups ($8.0 \times 10^{20} = 13.2 \times 10^6 \times 7.8 \times 10^6 \times 7.8 \times 10^6$) and for R1X(R2)R3 molecules ($6.7 \times 10^{23} = 3.1 \times 10^6 \times 0.6 \times 10^6 \times 0.6 \times 10^6 \times 0.6 \times 10^6$).
- (7) Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. Atomic physicochemical parameters for three-dimensional structure directed quantitative structure–activity relationships. 4. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163–172.
- (8) Ertl, P. Simple Quantum Chemical Parameters as an Alternative to the Hammett Sigma Constants in QSAR Studies. *Quant. Struct.-Act. Relat.* **1997**, *16*, 377–382.
- (9) Derwent World Drug Index WDI 2001.2, published by Derwent Information.
- (10) Bemis, G. W.; Murcko, M. A. Properties of Known Drugs. 2. Side Chains. *J. Med. Chem.* **1999**, *42*, 5095–5099.
- (11) Ormerod, A.; Willett, P.; Bawden, D. Comparison of Fragment Weighting Schemes for Substructural Analysis. *Quant. Struct.-Act. Relat.* **1989**, *8*, 115–129.

- (12) Ertl, P.; Jacob, O. WWW-based Chemical Information System. *J. Mol. Struct. (THEOCHEM)* **1997**, *419*, 113–120.
- (13) Ertl, P. QSAR Analysis through the World-Wide Web. *Chimia* **1998**, *52*, 673–677.
- (14) Ertl, P.; Miltz, W.; Rohde, B.; Selzer, P. Web-based Cheminformatics for Bench Chemists. *Drug Discovery World* **Fall 2000**, 45–50.
- (15) Patani, G. A.; LaVoie, E. J. Bioisosterism: A Rational Approach in Drug Design. *Chem. Rev.* **1996**, *95*, 3147–3176.
- (16) Ertl, P. World Wide Web-based System for the Calculation of Substituent Parameters and Substituent Similarity Searches. *J. Mol. Graph. Mod.* **1998**, *16*, 11–13.
- (17) JME molecular editor applet allowing creation or editing of molecules may be obtained directly from the author; see also the JME HomePage at www.molinspiration.com/jme/.

CI0255782