

Uniwersytet Śląski
Instytut Chemii
Zakład Chemii Organicznej

Agata Kurczyk

Polifarmakologiczna analiza leków aktywnych względem wirusa HIV

Rozprawa doktorska

Promotor: prof. dr hab. inż. Jarosław Polański



Katowice 2013

*Składam serdeczne podziękowania
Panu prof. dr hab. **Jarosławowi Polańskiemu**
za inspiracje naukowe,
opiekę i wsparcie w trakcie wykonywania prac badawczych,
a także wszelką pomoc i poświęcony czas.*



Dr BG 3435

Współpraca naukowa:

- I. Zakład Chemii Leków Instytutu Farmakologii Polskiej Akademii Nauk w Krakowie.

Składam serdeczne podziękowanie Panu mgr Dawidowi Warszyckiemu za inspirujące dyskusje oraz współpracę naukową w projekcie badawczym opracowania protokołu wirtualnego skriningu.

- II. Samodzielna Pracownia Wirusologii Narodowego Instytutu Leków w Warszawie.

Zespołowi badawczemu Samodzielnej Pracowni Wirusologii NIL pod kierownictwem Pana dr Andrzeja Piasek dziękuję za wykonanie oznaczeń aktywności biologicznej względem wirusa HIV.

Praca powstała przy wsparciu finansowym projektów stypendialnych:

- I. 2010/2011 – **Projekt UPGOW** (Uniwersytet Partnerem Gospodarki Opartej na Wiedzy) współfinansowany ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego;
- II. 2011/2012 – **Projekt UPGOW** (Uniwersytet Partnerem Gospodarki Opartej na Wiedzy) współfinansowany ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego;
- III. 2012/2013 – **Projekt SWIFT** (Stypendia Wspomagające Innowacyjne Forum Technologii) POKL.08.02.01-24-005/10 współfinansowany ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego.

Praca powstała przy wsparciu finansowym projektów stypendialnych:

- I. 2010/2011 – **Projekt UPGOW** (Uniwersytet Partnerem Gospodarki Opartej na Wiedzy) współfinansowany ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego;
- II. 2011/2012 – **Projekt UPGOW** (Uniwersytet Partnerem Gospodarki Opartej na Wiedzy) współfinansowany ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego;
- III. 2012/2013 – **Projekt SWIFT** (Stypendia Wspomagające Innowacyjne Forum Technologii) POKL.08.02.01-24-005/10 współfinansowany ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego.

SPIS TREŚCI

1	WSTĘP	8
2	IDEA I CEL PROJEKTU BADAWCZEGO.....	9
3	CZĘŚĆ LITERATUROWA ORAZ PODSTAWY TEORETYCZNE.....	11
3.1	Lek od pomysłu do klinicznego zastosowania	11
3.2	Skrining wirtualny	16
3.3	Metody uczenia maszynowego jako narzędzia wspomagające badania nad nowym lekiem	19
3.3.1	Sposoby reprezentacji molekuł	21
3.3.2	Czynniki wpływające na efektywność klasyfikacji	23
3.4	Wielokierunkowa eksploracja przestrzeni chemicznej	23
3.4.1	Bazy danych molekularno-strukturalnych jako źródło informacji chemicznej	24
3.4.1.1	Baza danych Drugs@FDA	25
3.4.1.2	Baza danych DrugBank	26
3.4.1.3	Baza danych ChemDB HIV/OI/TB	26
3.4.1.4	Baza danych PubChem.....	27
3.4.1.5	Baza danych Reaxys	28
3.4.2	Eksploracja chemicznych baz danych	28
3.4.2.1	Metody odkrywania wiedzy	29
3.4.2.2	Problemy związane z eksploracją baz danych	30
3.4.2.2.1	Meandry naukometrii	31
3.5	Polifarmakologia – model jednego leku oddziałującego w układzie wieloreceptorowym	35
3.6	Metody fragmentaryczne oraz koncepcja motywów uprzywilejowanych	38
3.6.1	Metody fragmentacji struktur chemicznych	41
3.6.2	Ewolucja pojęcia uprzywilejowania	42
3.6.3	Mechanizm powstawania zjawiska uprzywilejowania biochemicznego	43
3.6.3.1	Uprzywilejowanie struktur białkowych	45
3.7	Leki aktywne względem wirusa HIV	46
3.7.1	Produkty lecznicze zarejestrowane w terapii HIV/AIDS	48

4	BADANIA WŁASNE.....	56
4.1	Notacja oraz wykazów skrótów i symboli stosowanych w opisie badań	57
4.2	Studium lekotypii fragmentu mono- i diazanaftalenowego	59
4.2.1	Przestrzeń chemiczna i farmakologiczna azanaftalenów	61
4.2.2	Przestrzeń bioaktywności azanaftalenów	64
4.2.3	Analiza danych dotyczących fragmentu azanaftalenowego ewidencjonowanych w bazie PubChem Compound w funkcji czasu	66
4.2.4	Uprzywilejowana architektura podstawień azanaftalenów	73
4.2.5	Badanie wskaźników dostępności syntetycznej	79
4.2.5.1	Dostępność syntetyczna azanaftalenów	79
4.2.5.2	Odkrywanie wiedzy ukrytej w chemicznych bazach danych.....	88
4.3	Fragmentacja dekrementacyjna leków aktywnych względem enzymu integrazy wirusa HIV	91
4.3.1	Algorytm procedury fragmentacji dekrementacyjnej	91
4.3.2	Topologia lekotypii fragmentów molekularnych wydzielonych w strukturze cząsteczki raltegrawiru	93
4.3.3	Topologia lekotypii fragmentów molekularnych wydzielonych w strukturze cząsteczki elwitegrawiru	99
4.4	Opracowanie i weryfikacja protokołu wirtualnego skriningu inhibitorów enzymu integrazy wirusa HIV-1	106
4.4.1	Biblioteka inhibitorów enzymu integrazy wirusa HIV-1.....	107
4.4.2	Kompozycja i konstrukcja zbiorów uczących	107
4.4.3	Reprezentacja numeryczna cząsteczek chemicznych	110
4.4.4	Badanie efektywności metod uczenia maszynowego i wybór klasyfikatora o najwyższych parametrach predykcji.....	112
4.4.5	Badanie możliwości implementacji metody MI-DSE w określaniu fragmentów molekularnych różnicujących zbiory związków aktywnych i nieaktywnych biologicznie	117
4.4.6	Próba wskazania nowych uprzywilejowanych motywów strukturalnych w zbiorze inhibitorów enzymu integrazy wirusa HIV-1	121
4.4.7	Konstrukcja i weryfikacja protokołu wirtualnego skriningu inhibitorów enzymu integrazy wirusa HIV-1	124
4.4.7.1	Weryfikacja skuteczności opracowanego protokołu wirtualnego skriningu.....	125
4.4.7.1.1	Klasyfikacja inhibitorów enzymu integrazy z wykorzystaniem metod uczenia maszynowego – moduł ML.....	127
4.4.7.1.2	Moduł priorytetyzacji związków chemicznych w oparciu o wskaźnik uprzywilejowania – Moduł PS	128

4.4.7.1.3	Kaskadowe połączenie modułów klasyfikacji oraz priorytetyzacji.....	129
4.4.8	Skrining baz komercyjnie dostępnych związków w oparciu o skonstruowany protokół VS.....	130
4.4.9	Testy biologiczne <i>in vitro</i> związków wyłonionych na drodze wirtualnego skriningu	136
4.5	Podsumowanie	139
5	METODY.....	142
5.1	Przeszukiwanie baz danych w oparciu o podstruktury	142
5.2	Procedura generowania zbioru związków nieaktywnych DUD	143
5.3	Metryki oceny efektywności klasyfikatorów binarnych	144
6	CZĘŚĆ EKSPERYMENTALNA.....	147
6.1	Studium lekotypii fragmentu mono- i diazanaftalenowego – wyniki eksploracji chemicznych baz danych	147
6.2	Fragmentacja dekrementacyjna leków aktywnych względem enzymu integrazy wirusa HIV – dane liczbowe	155
6.3	Oznaczenie cytotoksyczności i aktywności względem wirusa HIV	163
7	SPIS ILUSTRACJI.....	165
8	BIBLIOGRAFIA.....	171
9	ZAŁĄCZNIKI.....	184

1 WSTĘP

Chemia to nauka, której celem jest badanie, modyfikowanie oraz konstruowanie materiałów i związków o praktycznym zastosowaniu w gospodarce. Związki o pożądanych właściwościach, takie jak leki, środki konserwujące bądź modyfikujące smak, zapach itp. otrzymuje się dzięki badaniom skuteczności wybranych połączeń atomów tworzących poszczególne elementy podstruktur cząsteczki. Istnieją dwie możliwe strategie badawcze: eksperymentalna (weryfikacja w układach rzeczywistych) oraz obliczeniowa (symulacja *in silico* w układach modelowych).

Poszukiwanie nowych leków to proces złożony, wieloletni i bardzo kosztowny. Zastosowanie różnorodnych metod informatycznych w połączeniu z danymi eksperymentalnymi umożliwia konstrukcję zaawansowanych algorytmów służących jako narzędzia wspomagające badania nad nowym lekiem. Podstawowym celem badań prowadzonych w zakresie chemii medycznej jest projektowanie cząsteczek posiadających optymalny profil farmakodynamiczny i farmakokinetyczny, a przy tym niską toksyczność. Tylko niewielka część badanej populacji związków przechodzi do etapu badań przedklinicznych, jeszcze mniejsza część zostaje zakwalifikowana do fazy badań klinicznych, a corocznie zaledwie pojedyncze związki zostają zatwierdzone jako leki i wprowadzone do obrotu. Zasadniczo leki użyteczne klinicznie nie są odkrywane bezpośrednio w jednym etapie. Bardziej prawdopodobne jest odkrycie struktury wiodącej (ang. lead compound), czyli prototypowego związku, który wykazuje oczekiwane właściwości farmakologiczne, ale równocześnie może też posiadać wiele cech niepożądanych. Modyfikacje struktury wiodącej mają na celu zwiększenie aktywności i zmniejszenie lub wręcz wyeliminowanie działań niepożądanych. Ze względu na złożoność procesu badań nad nowym lekiem, a tym samym spore ryzyko niepowodzenia na każdym z jego etapów, kluczowym aspektem staje się odpowiedni wybór „punktu startowego”, a więc struktury wiodącej. Zastosowanie metod informatycznych we wczesnych etapach badań nad nowym lekiem jest obiecującą alternatywą, cieszącą się dużym zainteresowaniem w biznesowym środowisku korporacji farmaceutycznych.

2 IDEA I CEL PROJEKTU BADAWCZEGO

Z chemoinformatycznego punktu widzenia chemia organiczna to nieprzebrane zasoby informacji. Chemicy skonstruowali i opisali ponad 60 milionów związków. Związkom tym przypisano szereg właściwości, w tym w szczególności aktywność biologiczną. Dane takie niezbędne są dla efektywnego procesu projektowania molekularnego. Z drugiej strony badania chemiczne mają charakter rozproszony. Realizowane są w różnych laboratoriach i publikowane w różnych czasopismach. Stąd na przykład badania potencjalnych leków przeciw HIV, tytułowy problem mojej pracy, opisują zróżnicowane dane różnej jakości. Dotyczą one często różnych celów działania tych związków, nawet w wypadku kiedy zainteresowanie dotyczy konkretnego pojedynczego enzymu. W innych wypadkach brak jest danych o aktywności badanych związków.

Moje zainteresowania dotyczyły głównie potencjalnych celów przeciw HIV, których kierunkiem działania jest integraza HIV. Najczęściej jednak dostępne dane opisują polifarmakologię tych potencjalnych leków, czyli układ, w którym interesuje nas wiele celów (enzymów, receptorów) działania leku. Także praktycznie przeprowadzone testy aktywności zaprojektowanej biblioteki potencjalnych leków wykonane zostały w układzie, w którym potencjalnym celem działania niekoniecznie jest wyłącznie integraza HIV. Tak więc centralnym problemem mojej pracy stała się polifarmakologia leków przeciw HIV.

Zakres prowadzonych przeze mnie badań obejmuje projektowanie i poszukiwanie związków biologicznie aktywnych z wykorzystaniem metod obliczeniowych i technologii informatycznych ze szczególnym uwzględnieniem eksploracji baz danych molekularno-strukturalnych.

Celem badań było:

1. opracowanie nowych metod eksploracji chemicznych baz danych w celu poszukiwania wiedzy koniecznej dla znajdowania fragmentów molekularnych przydatnych do projektowania leków;
2. wykorzystanie metod eksploracji chemicznych baz danych w celu przeprowadzenia polifarmakologicznej analizy leków aktywnych względem wirusa HIV w odniesieniu do koncepcji lekotypii oraz uprzywilejowania wybranych motywów strukturalnych;
3. poszukiwanie nowych związków aktywnych względem wirusa HIV.

Integralną część pracy stanowią dwie załączone publikacje (**Załącznik 2a** oraz **Załącznik 2b**), które opisują zastosowanie opracowanych przeze mnie metod:

1. Polanski J., Kurczyk A., Bak A., Musiol R.; Privileged structures – dream or reality: preferential organization of azanaphthalene scaffold., *Curr. Med. Chem.*, 2012, 19(13), 1921-1945 (IF₂₀₁₁ = 4,859);
2. Bak A., Magdziarz T., Kurczyk A., Serafin K., Polanski J.; Probing a Chemical Space for Fragmental Topology-Activity Landscapes (FRAGTAL): Application for Diketo Acid and Catechol HIV Integrase Inhibitor Offspring Fragments., *Comb. Chem. High Throughput Screen.*, 2013, 16(4), 274-287 (IF₂₀₁₂ = 2,000).

3 CZĘŚĆ LITERATUROWA ORAZ PODSTAWY TEORETYCZNE

Przez pojęcie aktywności biologicznej określa się szczególny rodzaj działania farmakologicznego, na przykład aktywność przeciwbakteryjna, przeciwnowotworowa czy przeciwbólowa. Konstruowanie cząsteczek nowych leków o określonym profilu właściwości chemicznych lub aktywności biologicznej nazywane jest projektowaniem molekularnym (ang. molecular design). Złożoność tego typu badań związana jest z koniecznością uwzględniania wielu czynników zależnych od specyfiki rozpatrywanego układu lek-receptor, gdzie receptor jest miejscem działania leku, które ostatecznie odpowiada za efekt farmakologiczny. W związku z tym nie istnieje jeden uniwersalny schemat postępowania. Możliwe strategie i koncepcje oraz teoretyczne podstawy mechanizmów i praw rządzących tworzeniem się kompleksu cząsteczki leku z makromolekułą są szeroko opisywane w literaturze [1–4].

3.1 LEK OD POMYSŁU DO KLINICZNEGO ZASTOSOWANIA

W drugiej połowie XIX wieku przeprowadzono pierwsze udane próby wyizolowania z roślin substancji biologicznie aktywnych. Sukces w tej dziedzinie zapoczątkował badania nad poszukiwaniem syntetycznych analogów substancji naturalnych. Jednak w znacznej mierze badania te oparte były na strategii prób i błędów. Rozwój nauk biologicznych oraz wyjaśnienie wielu procesów fizjologicznych na poziomie komórkowym i molekularnym przyczyniły się do zrewolucjonizowania modelu obowiązującego w projektowaniu leków. Obecnie pierwszym etapem badań nad cząsteczką nowego leku jest określenie miejsca działania leku (ang. drug target). W kolejnym etapie poszukuje się substancje oddziałujące z tym miejscem. Rysunek 3.1 przedstawia proces poszukiwania nowego leku.

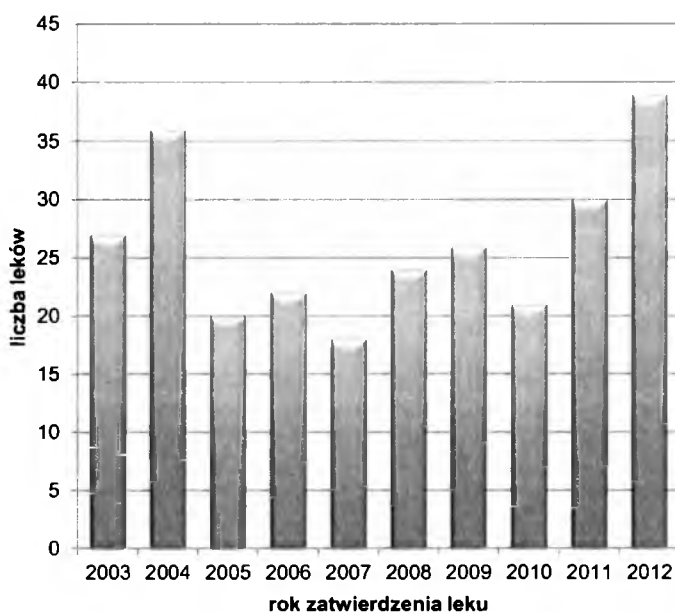


Rysunek 3.1 Lek od pomysłu do wdrożenia. Skrót SAR odnosi się do badań zależności pomiędzy strukturą cząsteczki, a aktywnością biologiczną.

Poszukiwanie nowych leków to proces złożony i kosztowny, a każdemu z jego etapów towarzyszą „pułapki”, które mogą zadecydować o ostatecznym niepowodzeniu. W związku z tym kierunki badań realizowanych przez koncerny farmaceutyczne są silnie skorelowane z czynnikiem ekonomicznym. Realizowane są głównie projekty dotyczące chorób, dla których istnieje duże zapotrzebowanie na coraz doskonalsze środki lecznicze. Wybór jednostki chorobowej, a więc kierunku badań, determinuje kluczową strategię, gdyż w bezpośredni sposób przekłada się na stopień prawdopodobieństwa wdrożenia leku na rynek, a tym samym stopień prawdopodobieństwa zwrotu kosztów inwestycji i wypracowania zysku ze sprzedaży.

Postępy genetyki i biologii molekularnej pozwoliły na odkrywanie coraz większej liczby receptorów i enzymów, a więc potencjalnych miejsc działania nowych leków. Chemia kombinatoryczna w połączeniu z metodami wysokowydajnego badania przesiewowego (ang. high-throughput screening, HTS) miały stać się skutecznym narzędziem służącym wydajnej syntezie i efektywnej identyfikacji cząsteczek o potencjalnym zastosowaniu farmakologicznym. Początki i rozkwit tych metod przypadają na lata dziewięćdziesiąte XX wieku, jednak ich wprowadzenie nie przyniosło spodziewanych rezultatów. Liczba nowych leków (ang. new molecular entities, NMEs) zarejestrowanych przez

Amerykańską Agencję ds. Żywności i Leków (U. S. Food and Drug Administration, FDA) w roku 2007 nie przekroczyła dwudziestu, co stanowiło najniższą wartość farmaceutycznej produktywności począwszy od 1983 roku [5]. W kolejnych latach wartość ta tylko nieznacznie wzrosła (Rysunek 3.2), ale nadal nie istnieje bezpośrednie przełożenie zwiększających się nakładów finansowo-inwestycyjnych w badania nad nowym lekiem na efektywność tych badań wyrażoną liczbą wprowadzonych na rynek leków [5–9].



Rysunek 3.2 Liczba leków zatwierdzonych przez FDA i wprowadzonych na rynek w ostatnich dziesięciu latach. Dane przedstawione na wykresie obejmują tylko leki zawierające nowe struktury chemiczne jako substancje aktywne (leki typu NME). Pominięto informacje dotyczące liczby nowych leków generycznych.

W celu poprawienia skuteczności metod HTS przeprowadzone zostały analizy porównujące cechy strukturalne leków z odpowiednimi strukturami wiodącymi [10–12]. Uzyskane wyniki wskazują jednoznacznie, iż struktury wiodące są znacznie mniej złożone niż wywodzące się z nich leki, posiadają niższe wartości mas molowych i są mniej hydrofobowe. Zasugerowano więc, iż tworzenie bibliotek kombinatorycznych zgodnie z charakterystyką odpowiadającą strukturom wiodącym powinno uskutecznić wyniki skringingu metodami HTS [11,12]. Wykazano również, że wzrost złożoności strukturalnej cząsteczek w bezpośredni sposób przekłada się na zmniejszenie

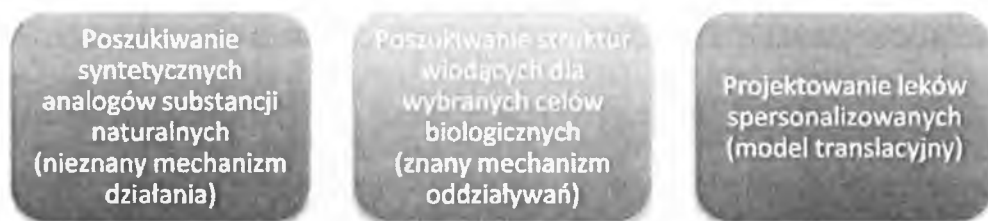
prawdopodobieństwa uzyskania pozytywnego wyniku w teście biologicznym [10]. Można więc przypuszczać, że niska skuteczność metod HTS w głównej mierze mogła wynikać z konstrukcji bibliotek o zbyt wysokim stopniu złożoności.

Model obowiązujący w projektowaniu leków podlega ciągłej ewolucji [13]. Brak korelacji pomiędzy nakładami finansowymi na badania, a ich efektywnością wyrażoną jako liczba wdrożeń nowych leków wskazuje na konieczność reorganizacji schematu badawczego obowiązującego w projektowaniu leków. Antidotum na niski stopień produktywności farmaceutycznej zaczęto upatrywać w skutecznym wdrażaniu założeń tzw. medycyny translacyjnej.

Ideą badań translacyjnych jest szybkie i skuteczne przełożenie wyników eksperymentów laboratoryjnych do praktycznych zastosowań terapeutycznych [14]. Celem medycyny translacyjnej jest stworzenie struktury organizacyjnej, która połączy badania biomedyczne prowadzone w instytucjach naukowych z rzeczywistym zapotrzebowaniem klinicznym. Realizacja tej idei wymaga przede wszystkim skutecznej współpracy ośrodków naukowo-badawczych z koncernami farmaceutycznymi, firmami biotechnologicznymi oraz instytucjami rządowymi [15]. Strategia badań translacyjnych w poszukiwaniu nowych leków preferuje model terapii spersonalizowanej, a więc ukierunkowanej do zastosowania dla ściśle określonej populacji pacjentów. W podejściu translacyjnym do badań nad nowym lekiem można wyróżnić trzy kluczowe aspekty: eksperymenty laboratoryjne, badania kliniczne oraz obserwacje skuteczności terapii dla wybranych grup pacjentów. Badania biomedyczne powinny być prowadzone w taki sposób, aby możliwe było ich szybkie uskutecznienie i przejście do badań klinicznych (etap opisywany w literaturze jako T1). Z kolei efekty kliniczne powinny znaleźć szybkie zastosowanie w regularnej terapii wdrożonej dla docelowej grupy pacjentów (etap T2) [16].

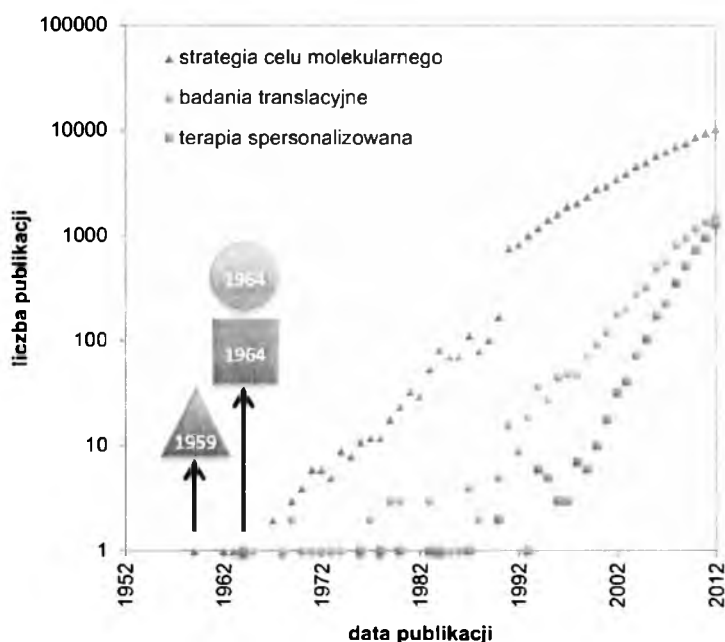
Rysunek 3.3 przedstawia ewolucję modelu obowiązującego w projektowaniu leków. Początkowe koncepcje poszukiwania syntetycznych analogów substancji naturalnych przekształciły się w strategię poszukiwania struktur wiodących dla wybranych celów biologicznych. Transformacja ta była podyktowana rozwojem nauk biologicznych i genetyki. Rozwój ten pozwolił na odkrywanie coraz większej liczby receptorów

i enzymów. Obecnie można zaobserwować stopniowe odejście od tradycyjnego projektowania leków w kierunku modelu translacyjnego, zakładającego spersonalizowane podejście w poszukiwaniu nowych środków leczniczych.



Rysunek 3.3 Ewolucja modelu obowiązującego w badaniach nad nowym lekiem.

Opisany powyżej proces przemian ma swoje odzwierciedlenie w liczbie publikacji traktujących o racjonalnym projektowaniu leków lub z wykorzystaniem założeń modelu translacyjnego, jaka ukazała się na przestrzeni kilkudziesięciu ostatnich lat (Rysunek 3.4). W roku 1991 nastąpił wzrost liczby artykułów opisujących ścieżki poszukiwań nowych leków realizowane strategią celu molekularnego. Tendencja wzrostowa utrzymuje się do chwili obecnej, co wskazuje na duże rozpowszechnienie w chemii medycznej tradycyjnego modelu projektowania farmaceutyków. Jednak od kilku lat zauważalny jest systematyczny wzrost liczby doniesień naukowych odwołujących się do założeń medycyny translacyjnej.



Rysunek 3.4 Liczba publikacji traktujących o badaniach nad poszukiwaniem nowych leków z uwzględnieniem podziału w zależności od przyjętej strategii badawczej. Dane uzyskano metodą eksploracji literaturowej bazy Web of Science.¹ Wyróżniono pierwsze publikacje dla każdej strategii. (Badania własne – materiał przygotowywany do druku).

3.2 SKRINING WIRTUALNY

Skrining wirtualny (ang. virtual screening, VS) jest integralną częścią procesu projektowania leków. VS to metoda analizy wirtualnych bibliotek związków chemicznych, mająca na celu wskazanie molekuł o potencjalnej aktywności biologicznej.

Metody VS dzielą się na dwie kategorie: oparte o budowę znanych ligandów (ang. ligand-based VS) lub bazujące na znajomości struktury receptora (ang. structure-

¹ Składnia zapytań: dla strategii celu molekularnego [Topic=((receptor* OR enzyme* OR ligand* OR molecule* OR drug*) NEAR/5 target*) OR Title=((receptor* OR enzyme* OR ligand* OR molecule* OR drug*) NEAR/5 target*)], dla badań translacyjnych [Topic=((research* OR medicin* OR drug*) NEAR/5 transl*) OR Title=((research* OR medicin* OR drug*) NEAR/5 transl*)] oraz dla terapii spersonalizowanej [Topic=((treat* OR medicin* OR drug*) NEAR/5 personaliz*) OR Title=((treat* OR medicin* OR drug*) NEAR/5 personaliz*)]. Znaczenie operatorów: operator Boole'a OR (LUB) – oznacza sumę zbiorów, czyli wszystkie elementy należące do jednego lub do drugiego zbioru; operator sąsiedztwa NEAR/x – określa maksymalną odległość (mierzoną w liczbie wyrazów „x”), w jakiej znajdują się połączone nim wyrażenia; wieloznacznik w postaci symbolu gwiazdki * – oznacza dowolną liczbę (w tym zero) dowolnych znaków.

based VS), co przedstawia Rysunek 3.5. W zakres pierwszej klasy metod wchodzi zadania realizowane w oparciu o modele farmakoforowe lub podobieństwo strukturalne do związków referencyjnych. Zgodnie z regułą podobieństwa (ang. similarity principle) sformułowaną w roku 1990 przez Johnsona i Maggiore molekuły wykazujące podobieństwo strukturalne często wykazują również podobną aktywność farmakologiczną [17]. Druga klasa metod VS wykorzystuje algorytmy dokowania molekularnego małych cząsteczek chemicznych do miejsca wiążącego receptora. Techniki dokowania pozwalają sprawdzić możliwość powstawania kompleksu ligand-receptor, a także ocenić stopień dopasowania pozy liganda (położenie i konformacja) do kieszeni receptora.



Rysunek 3.5 Podział metod w zadaniach wirtualnego skringu.

Badanie przesiewowe wirtualnych bibliotek struktur chemicznych pozwala na wytypowanie związków, dla których prawdopodobieństwo aktywności biologicznej jest największe. W kolejnym etapie związki te zostają przebadane w rzeczywistym eksperymencie laboratoryjnym w celu weryfikacji uzyskanych drogą obliczeń wyników. Test biologiczny (lub badania przesiewowe) pozwala na porównanie aktywności danego związku z aktywnością związku odniesienia w odpowiednim układzie biologicznym. Zaletą VS jest przede wszystkim możliwość znacznego ograniczenia kosztów związanych z przeprowadzeniem testów biologicznych, ponieważ eksperymenty laboratoryjne wykonywane są tylko dla wytypowanych związków wyjściowej populacji. Rysunek 3.6 przedstawia schemat metodologii VS. Jako etap

wstępny procedury stosuje się filtrowanie wyjściowej biblioteki związków w celu odrzucenia struktur o niekorzystnych parametrach farmakologicznych lub posiadających tzw. niepożądane dla projektowania leków elementy podstruktur (grupy reaktywne lub zmniejszające rozpuszczalność związku). Jako filtry często stosuje się wartości graniczne różnorodnych wskaźników lekotypii.



Rysunek 3.6 Schemat procedury wirtualnego skryningu.

Koncepcja lekotypii (ang. drug-likeness) [3,18–21] zakłada istnienie specyficznego zespołu cech strukturalnych i właściwości fizykochemicznych charakterystycznych dla populacji leków. Najbardziej znana koncepcja definiująca cechy molekularne, które czynią cząsteczkę bardziej lub mniej lekopodobną, to reguła pięciu Lipińskiego (ang. rule of five, Ro5) [22]. Reguła ta została zaproponowana w 1997 roku na podstawie analizy 2245 związków, które zostały dopuszczone do drugiej fazy badań klinicznych. Zasada podaje zakresy wartości właściwości molekularnych, które są istotne dla farmakokinetyki potencjalnego leku w organizmie żywym. Według niej leki preferencyjnie spełniają następujące kryteria:

- masa cząsteczkowa poniżej 500 Da;
- lipofilowość poniżej 5 (w skali clogP);
- poniżej pięciu centrów donorowych wiązania wodorowego (suma grup OH i NH);
- poniżej dziesięciu centrów akceptorowych wiązania wodorowego (liczba atomów O i N).

Przyjmuje się, że cząsteczki których profil właściwości molekularnych nie spełnia dwóch z podanych powyżej warunków, mają znacznie mniejsze szanse zastosowania jako leki. Ponieważ trudno poddać analizie całą przestrzeń związków chemicznych, trudno też ocenić, czy reguły pozwalające odróżnić cząsteczki leków od „nieleków” są rzeczywiście prawdziwe. Wykazano na przykład, że rozkład mas cząsteczkowych związków opisanych w Encyklopedii Beilsteina jest dokładnie taki sam jak rozkład mas cząsteczkowych leków [23]. Reguła pięciu Lipińskiego nie jest więc prawdziwa, co do kryterium mas cząsteczkowych. W innych analizach wykazano, że około 80% cząsteczek związków chemicznych zdeponowanych w chemicznych bazach danych spełnia kryteria reguły Ro5 [24].

Wyniki uzyskane metodą VS weryfikowane są również pod względem różnorodności strukturalnej (ang. molecular diversity), ponieważ odkrywanie nowych leków wymaga eksploracji różnych obszarów przestrzeni chemicznej. Konieczna więc staje się konstrukcja molekuł wypełniających możliwie równomiernie takie obszary. Z wyników uzyskanych na drodze wirtualnego skriningu odrzucane są także związki znane i opatentowane.

W zadaniach wirtualnego skriningu znajdują także zastosowanie metody oparte o algorytmy uczenia maszynowego [25–27].

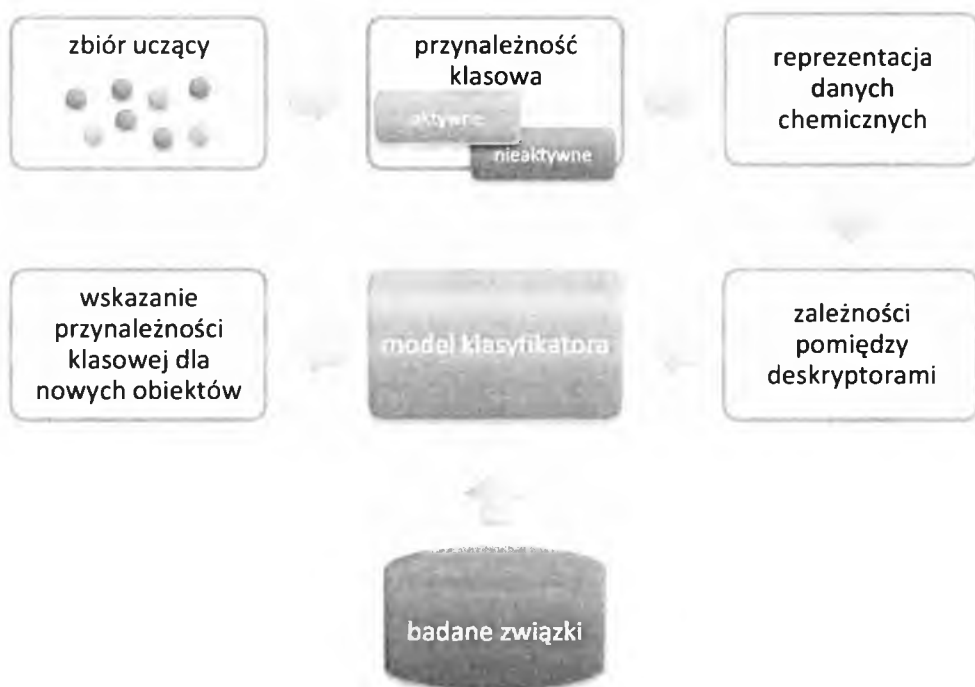
3.3 METODY UCZENIA MASZYNOWEGO JAKO NARZĘDZIA WSPOMAGAJĄCE BADANIA NAD NOWYM LEKIEM

Metody uczenia maszynowego (ang. machine learning, ML) zaliczane są do dziedziny nauk zajmujących się problematyką sztucznej inteligencji. Techniki ML znajdują szereg praktycznych zastosowań w różnorodnych dziedzinach nauki i gałęziach przemysłu. Oprogramowanie do rozpoznawania mowy, automatyczna nawigacja i sterowanie, automatyzacja systemów produkcji, klasyfikacja i analiza danych – w tym przewidywanie trendów na rynkach finansowych, ale także modelowanie i rozwijanie terapii lekowych to tylko wybrane przykłady możliwych implementacji algorytmów ML. W literaturze odnaleźć można przykłady badań, których wyniki dowodzą wysokiej

skuteczności metod ML jako narzędzi wspierających proces badań nad poszukiwaniem nowego leku [28–32].

Systemy uczące się realizują zadania oparte na klasyfikacji lub regresji. Klasyfikacja polega na przypisaniu obiektów (instancji) do określonych kategorii (klas). Z kolei zadania regresyjne umożliwiają predykcję wartości numerycznych (ściślej zakresów wartości) rozpatrywanych parametrów. Proces uczenia może przebiegać bez nadzoru lub z nadzorem. Pierwszy z wymienionych sposobów wykorzystywany jest do poszukiwania wzajemnych relacji w wyjściowym zbiorze danych. Natomiast uczenie z nadzorem wymaga określenia przynależności klasowej obiektów zbioru uczącego. Klasa związków posiadających rozpatrywaną cechę stanowi zbiór przykładów pozytywnych (ang. positive set, P), natomiast klasa jej przeciwstawna definiowana jest jako zbiór instancji negatywnych (ang. negative set, M). Procedura klasyfikacji lub regresji z wykorzystaniem metod ML zawiera dwa następujące po sobie etapy: proces uczenia i następnie testowanie modelu. Podczas pierwszego z nich, w oparciu o przynależność klasową oraz zależności pomiędzy poszczególnymi atrybutami (cechami) obiektów, algorytm buduje model klasyfikatora. W drugim etapie następuje klasyfikacja lub priorytetyzacja obiektów zbioru testowego.

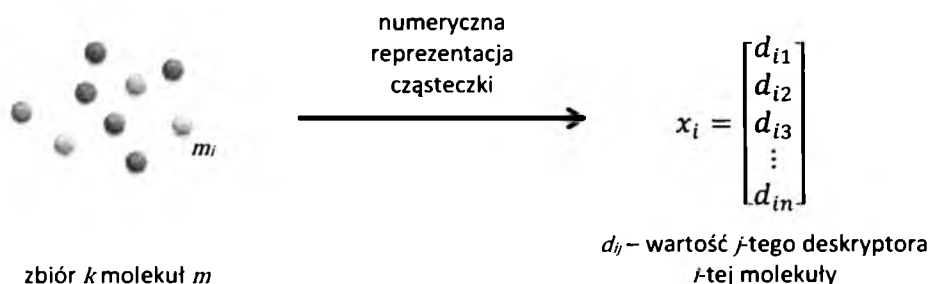
W zastosowaniach chemoinformatycznych najczęściej stosowanym kryterium klasyfikacyjnym jest aktywność biologiczna. Modele regresyjne wykorzystywane są do predykcji wartości wybranych właściwości fizykochemicznych lub wartości związanej z aktywnością biologiczną (np. wartość parametru IC_{50}). Rysunek 3.7 przedstawia schemat klasyfikatora binarnego, którego zadaniem jest dychotomiczny podział badanej grupy związków na aktywne lub nieaktywne w wybranym kierunku stymulacji farmakologicznej.



Rysunek 3.7 Schemat klasyfikatora binarnego w rozwiązywaniu zadań chemoinformatycznych.

3.3.1 SPOSOBY REPREZENTACJI MOLEKUŁ

Zastosowanie metod klasyfikacji lub regresji w rozwiązywaniu zadań chemoinformatycznych wymaga odpowiedniej reprezentacji danych chemicznych. Do opisu struktury i właściwości fizykochemicznych stosuje się n -wymiarowe wektory, których składowe wyrażają wartości numeryczne deskryptorów – co schematycznie przedstawia Rysunek 3.8. Deskryptor molekularny to wynik operacji logicznej lub matematycznej, która przekształca informację chemiczną kodowaną w symbolicznej reprezentacji cząsteczki w jej postać numeryczną [33]. W etapie uczenia i konstrukcji modelu klasyfikatora reprezentacje numeryczne molekuł zbioru treningowego opatrzone są etykietą przynależności klasowej, która oznaczana jest często jako +1 (przynależność do klasy P) lub -1 (przynależność do klasy M).



Rysunek 3.8 Reprezentacja numeryczna zbioru k molekuł w postaci n -wymiarowego wektora.

Deskryptory można podzielić ze względu na ich zakres zastosowania na dwa rodzaje: (i) deskryptory kodujące oraz (ii) deskryptory niekodujące [3]. Pierwsze z wymienionych umożliwiają odtworzenie na ich podstawie pełnej konstytucji lub stereochemii cząsteczki. Przykładem takiego deskryptora może być kod SMILES lub macierz połączeń. Drugi rodzaj stanowią deskryptory, których zadaniem jest charakterystyka tylko wybranych cech molekularnych. Operacja obliczania takich deskryptorów jest jednokierunkowa, a na ich podstawie nie ma możliwości odtworzenia wyjściowej struktury cząsteczki.

Deskryptory obliczane na podstawie fragmentów molekularnych definiują wybrane motywy strukturalne cząsteczki, na przykład liczbę grup hydroksylowych, liczbę grup karboksylowych, liczbę drugorzędowych atomów węgla itp. Przykładem fragmentarycznego deskryptora jest deskryptor używany w modelowaniu HQSAR, którego obliczenie polega na zliczeniu wszystkich możliwych fragmentów molekularnych o określonej liczbie atomów. Daktylogramy molekularne (ang. molecular fingerprints) to specjalny rodzaj deskryptorów fragmentarycznych, które obliczane są metodą analizy elementów podstruktury. Deskryptory daktyloskopowe mogą być przedstawiane w postaci wektorów binarnych kodujących obecność (1) lub brak (0) odpowiedniej cechy strukturalnej lub w postaci wektorów zliczeń odpowiednich elementów.

3.3.2 CZYNNIKI WPŁYWAJĄCE NA EFEKTYWNOŚĆ KLASYFIKACJI

Efektywność procesu klasyfikacji zależy od wielu czynników, między innymi od sposobu reprezentacji molekuł [34], od metody uczenia i jej parametryzacji [35,36], a także kompozycji zbioru uczącego w zależności od liczebności obiektów w klasach P i N [37]. Skuteczność modelu klasyfikatora w głównej mierze zależy od rzetelności danych eksperymentalnych dotyczących bioaktywności. Bazy danych molekularno-strukturalnych często stanowią podstawowe źródło informacji na temat badanej podprzestrzeni chemicznej. O ile jednak stosunkowo łatwo można w nich odnaleźć adnotacje na temat aktywności biologicznej, o tyle znacznie trudniej jest wskazać związki, które będą mogły służyć do konstrukcji zbiorów związków niewykazujących działania w analizowanym kierunku stymulacji farmakologicznej. W chemicznych bazach danych zdecydowanie częściej katalogowane są wyniki eksperymentalne potwierdzające istnienie aktywności, niż te, które wskazywałyby jej brak. Skuteczność klasyfikacji w dużej mierze zależy od informacji jakie zostaną dostarczone w etapie uczenia, a więc od kompozycji i konstrukcji – można powiedzieć, że od „jakości” – zbioru uczącego. Niewystarczająca liczba danych eksperymentalnych potwierdzających brak aktywności biologicznej wymusza definiowanie nieaktywności często na zasadzie przypuszczeń i założeń. Ciekawy przykład stanowi praca, w której w wyniku szeregu eksperymentów porównawczych wykazano wpływ sposobu definiowania i konstrukcji zbioru związków nieaktywnych na efektywność procesu klasyfikacji [38].

3.4 WIELOKIERUNKOWA EKSPLOACJA PRZESTRZENI CHEMICZNEJ

W ostatnich latach model realizacji procedury wirtualnego skriningu uległ przeobrażeniu [26]. Metodologia bazująca na przeszukiwaniu bibliotek związków w oparciu o podobieństwo strukturalne do związku referencyjnego została wyparta przez bardziej złożone techniki wykorzystujące zaawansowane algorytmy eksploracji danych, a także metody uczenia maszynowego. Ewolucja ta częściowo jest wynikiem szybkiego rozwoju chemo- i bioinformatycznych baz danych. Nieograniczony dostęp do

wielomilionowych repozytoriów danych chemicznych w sposób naturalny wymusił rozwój metod umożliwiających wydajne zarządzanie informacją masową.

3.4.1 BAZY DANYCH MOLEKULARNO-STRUKTURALNYCH JAKO ŹRÓDŁO INFORMACJI CHEMICZNEJ

Eksploracja baz danych molekularno-strukturalnych (ang. database mining) dostarczyć może wielu informacji na temat badanej podprzestrzeni chemicznej. Bazy danych, przechowujące informacje strukturalne, fizykochemiczne, a także farmakologiczne, w tym w szczególności adnotacje farmakokinetyczne i farmakodynamiczne małych cząsteczek chemicznych, tzw. ligandów i makrocząsteczek, stanowiących cele działania leku (receptory, enzymy, DNA) umożliwiają zebranie i odpowiednie hierarchizowanie danych dotyczących rozpatrywanej grupy związków. Dzięki temu są dobrym źródłem informacji niezbędnej do przeprowadzenia analizy zależności między strukturą a aktywnością farmakologiczną (ang. structure-activity relationships, SAR). Większość bioinformatycznych baz dostępna jest on-line. Tabela 3.1 przedstawia wybrane przykłady chemicznych baz danych, które dostępne są nieodpłatnie w trybie on-line oraz posiadają interfejs umożliwiający przeszukiwanie w oparciu o podstruktury. Korzystanie z niektórych zasobów wymaga jednak specjalnych licencji i zainstalowania oprogramowania do nawiązania połączenia z serwerem bazy. Na przykład korzystanie z bazy Beilstein CrossFire plus Reactions (obecnie już niedostępna) wymagało zainstalowania klienckiego oprogramowania o nazwie Beilstein Commander, łączącego się w trybie telnetowym z serwerem bazy danych poprzez specjalnie opracowany protokół XFire.

Tabela 3.1 Wybrane bazy danych przechowujące informację chemiczną.

Rodzaj bazy danych	Przykład bazy danych	Liczba zdeponowanych związków	Referencje
Baza leków	DrugBank; http://www.drugbank.ca/	> 6 711	[39–41]
Bazy ligandów	ZINC; http://zinc.docking.org/	> 21 000 000	[42,43]
	ChemDB; http://cdb.ics.uci.edu/	4 100 000	[44]
	PubChem Compound; http://pubchem.ncbi.nlm.nih.gov/	> 32 000 000	[45]
	ChemBank; http://chembank.broadinstitute.org/	> 1 200 000	[46]
	ChEMBL; http://www.ebi.ac.uk/chembl/	1 296 266	[47,48]
Dane liczbowe na dzień 5 lipca 2012.			

W rozdziałach od 3.4.1.1 do 3.4.1.5 sporządzono krótką charakterystykę wybranych repozytoriów.

3.4.1.1 BAZA DANYCH DRUGS@FDA

Baza danych Drugs@FDA dostępna pod adresem <http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm> zawiera informacje na temat zatwierdzonych przez Amerykańską Agencję ds. Żywności i Leków (U. S. Food and Drug Administration, FDA) środków farmakologicznych o przeznaczeniu terapeutycznym i leków dopuszczonych do sprzedaży w Stanach Zjednoczonych. Baza ewidencjonuje leki wprowadzone na rynek amerykański począwszy od roku 1939 do chwili obecnej i jest na bieżąco aktualizowana zgodnie z informacjami publikowanymi w dokumentacji medyczno-farmaceutycznej zwyczajowo nazywanej „Orange Book”. Pełna nazwa publikacji brzmi „Approved Drug Products with Therapeutic Equivalence Evaluations” i jest dostępna odpłatnie w formie papierowej za pośrednictwem U. S. Government Printing Office lub w wersji elektronicznej poprzez serwis Electronic Orange Book (<http://www.accessdata.fda.gov/scripts/cder/ob/default.cfm>).

3.4.1.2 BAZA DANYCH DRUGBANK

DrugBank to bio- i chemoinformatyczna baza łącząca szczegółową charakterystykę profilu chemicznego, farmaceutycznego i farmakologicznego leków z danymi receptorowymi jak sekwencja, struktura czy szlaki biochemiczne. Zawiera około 6800 związków chemicznych przeznaczonych do zastosowania w farmakologii, w tym ponad 1400 leków zatwierdzonych przez FDA. Projekt DrugBank jest prowadzony i rozwijany przez University of Alberta (David Wishart, Departments of Computing Science & Biological Sciences). Na stronie domowej bazy DrugBank (<http://www.drugbank.ca/>) za pośrednictwem rozbudowanego interfejsu graficznego, istnieje możliwość wielokontekstowego przeszukiwania przechowywanych w serwisie danych. Formułowanie zapytań do bazy może być kierowane w trybie tekstowym lub w formie danych strukturalnych. Zaimplementowana aplikacja ChemQuery umożliwia przeszukiwanie w oparciu o podstruktury lub podobieństwo chemiczne do struktury referencyjnej. Karta informacyjna każdego leku zawiera ponad 150 pól, w których zawarto informacje dotyczące zarówno cząsteczki leku, jak i struktur białkowych stanowiących jego cele działania.

3.4.1.3 BAZA DANYCH CHEMDB HIV/OI/TB

Baza danych ChemDB HIV, Opportunistic Infection and Tuberculosis Therapeutics (ChemDB HIV/OI/TB) to przedsięwzięcie rozwijane przez The National Institute of Allergy and Infectious Diseases (NIAID) w celu rejestrowania aktualnego stanu wiedzy na temat związków o potencjalnym zastosowaniu w terapii przeciw HIV/AIDS oraz towarzyszącym tej chorobie infekcjom. Baza jest na bieżąco aktualizowana o informacje strukturalne i aktywności związków testowanych w kierunku przydatności jako potencjalne środki przeciw HIV. Struktura bazy oparta jest na hierarchizacji danych w taki sposób, by możliwe było kontekstowe przeszukiwanie odpowiednio pod względem właściwości chemicznych lub aktywności farmakologicznej. Ponadto aktywność biologiczna w organizacji bazy rozdzielona została na trzy odrębne klasy w zależności od specyfiki i mechanizmu stymulacji farmakologicznej. W ten sposób odseparowano dane na temat aktywności względem wirusa HIV stwierdzonej

w układzie komórkowym od aktywności stwierdzonej w testach przeprowadzonych wobec enzymów wirusa HIV poza układem komórki. Trzecią kategorię stanowią informacje o stymulacji względem oportunistycznych czynników chorobotwórczych towarzyszącym infekcjom wywołanym przez wirusa HIV, a więc baza zawiera również dane dotyczące takich patogenów jak: *Mycobacterium* spp., *Pneumocystis* spp., *Cryptococcus* spp., *Candida* spp., *Aspergillus* spp., *Microsporidia*, *Toxoplasma gondii*, *Cryptosporidium parvum*, Hepatitis A, Hepatitis B, Hepatitis C, a także Cytomegalovirus. Baza dostępna jest on-line pod adresem internetowym: <http://chemdb.niaid.nih.gov/AboutChemDB.aspx>.

3.4.1.4 BAZA DANYCH PUBCHEM

PubChem (<http://pubchem.ncbi.nlm.nih.gov/>) to baza danych zarządzana przez National Center for Biotechnology Information (NCBI), który stanowi część składową National Library of Medicine – instytucji podległej United States National Institutes of Health (NIH). Celem projektu jest gromadzenie informacji na temat aktywności biologicznej małych cząsteczek chemicznych. Wewnętrzna konstrukcja bazy PubChem zorganizowana jest w obrębie trzech kategorii tematycznych: informacji na temat substancji, danych strukturalnych związków, a także ich aktywności biologicznej. Wymienione kategorie stanowią odrębne sekcje bazy macierzystej, kolejno nazywane jako PubChem Substance (ponad 119 milionów rekordów), PubChem Compound (ponad 47 milionów unikatowych struktur) oraz PubChem BioAssay (ponad 600 tysięcy rodzajów oznaczeń biologicznych).² W zależności od rodzaju poszukiwanych informacji, struktura bazy PubChem umożliwia na płynne przechodzenie pomiędzy poszczególnymi sekcjami. Niewątpliwą zaletą serwisu jest rozbudowany interfejs graficzny umożliwiający dogodne formułowanie szczegółowych zapytań zarówno tekstowych, jak i strukturalnych. Ponadto projekt PubChem wspierany jest przez liczne mniejsze serwisy i jednostki partnerskie, dzięki czemu baza jest dynamicznie rozwijana oraz systematycznie aktualizowana.

² Dane liczbowe na dzień 2 czerwca 2013 roku.

3.4.1.5 BAZA DANYCH REAXYS

W styczniu 2009 firma Elsevier uruchomiła bazę danych Reaxys, łączącą bazy Beilstein, Gmelin i Patent Chemistry Database. Po zakupieniu licencji dostęp do bazy Reaxys możliwy jest za pomocą standardowej przeglądarki internetowej. Od roku 2011 dostęp do baz danych poprzez CrossFire został zakończony. Zakres tematyczny bazy obejmuje chemię i dyscypliny pokrewne. Dane pochodzą z następujących źródeł: czasopism chemicznych (od 1771 roku) z zakresu chemii organicznej, nieorganicznej i metaloorganicznej, światowych publikacji patentowych z zakresu chemii organicznej (od 1869 do 1980 roku) oraz anglojęzycznych publikacji patentowych (od 1976 roku) głównie z zakresu chemii organicznej, biocydów, agrochemii i barwników.

Baza Reaxys posiada rozbudowany interfejs graficzny umożliwiający wielokontekstowe przeszukiwanie przechowywanych w serwisie danych. Formułowanie zapytań do bazy może być kierowane w trybie tekstowym lub w formie danych strukturalnych za pośrednictwem wbudowanego edytora molekularnego. Informacja o każdym ze związków chemicznych katalogowanych w bazie obejmują dane strukturalne, nazwy zwyczajowe i handlowe, reakcje z udziałem danego związku, właściwości fizykochemiczne, a także dane odnośnie aktywności farmakologicznej, toksyczności i wpływu na środowisko naturalne. Każdy rekord opatrzony jest zbiorem odnośników literaturowych.

3.4.2 EKSPLOACJA CHEMICZNYCH BAZ DANYCH

Nawet najlepsza baza danych molekularnych jest bezwartościowym repozytorium bez możliwości jej efektywnego i szybkiego skringu. Jest to szczególny problem w przypadku danych molekularnych. Częsteczka chemiczna jest bowiem układem dynamicznym o zmiennej konformacji, konfiguracji czy nawet konstytucji (efekt tautomerii). Konieczność wydajnego zarządzania informacją masową zaowocowała powstawaniem projektów, których nadrzędnym celem jest dostarczenie chemo- i bioinformatycznych narzędzi przeznaczonych do eksploracji wielomilionowych repozytoriów danych chemicznych. Niewątpliwą zaletą tego typu narzędzi jest przede

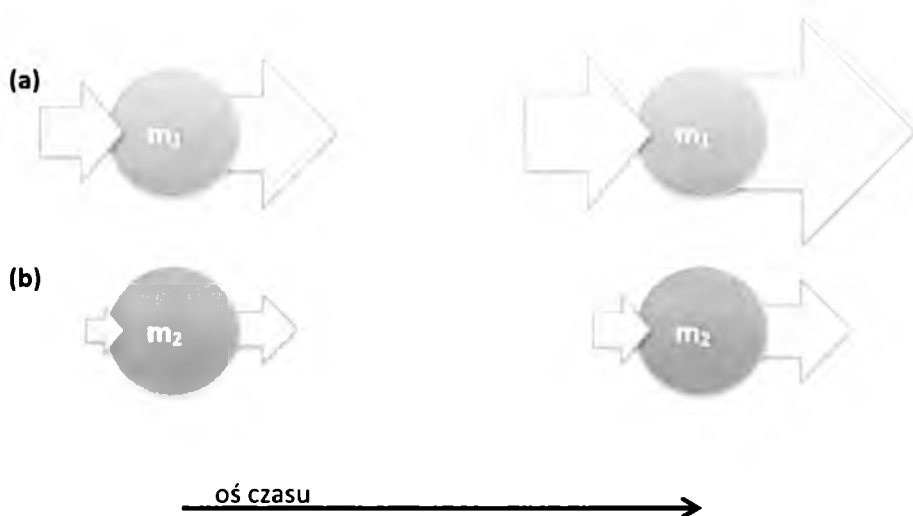
wszystkim możliwość formułowania nie tylko prostych zapytań tekstowych, czy strukturalnych, ale przede wszystkim zapytań wielokontekstowych z równoczesnym swobodnym trybem zapisu uzyskanych rekordów w dogodnym formacie pliku wyjściowego. Przykładem takiej aplikacji jest projekt MoStBioDat, który powstał przy współpracy Zakładu Chemii Organicznej Uniwersytetu Śląskiego z Austrian Research Centers. MoStBioDat to kompleksowe środowisko programowania umożliwiające formułowanie szczegółowych zapytań w celu dogodnej eksploracji wybranej podprzestrzeni chemicznej [49–52].

3.4.2.1 METODY ODKRYWANIA WIEDZY

Bazy danych są niezbędnym narzędziem pracy współczesnego chemika. Umożliwiają poszukiwanie i identyfikację związków organicznych oraz ich właściwości. Dostarczają informacje na temat reakcji chemicznych oraz pozwalają na planowanie on-line możliwych dróg syntezy (baza Reaxys). Baza danych może jednak poza dostarczaniem standardowej informacji pełnić rolę „nauczyciela”. Wykorzystanie technologii informatycznych do szybkiego przeszukiwania i analizy informacji pozwala na ujawnienie relacji, które są szczególnie trudne do odkrycia przez człowieka. Odkrywanie wiedzy z baz danych (ang. knowledge discovery) definiowane jest jako pozyskiwanie potencjalnie użytecznych informacji na podstawie analizy danych masowych. Przy czym pozyskiwanie to następuje w sposób nietrywialny i w końcowym efekcie prowadzi do odkrycia nowych, nieznanych wcześniej faktów [53,54].

Wykorzystując metody odkrywania wiedzy wykonano ciekawe statystyczne analizy danych zgromadzonych w bazie Beilstein, w wyniku których zbadano trendy i preferencje syntetyczne w chemii organicznej na przestrzeni przeszło dwustu lat [23,55]. W ten sposób ujawniono modułarną architekturę połączeń chemicznych. Znane i opisane w literaturze reakcje syntezy organicznej przedstawiono w postaci sieci wzajemnych połączeń. Okazuje się, że węzły (molekuły) tej sieci nie są w takim samym stopniu cenne. Preferowane węzły chemicznej sieci są źródłem większej liczby połączeń, przez co wyznaczają dominujące kierunki syntez. W tym samym przedziale czasowym molekuły o mniejszej „popularności” generują znacznie mniej nowych

połączeń (Rysunek 3.9). W ten sposób chemiczna sieć połączeń ewoluuje w czasie, a jej obraz wyraźnie wskazuje preferowane obszary syntetycznej działalności. Uzyskane wyniki są nie tylko ciekawe z naukowego punktu widzenia. Mogą również służyć do prognozowania stopnia przydatności wybranych struktur na przykład dla zastosowań w przemyśle lub do ekonomicznych prognoz szacujących zapotrzebowanie na wybrane substancje na najbliższe lata.



Rysunek 3.9 Preferencyjność wybranych molekuł do tworzenia większej liczby nowych połączeń w funkcji czasu. Częsteczka m_1 jest źródłem dużej liczby nowych szlaków syntetycznych (a), natomiast cząsteczka m_2 w znacznie mniejszym stopniu rozbudowuje sieć chemicznych połączeń (b). Preferencyjność jest zachowana w funkcji czasu.

3.4.2.2 PROBLEMY ZWIĄZANE Z EKSPLORACJĄ BAZ DANYCH

Bazy danych są niezwykle pomocnym narzędziem pracy we współczesnej chemii, ale efektywne korzystanie z ich zasobów wymaga wcześniejszego poznania struktury systemu bazodanowego oraz zasad formułowania składni zapytań. Większość baz danych pozwala na konstrukcję zaawansowanych zapytań przy użyciu operatorów logicznych (*AND*, *OR*, *NOT*), ale także swoistych dla danej bazy różnorodnych operatorów przeszukiwań, na przykład *SAME*, *NEAR/x* w literaturowej bazie Web of Knowledge (<http://apps.webofknowledge.com/>). Operatory swoiste z reguły można stosować tylko w wybranych polach zapytań bazy, a także nie zawsze można je w dowolny sposób łączyć w obrębie jednego zapytania. Próba przeszukiwania baz

danych bez znajomości tych zasad skutkuje uzyskiwaniem błędnych wyników. Poniższy rozdział jest komentarzem do danych prezentowanych na konferencji Dziekanów Chemii w Poznaniu.

3.4.2.2.1 MEANDRY NAUKOMETRII³

Termin *naukometria*, czy *G1/N12* zgrzyta jak stare hamulce. Nauka kojarzy się z misją, zaangażowaniem i pasją. Cóż kiedy globalny świat wymaga klasyfikacji, rankingów i medali. Publikuj lub giń, oprogramowanie *Publish or Perish* dostępne już w wersji 3.0 to odpowiedź na popyt w tym zakresie. Rankingi i klasyfikacje decydują o karierach, losach uczelni i wydziałów. Na zjeździe Dziekanów Chemii w Poznaniu jako przyczynek przed nadchodzącą kategoryzacją przedstawiono ranking wybranych jednostek grupy *G1/N12* uszeregowanych według liczby publikacji w latach 2009-2012 i cytowań tych prac. Niestety analiza baz danych bywa zdradziecka. Popołnione błędy spowodowały, że podane liczby nie miały się nijak do rzeczywistych. Ponieważ okres zbliżającej się kategoryzacji to także okres wzmożonego zainteresowania naukometrią poniżej podajemy korektę danych poznańskich uzupełnioną kilkoma uwagami o poszukiwaniu danych o afiliacji i cytowaniach w bazach Scopus (Sc) i Web of Knowledge (WoK).

Zastosowana w Poznaniu składnia zapytanie do bazy Scopus np. *AFFIL(jagiellonian university) AND AFFIL(chemistry OR chemical)* zwraca publikacje, gdzie dowolna jednostka *jagiellonian university* występuje obok dowolnej innej zawierającej w nazwie *chemistry OR chemical*. Ten warunek spełnia na przykład publikacja w *Astrophysical Journal* 761 (2), art. no. 91, afiliowana przez *Astronomical Observatory, Jagiellonian University*, oraz *Department of Chemistry and Physics, Purdue University*, nie mająca nic wspólnego z Wydz. Chemii UJ, lub nawet krakowską chemią. Baza Scopus oferuje specjalne pole zapytań o afiliację związaną z uczelniami, gdzie możemy precyzyjnie zdefiniować konkretną jednostkę *AF-ID*. Cóż kiedy nie można jej połączyć z innym zapytaniem oznaczającym „w tej samej linii afiliacji”. W bazie WoK formułując składnię zapytania, trzeba zaś zastosować operator *SAME*, który oznacza „w tej samej linii”.

³ A. Kurczyk, S. Krompiec, J. Polański; „*Meandry naukometrii, czyli sondowanie grupy G1/N12*”. Komentarz do danych przedstawionych na konferencji Dziekanów Chemii w Poznaniu; materiał w przygotowaniu do druku.

Trzeba także pamiętać, że zapytanie *chemistry OR chemical*, inaczej *chem**, zwróci tzw. hity związane ze wszystkimi jednostkami zawierającymi w nazwie *chem**, niekoniecznie związanymi z interesującym nas wydziałem. Z kolei zapytanie o *faculty of chemistry (fac chem* w WoK) nie pozwoli na znalezienie takich prac, których autorzy zapomnieli o wydziale. Prawidłowa liczba prac i ich cytowań będzie więc gdzieś pomiędzy tymi dwoma zapytaniami. Trudno musimy zaakceptować, ten stan niepewności. Zresztą bazy zawierają także inne błędy i nieścisłości.

W tabeli poniżej sondujemy stan podanych w Poznaniu jednostek, reprezentujących poprzednią grupę kategoryzacyjną *G1/N12*. Nie było naszą ambicją stworzenie rankingu tej grupy, lecz zwrócenie uwagi na pułapki naukometrii. Sport i rankingi uzależniają, więc z lokalnej ciekawości, ze względu na katowickie usytuowanie niektórych autorów, uzupełniamy wykaz o Uniwersytet Śląski. Podsumowując, poszukiwanie odpowiedzi na temat rozkładu prac i ich cytowań w bazach danych wymaga głębokiej znajomości struktury tych jednostek oraz składni zapytań. Statystycznie liczba cytowań koreluje w wysokim stopniu z liczbą prac. W pewnym sensie, kontrolnym parametrem może być więc iloraz cytowań do liczby publikacji C/P. Iloraz ten zawiera się w większości w zakresie 3-4, co oznacza, że średnią liczbę cytowań prac w polskiej chemii jest zbliżona dla większości ośrodków, a jej wahania mieszczą się w granicy błędu. Jedynym liderem w tej konkurencji jest składnia UW(Sc) - 4,3.

Tabela 3.2 Sondowanie jednostek grupy *G1/N12* w bazach Scopus (Sc) i Web of Knowledge (WoK), prace opublikowane w latach 2009-2012 i ich cytowania w latach 2009-2012.

Lp.	Zapytanie ^a	Publikacje (P)	Cytowania (C)	C/P
1	US(WoK1) ^b	1660	5229	3,15
2	UW(Sc1)	1164	5022	4,31
3	UW(Sc2)	1156	5006	4,33
4	US(WoK2)	1348	4677	3,47
5	US(Sc1) ^b	1459	4620	3,17
6	US(Sc2)	1363	4489	3,29
7	UJ(WoK1)	1018	3565	3,50
8	UJ(Sc1)	954	3438	3,60
9	UJ(Sc2)	949	3427	3,61
10	PW(WoK1)	917	3330	3,63
11	UW(WoK1)	904	3281	3,63

3. Część literaturowa oraz podstawy teoretyczne

12	UW(WoK2)	772	2980	3,86
13	UAM(Sc1)	1051	2963	2,82
14	UAM(Sc2)	1050	2963	2,82
15	UAM(WoK1)	1049	2957	2,82
16	PW(Sc1)	778	2921	3,75
17	UJ(WoK2)	774	2914	3,76
18	UAM(WoK2)	1021	2909	2,85
19	PW(Sc2)	742	2858	3,85
20	PWr(Sc1)	805	2839	3,53
21	PWr(WoK1)	890	2806	3,15
22	Uł(WoK1)	1042	2729	2,62
23	PW(WoK2)	676	2633	3,89
24	PG(Sc2)	593	2110	3,56
25	PG(Sc1)	595	2109	3,54
26	PWr(Sc2)	573	2059	3,59
27	UG(WoK1)	700	2015	2,88
28	PG(WoK1)	595	1970	3,31
29	UMK(WoK1)	551	1968	3,57
30	PWr(WoK2)	567	1930	3,40
31	UMK(Sc1)	477	1822	3,82
32	UMK(Sc2)	476	1816	3,82
33	PG(WoK2)	487	1811	3,72
34	UG(Sc1)	464	1684	3,63
35	UMK(WoK2)	422	1577	3,74
36	UMCS(WoK1)	564	1562	2,77
37	UMCS(Sc2)	578	1550	2,68
38	UMCS(Sc1)	580	1548	2,67
39	UG(Sc2)	425	1458	3,43
40	UG(WoK2)	470	1416	3,01
41	UMCS(WoK2)	488	1407	2,88
42	Pł(WoK1)	622	1391	2,24
43	Pł(Sc1)	541	1361	2,52
44	PP(Sc1)	333	1304	3,92
45	PP(WoK1)	258	1179	4,57
46	Uł(Sc1)	358	1137	3,18
47	Uł(Sc2)	349	1113	3,19
48	ZUT(WoK1)	521	1075	2,06
49	PSI(Sc1)	357	1059	2,97
50	PSI(Sc2)	352	1056	3,00
51	ZUT(Sc1)	407	992	2,44
52	PSI(WoK1)	336	963	2,87
53	PP(Sc2)	253	859	3,40
54	WAT(WoK1)	332	856	2,58
55	WAT(Sc1)	312	813	2,61
56	ZUT(WoK2)	330	788	2,39
57	ZUT(Sc2)	295	783	2,65
58	PSI(WoK2)	232	743	3,20
59	PP(WoK2)	172	725	4,22
60	WAT(Sc2)	277	696	2,51
61	WAT(WoK2)	289	694	2,40

3. Część literaturowa oraz podstawy teoretyczne

62	UŁ(WoK2)	176	637	3,62
63	PRz(Sc1)	179	500	2,79
64	PRz(Sc2)	179	500	2,79
65	PRz(WoK1)	203	499	2,46
66	PŁ(Sc2)	190	476	2,51
67	PRz(WoK2)	132	426	3,23
68	UO(Sc1)	118	381	3,23
69	PK(Sc1)	203	327	1,61
70	UO(Sc2)	96	321	3,34
71	PŁ(WoK2)	93	313	3,37
72	PK(WoK1)	155	283	1,83
73	UO(WoK1)	82	252	3,07
74	UTP(WoK1)	149	241	1,62
75	UTP(Sc1)	158	240	1,52
76	UTP(Sc2)	154	237	1,54
77	PK(Sc2)	126	213	1,69
78	UTP(WoK2)	112	210	1,88
79	UO(WoK2)	60	195	3,25
80	PK(WoK2)	89	165	1,85

^a US - Uniwersytet Śląski, UW - Uniwersytet Warszawski, UJ - Uniwersytet Jagielloński, PW - Politechnika Warszawska, UAM - Uniwersytet im. Adama Mickiewicza, PWr - Politechnika Wrocławska, UŁ - Uniwersytet Łódzki, PG - Politechnika Gdańska, UG - Uniwersytet Gdański, UMK - Uniwersytet Mikołaja Kopernika, UMCS - Uniwersytet Marii Curie-Skłodowskiej, PŁ - Politechnika Łódzka, PP - Politechnika Poznańska, ZUT - Zachodniopomorski Uniwersytet Technologiczny, PSI - Politechnika Śląska, WAT - Wojskowa Akademia Techniczna im. Jarosława Dąbrowskiego, PRz - Politechnika Rzeszowska im. Ignacego Łukasiewicza, UO - Uniwersytet Opolski, PK - Politechnika Krakowska im. Tadeusza Kościuszki, UTP - Uniwersytet Technologiczno-Przyrodniczy im. Jana i Jędrzeja Śniadeckich.

Przykładowa składnia zapytań do baz na przykładzie UW:

UW(WoK1) - Address=(Univ Warsaw SAME Chem);

UW(WoK2) - Address=(Univ Warsaw SAME Fac Chem) or Univ Warsaw SAME Dept Chem;

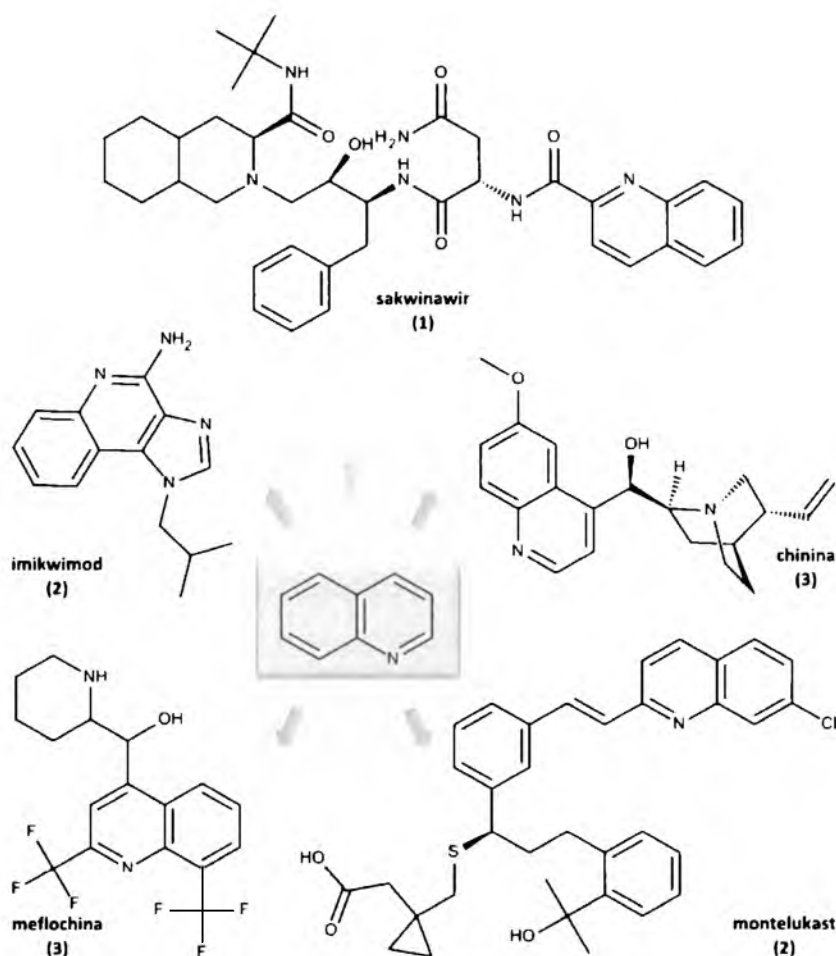
UW(Sc1) - AF-ID("Uniwersytet Warszawski" 60013756) AND AFFIL(univ* AND warsaw AND chem*);

UW(Sc2) - AF-ID("Uniwersytet Warszawski" 60013756) AND (AFFIL(univ* AND warsaw AND fac* AND chem*) OR AFFIL(univ* AND warsaw AND dep* AND chem*)); dla uproszczenia pominięto limit lat.

^b Znając strukturę UŚ, orientujemy się, że składnia WoK1; Sc1 zwraca prace wykonane poza wydziałem M-F-Ch; WoK2 dla Instytutu Chemii/ (N=56/2012) 463/1349;2,91.

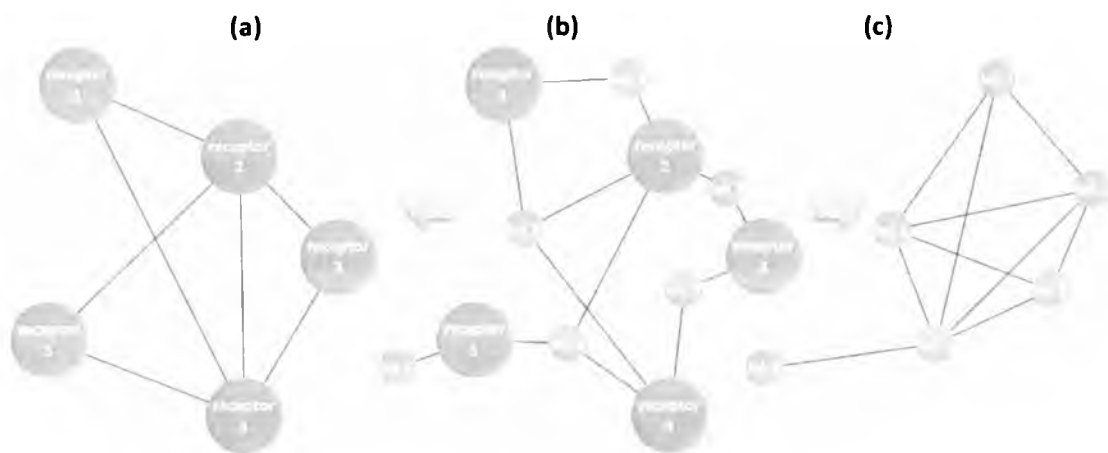
3.5 POLIFARMAKOLOGIA – MODEL JEDNEGO LEKU ODDZIAŁUJĄCEGO W UKŁADZIE WIELORECEPTOROWYM

Bazy danych molekularno-strukturalnych, umożliwiające wielokierunkową eksplorację przestrzeni chemicznej są niezwykle pomocne w rozpatrywaniu problemów polifarmakologii [56]. Od leku oczekuje się, że będzie selektywnie oddziaływał z wybranym receptorem, wywołując pożądany efekt farmakologiczny. Selektywność receptorowa jest jednak trudna do osiągnięcia. Wiele spośród aktualnych leków aktywuje kilka receptorów (co potwierdza przykład czterech związków – montelukast, chinina, meflochina oraz imikwimod (Rysunek 3.10). Innym przykładem są neuroleptyki, których efekt kliniczny wynika z antagonizmu w stosunku do receptorów dopaminowych. Jednak ta grupa leków blokuje także receptory cholinergiczne i α -adrenergiczne, co wywołuje objawy uboczne w postaci obniżenia ciśnienia krwi i nadmiernego uspokojenia [57].



Rysunek 3.10 Przykłady związków zawierających podstrukturę chinoliny i należących do grupy leków zatwierdzonych przez FDA, U. S. Food and Drug Administration. W nawiasach podano liczbę receptorów z którymi leki te oddziałują. Dane uzyskano metodą eksploracji bazy DrugBank. Dane na dzień 14 luty 2013.

Polifarmakologia to innowacyjna koncepcja, która w procesie modelowania mechanizmu oddziaływania lek-receptor skupia się nie tylko na pojedynczym układzie, ale rozpatruje sieć wszystkich połączeń i interakcji wybranego leku z wieloma celami molekularnymi [58]. Rysunek 3.11 przedstawia schemat jedno- i dwuczęściowej sieci zależności polifarmakologicznych. Analiza polifarmakologiczna umożliwia szerokie i kompleksowe badanie aktywności farmakologicznej związku w kontekście układu biologicznego jako całości oraz dostarcza informacji na temat wzajemnej zależności podprzestrzeni chemicznej, biologicznej i farmakologicznej. Strategia ta ponadto często prowadzi do odkrycia nowych miejsc działania znanych leków [59].



Rysunek 3.11 Polifarmakologiczna dwuczęściowa sieć oddziaływań lek-receptor (b) oraz skonstruowane sieci jednoczęściowe lek-lek (c) i receptor-receptor (a). Dwa węzły sieci jednoczęściowej połączone są jeżeli w sieci dwuczęściowej posiadają elementy wspólne.

Polifarmakologia leków w terapii skutkuje w dwojaki sposób – w niektórych przypadkach ujawnia się w postaci niepożądanych efektów ubocznych, ale często wielokierunkowość działania leku i aktywacja kilku różnych receptorów jest niezbędna w celu uzyskania pożądanej odpowiedzi farmakologicznej leków wieloreceptorowych. Selektyność receptorowa jest wskazana tylko wtedy, gdy patogeneza obejmuje działanie pojedynczego czynnika chorobotwórczego. W pozostałych przypadkach, gdy mechanizm powstawania choroby obejmuje równoczesne działanie kilku czynników, skuteczna terapia wymaga zastosowania leków o wielokierunkowym działaniu farmakologicznym, równocześnie aktywujących kilka różnych receptorów (ang. multitarget-directed ligands, MTDLs). Przykładami zaburzeń o złożonej patogenezie są schorzenia neurodegeneracyjne czy zakażenia pierwotniakami [60,61].

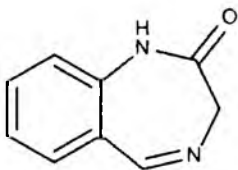
Polifarmakologia leków jest determinowana strukturą chemiczną cząsteczki. Znaczna część leków wieloreceptorowych zawiera w swojej strukturze uprzywilejowany fragment molekularny (ang. privileged structures, PS) [62]. Fakt ten skłonił autorów do sformułowania przypuszczenia, iż uprzywilejowane trzony molekularne mogą z powodzeniem stanowić podstawę i punkt wyjścia racjonalnego projektowania cząsteczek, których pożądana aktywność farmakologiczna wymaga oddziaływania z wieloma celami molekularnymi. Postulat ten odniesiono na przykład do motywu strukturalnego poliamin [63].

Ze względu na rosnącą popularność strategii fragmentarycznych we współczesnej chemii leków, a także wykorzystanie koncepcji uprzywilejowanych motywów strukturalnych w rozpatrywaniu problemów polifarmakologii, w Rozdziale 3.6 opisano genezę i teoretyczne podstawy wymienionych zagadnień.

3.6 METODY FRAGMENTARYCZNE ORAZ KONCEPCJA MOTYWÓW UPRIZYWILEJOWANYCH

Pojęcie uprzywilejowanych motywów w kontekście fragmentów strukturalnych (ang. privileged structures, PS) po raz pierwszy pojawiło się w publikacji z 1988, w której grupa badawcza Evansa opisała badania nad poszukiwaniem antagonistów cholecystokininy, a więc antagonistów rodziny hormonów peptydowych, działających w obrębie układu pokarmowego oraz ośrodkowego układu nerwowego [64]. W opisanych rozważaniach postulowano, iż w niektórych przypadkach architektura związków oparta na wybranych trzonach molekularnych może stać się punktem wyjścia dla skutecznego projektowania molekuł o pożądanym profilu aktywności farmakologicznej: *„certain privileged structures are capable of providing useful ligands for more than one receptor and that judicious modification of such structures could be a viable alternative in the search for new receptor agonists and antagonists”*.

Wnioski powyższe sformułowano na podstawie obserwacji, iż fragment molekularny 1,4-benzodiazepin-2-onu (Rysunek 3.12) wykazuje powinowactwo nie tylko względem badanego celu molekularnego – cholecystokininy – ale także względem kilku innych receptorów.

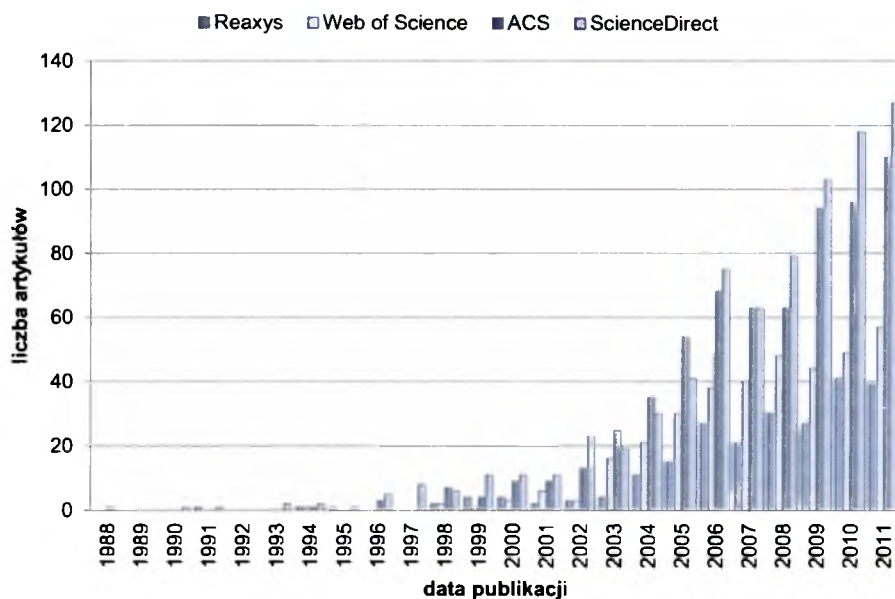


Rysunek 3.12 Prototyp uprzywilejowanych motywów strukturalnych [64].

Biblioteka struktur zbudowana w oparciu o uprzywilejowane trzony molekularne może więc stanowić interesujące źródło związków aktywnych względem różnorodnych

receptorów. W literaturze odnaleźć można opisy konstrukcji i syntez bibliotek związków, których centralny element stanowił jeden z uprzywilejowanych fragmentów molekularnych, na przykład benzopiran [65–67], puryna [68–71], bądź biblioteki związków glikozydowych [72–74]. Inne algorytmy kompozycji [75] i syntezy [76–78] bibliotek związków w oparciu o motywy uprzywilejowane można odnaleźć w literaturze szczegółowej.

Od momentu wprowadzenia terminologii opartej na uprzywilejowaniu wybranych motywów strukturalnych pojawiło się wiele prac porządkujących stan wiedzy w tej dziedzinie [79–84] oraz wskazujących fragmenty molekularne o potencjalnie uprzywilejowanej konstytucji w kontekście odpowiedniego celu molekularnego. Rosnąca liczba corocznie publikowanych artykułów traktujących o strukturach uprzywilejowanych dowodzi coraz większej popularności koncepcji w rozpatrywaniu zagadnień związanych z badaniami nad nowym lekiem (Rysunek 3.13).

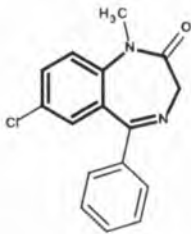
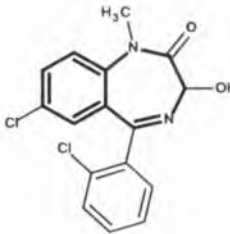
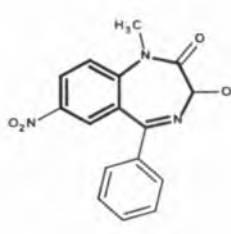
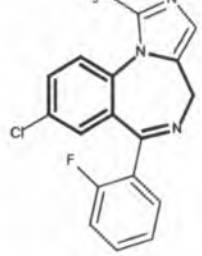
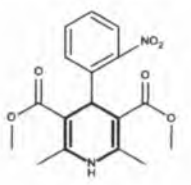
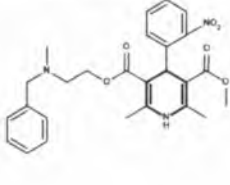
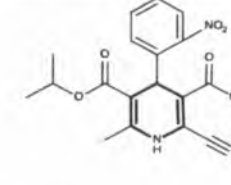
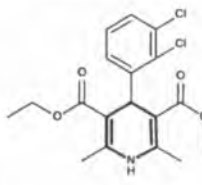
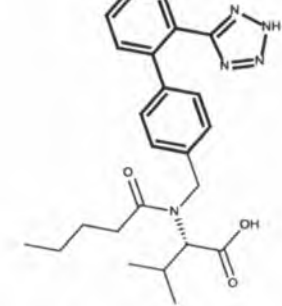
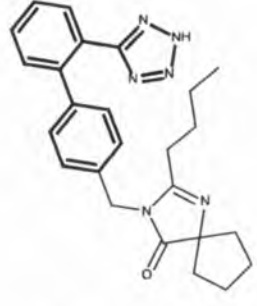
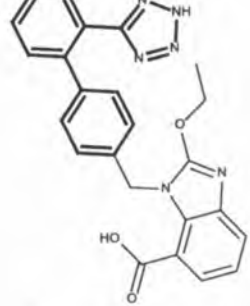


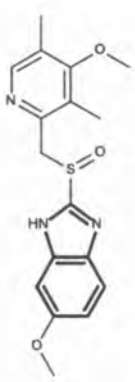
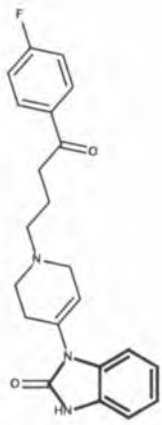
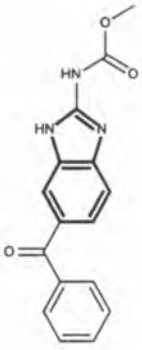
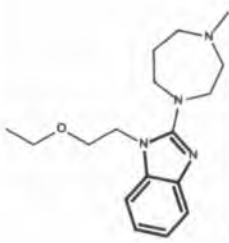
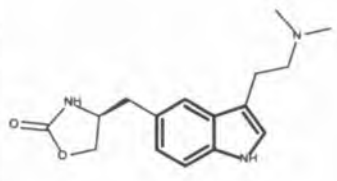
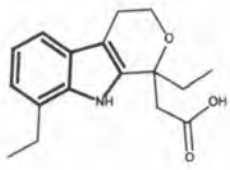
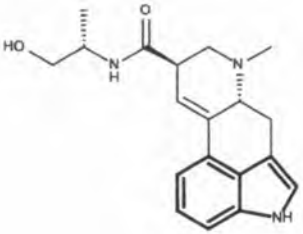
Rysunek 3.13 Liczba publikowanych corocznie artykułów traktujących o strukturach uprzywilejowanych zindeksowanych w wybranych literaturowych bazach danych. Wyniki uzyskano formułując zapytanie tekstowe jako kombinacje określenia „privileged” w połączeniu z następującymi wyrazami „structure/ -s”, „substructure/ -s”, „scaffold/ -s”, „fragment/ -s”. Baza ScienceDirect oraz ACS posiadają opcję przeszukiwania pełnotekstowego (za wyjątkiem referencji). W przypadku pozostałych baz możliwe było przeszukiwanie jedynie w obrębie tytułu/tematu (Web of Science) lub tytułu/słów kluczowych/abstraktu (Reaxys).

3. Część literaturowa oraz podstawy teoretyczne

W szczególności jako uprzywilejowane opisano między innymi następujące struktury: 1,4-benzodiazepin-2-on [64], benzimidazol [84], benzopiran [65–67,85], bifenyl [86,87], 1,4-dihydropirydyna [88], benzylopiperydyna [75,86], difenylotetrazol [75,86], indol [75,86,89]. Tabela 3.3 przedstawia wybrane uprzywilejowane trzony molekularne jako podstruktury leków.

Tabela 3.3 Uprzywilejowane fragmenty molekularne jako podstruktury leków. Adaptacja z publikacji [90].

				(a)
Diazepam [86]	Lorazepam [91]	Nimetazepam [92]	Midazolam [91]	
				(b)
Nifedipine [88,93,94]	Nicardipine [95]	Nilvadipine [96,97]	Felodipine [98]	
				(c)
Valsartan [86,99]	Irbesartan [100]	Candesartan [100]		

				(d)
Omeprazole [101]	Droperidol [102]	Mebendazol [103]	Emedastine [104]	
				(e)
Zolmitriptan [105]	Etodolac [106]	Ergonovine [107]		
Benzodiazepina [84,86] (a), 1,4-dihydropirydyna [88] (b), difenylotetrazol [75,86] (c), benzimidazol [84] (d), indol [75,86,89] (e).				

3.6.1 METODY FRAGMENTACJI STRUKTUR CHEMICZNYCH

Chociaż koncepcja struktur uprzywilejowanych znana jest w literaturze od 1988 roku [64], a od tego czasu udowodniono metodami eksperymentalnymi istnienie uprzywilejowanych elementów podstruktury ligandów wybranych rodzin receptorów [87], to jednak nadal brakuje zaawansowanych metod badawczo-obliczeniowych, pozwalających *in silico* określać i wskazywać nowe uprzywilejowane motywy odpowiedzialne za ułatwienie tworzenia kompleksu lek-receptor.

Algorytmy podejmujące próbę identyfikacji nowych uprzywilejowanych struktur w większości oparte są na analizie częstotliwości występowania różnorodnych fragmentów molekularnych w wybranych zbiorach ligandów, populacji leków, bądź innych grup związków biologicznie aktywnych. Na przykład Bemis i Murcko wprowadzili formalną koncepcję podziału molekuły na stałe segmenty: układy

pierścieni i łączniki, tworzące trzon molekularny (ang. scaffold) oraz łańcuchy boczne [108,109]. Zautomatyzowana analiza trzonów molekularnych pozwala w szybki sposób oszacować różnorodność molekularną badanego zbioru związków. Opracowano strategię fragmentacji struktur zbioru leków (ang. retrosynthetic combinatorial analysis procedure, RECAP) zgodnie z uprzednio zdefiniowanymi regułami rozłamu wiązań, uzyskując w ten sposób zestaw swoistych bloków strukturalnych, które z kolei mogą zostać użyte do wygenerowania kombinatorycznych bibliotek związków [110]. Powstały także algorytmy, według których biblioteki uprzywilejowanych struktur tworzone są z zachowaniem podobieństwa topologicznego ze znanymi lekami oraz z uwzględnieniem narzuconych wymogów fizykochemicznych [111].

3.6.2 EWOLUCJA POJĘCIA UPRZYWILEJOWANIA

Koncepcja uprzywilejowania wybranych trzonów molekularnych od momentu wprowadzenia w 1988 roku po dzień dzisiejszy jest przedmiotem żywej dyskusji nad możliwościami aplikacyjnymi, szczególnie w dziedzinie badań nad identyfikacją nowych, biologicznie aktywnych związków poprzez konstruowanie odpowiednich chemicznych bibliotek [84,112]. Pojęcie „uprzywilejowania” stosowano w różnym kontekście, co niewątpliwie może prowadzić do chaosu. Pierwotne pojęcie uprzywilejowanych motywów strukturalnych odnosiło się do podstruktur, które obecne były wśród ligandów aktywujących różne receptory, ale były to receptory należące do tej samej rodziny. Pomimo jednak, iż Evans i Patchett koncepcję uprzywilejowania sformułowali w odniesieniu do ligandów względnie wąskiej klasy receptorów przynależnej do rodziny receptorów sprzężonych z białkami G, to z czasem pojęcie uprzywilejowania zaczęto odnosić względem innych celów molekularnych postulując równocześnie preferencyjność i selektywność receptorową w obrębie badanej rodziny receptorów. Wykazano jednak, iż niektóre uprzywilejowane jednostki podstruktur wbudowane są również w struktury cząsteczek aktywujących receptory należące do różnych rodzin [113]. Co więcej, pojedynczy motyw strukturalny może być obecny w tysiącach różnorodnych substancji chemicznych włączając substancje pochodzenia naturalnego, których spektrum aktywności farmakologicznej może być

bardzo szerokie. Pojęcie motywu uprzywilejowanego zaczęło więc ewoluować, a jego początkowe określenie przypisywane względem struktur, które umożliwiają wiązanie się liganda z wybranym receptorem lub rodziną receptorów rozszerzono do przypadków, w których ułatwienie tworzenia kompleksu ligand-receptor niekoniecznie oznaczało selektywność receptorową. Przykładem mogą być dwie heterocykliczne struktury: chinolina oraz akrydyna, które występują ze znaczną częstotliwością wśród związków o różnym kierunku aktywności farmakologicznej [114].

3.6.3 MECHANIZM POWSTAWANIA ZJAWISKA UPRZYWILEJOWANIA BIOCHEMICZNEGO

Ze względu na spory potencjał i aplikacyjność uprzywilejowanych trzonów molekularnych w planowaniu i syntezie kombinatorycznych bibliotek związków o potencjalnym zastosowaniu farmakologicznym, a także implementację jako struktury wyjściowe w procesie racjonalnego projektowania farmaceutyków o pożądanej aktywności farmakologicznej (w tym także leków wieloreceptorowych) koncepcja motywów uprzywilejowanych jest żywo dyskutowana w kontekście zwiększenia skuteczności badań nad nowym lekiem. Statystyczna analiza wyników badań nad tworzeniem kompleksów małych cząsteczek chemicznych z jedenastoma różnymi celami molekularnymi ujawniła preferencyjność wiązania cząsteczek posiadających jako podstrukturę motyw bifenylu [87].

Spekulacje na temat preferencji tworzenia kompleksów ligand-receptor obejmują wpływ trzech czynników: (i) topologii małych cząsteczek chemicznych wiążących się specyficznie do (ii) preferowanych rejonów struktur receptora, a także (iii) czynnik ewolucyjny w naturalny sposób weryfikujący skuteczność wybranych połączeń atomów na szlakach biochemicznych (Rysunek 3.14). Szczególnie interesujący wydaje się ostatni aspekt. Wiele spośród struktur zdefiniowanych jako uprzywilejowane odnaleźć można jako podstruktury związków naturalnych i pochodzenia naturalnego [84,90,115]. Fragment molekularny 1,4-benzodiazepin-2-onu, prototyp uprzywilejowanych motywów strukturalnych, stanowi podstrukturę jednej z mykotoksyn – produktu naturalnego o nazwie *Asperlicin* [116], której cząsteczka stanowi strukturę wiodącą

w badaniach nad poszukiwaniem antagonistów cholecystokininy [117–120]. Być może Natura na drodze ewolucji dokonuje selekcji, w wyniku której preferowane są cząsteczki o wysokim powinowactwie do struktur białkowych. Bądź też preferencja ta jest naturalną konsekwencją procesów biochemicznych, w których struktury te zostały utworzone [79]?



Rysunek 3.14 Uprzywilejowanie struktur chemicznych w układach biologicznych.

Źródła biochemicznego uprzywilejowania, podobnie jak i jego szczegółowe mechanizmy działania w organizmach żywych, pozostają nadal w domenie spekulacji. Wśród badaczy dominuje przekonanie, iż za niespecyficzne wiązanie i powinowactwo małych cząsteczek chemicznych względem szerokiego spektrum struktur białkowych odpowiedzialna jest topologia trzonu molekularnego. Odpowiednie połączenie atomów – tworzących strukturę trzonu – umożliwia uzyskanie takich cech fizykochemicznych, które zapewniają wytworzenie niekowalencyjnych oddziaływań pomiędzy małą cząsteczką liganda a receptorem białkowym [79]. Być może dlatego większość struktur, które określono uprzywilejowanymi to fragmenty stanowiące aromatyczne, bądź heteroaromatyczne układy, których lipofilowość umożliwia tworzenie oddziaływań van der Waalsa z niepolarnymi resztami aminokwasowymi, na przykład fenyloalaniną hydrofobowych kieszeni receptora. Szczególną rolę w procesie tworzenia kompleksu ligand-receptor upatruje się w oddziaływaniach pierścieni

aromatycznych, tzw. π -stacking [121]. W ten sposób obecność aromatycznych układów w strukturze liganda może w znacznym stopniu tłumaczyć preferencyjność tych związków w tworzeniu kompleksów z szeregiem różnorodnych receptorów [87]. Topologia trzonu molekularnego w znacznym stopniu odpowiedzialna jest za niespecyficzność i powinowactwo związku względem struktur białkowych. Natomiast specyficzność wiązania zależna jest bezpośrednio od architektury podstawienia trzonu [75,122]. Proces rozpoznania molekularnego (ang. molecular recognition) wymaga więc komplementarności powierzchni molekularnej receptora i liganda. Przy czym komplementarność ta obejmuje zarówno uwarunkowania steryczne, jak i elektrostatyczne powierzchni obydwu cząsteczek. W szczególności komplementarność ligand-receptor odniesiono do rodziny receptorów sprzężonych z białkami G sugerując, iż miejsca aktywne tych makrocząsteczek posiadają niejako „zdolność rozpoznawania” odpowiedniej aranżacji łańcuchów bocznych trzonu molekularnego liganda [73,123].

3.6.3.1 UPRZYWILEJOWANIE STRUKTUR BIAŁKOWYCH

W literaturze odnaleźć można opinie sugerujące iż na preferencyjność tworzenia kompleksów ligand-receptor istotny wpływ ma nie tylko topologia małych cząsteczek chemicznych, ale także topografia struktur białkowych. Wiele białek posiada zdolność asocjacji szeregu różnorodnych ligandów – co ciekawe – asocjacja ta następuje jedynie w wybranych rejonach struktur białkowych [124,125]. Chemiczna różnorodność powierzchni białkowych jest ogromna – zdefiniowana przestrzenną aranżacją kombinacji dwudziestu aminokwasów. Pomimo tego swoiste rozpoznawanie antygenów następuje jedynie za pośrednictwem wybranych fragmentów wiążących obecnych w strukturze przeciwciał. Pozostała część powierzchni immunoglobulin w trakcie tworzenia kompleksu pozostaje praktycznie bezwładna [126]. W tym kontekście wiążący fragment przeciwciał można interpretować jako uprzywilejowany motyw strukturalny umożliwiający wiązanie szeregu antygenów [127].

3.7 LEKI AKTYWNE WZGLĘDEM WIRUSA HIV

Ludzki wirus niedoboru odporności HIV (ang. human immunodeficiency virus) należy do rodziny retrowirusów. Dotychczas poznano dwa typy wirusa: HIV-1 oraz HIV-2. Wirus HIV atakuje głównie limfocyty T, a w końcowym stadium zakażenia prowadzi do wyniszczenia układu immunologicznego i wykształcenia zespołu nabytego niedoboru odporności (ang. Acquired Immune Deficiency Syndrome, AIDS). Pacjenci z AIDS posiadają zwiększoną zapadalność na inne choroby, między innymi nowotwory, zapalenie płuc i grzybice. Powikłania te często prowadzą do śmierci pacjenta.

Obecnie w leczeniu infekcji wywołanych wirusem HIV stosowane są środki, które ze względu na odmienny mechanizm stymulacji farmakologicznej można podzielić na dwie podstawowe grupy. Do pierwszej zaliczane są inhibitory enzymów cyklu życiowego wirusa. Drugą stanowią związki, których efekt terapeutyczny polega na zakłócaniu interakcji wirusa HIV z otoczeniem. Systematykę środków terapeutycznych ze względu na cel molekularny z jakim oddziałują przedstawia Rysunek 3.15.



Rysunek 3.15 Klasyfikacja środków terapeutycznych stosowanych w leczeniu infekcji HIV/AIDS ze względu na mechanizm stymulacji farmakologicznej.

W obrębie wyróżnionych powyżej dwóch podstawowych rodzin wyróżnia się pięć głównych klas antyretrowirusowych środków terapeutycznych obecnie stosowanych w leczeniu infekcji HIV/AIDS:

1. inhibitory odwrotnej transkryptazy;

- a. nukleozydowe/nukleotydowe inhibitory odwrotnej transkryptazy;
- b. nienukleozydowe inhibitory odwrotnej transkryptazy;
2. inhibitory proteazy;
3. inhibitory fuzji;
4. inhibitory integrazy;
5. leki wieloskładnikowe, powstałe z połączenia środków należących do różnych klas.

Inhibitory odwrotnej transkryptazy (ang. reverse transcriptase inhibitors, RT) zakłócają kluczowy etap cyklu życiowego wirusa HIV, w trakcie którego enzym RT przeprowadza odwrotną transkrypcję pojedynczej nici RNA na podwójną nić wirusowego DNA. Nukleozydowe/nukleotydowe inhibitory enzymu RT stanowią jednostki budulcowe DNA, które po połączeniu z wirusowym DNA uniemożliwiają replikację wirusa w komórce. Natomiast nienukleozydowe inhibitory wiążąc się z enzymem RT uniemożliwiają transkrypcję wirusowego RNA do DNA.

Proteaza to enzym, który inicjuje proces rozcinania zainfekowanego łańcucha peptydowego na potomne, w pełni sprawne fragmenty wirusowe. Inhibitory enzymu proteazy zatrzymują więc proces dojrzewania funkcjonalnych białek namnażającego się wirusa HIV.

Inhibitory fuzji blokują wnikanie wirusa HIV przez błony do wnętrza komórki gospodarza, natomiast mechanizm inhibitorów integrazy polega na hamowaniu enzymu odpowiedzialnego za połączenie materiału genetycznego wirusa z komórkowym DNA.

Leki wieloskładnikowe, które łączą w jednym produkcie substancje czynne różnych klas środków terapeutycznych mają na celu zapobieganie powstawania lekooporności. Wysokoaktywna terapia antyretrowirusowa (ang. highly active antiretroviral therapy, HAART) polegająca na podawaniu jednocześnie kilku rodzajów farmaceutyków uważana jest obecnie za najbardziej skuteczną metodę leczenia infekcji HIV/AIDS.

W fazie badań znajduje się szósta klasa środków (ang. maturation inhibitors), których aktywność farmakologiczna polegałaby na zatrzymywaniu procesu tzw. pączkowania,

w trakcie którego niedojrzałe wirusy wydostają się na zewnątrz zainfekowanej komórki zabierając część membrany w celu utworzenia ochronnej otoczki.

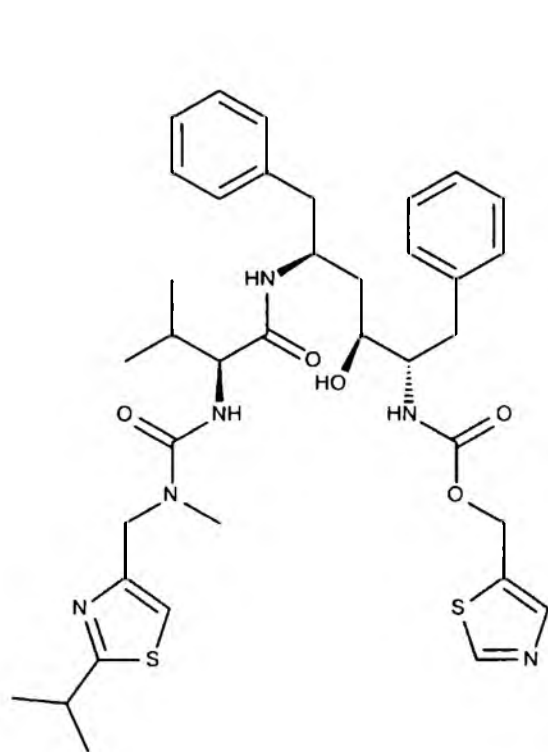
3.7.1 PRODUKTY LECZNICZE ZAREJESTROWANE W TERAPII HIV/AIDS⁴

Pula zatwierdzonych przez U. S. Food and Drug Administration leków przeznaczonych do terapii infekcji wywołanych wirusem HIV stanowi trzydzieści sześć produktów leczniczych. Rysunek 3.16 przedstawia ich nazwy handlowe oraz ilustruje przynależność specyfiku do wybranej klasy farmaceutyków zgodnie z mechanizmem stymulacji farmakologicznej substancji czynnej leku. Struktury chemiczne substancji czynnych przedstawiono na pięciu kolejnych rysunkach (Rysunek 3.17, Rysunek 3.18, Rysunek 3.19, Rysunek 3.20 oraz Rysunek 3.21).

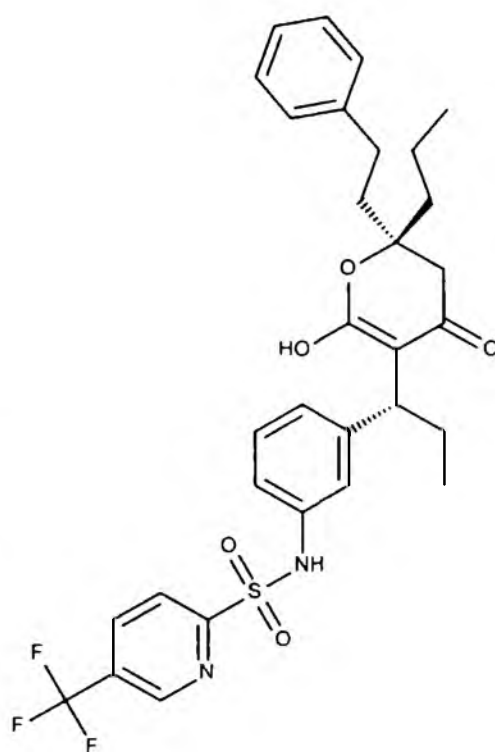
⁴ Rozdział obejmuje leki, które zostały zarejestrowane przez FDA do końca roku 2012.

Combivir	Hivid	Fortovase	Kaletra	Rescriptor	Stribild	
Trizivir	Viread	Viracept	Reyataz	Edurant	Atripla	
Videx EC	Ziagen	Lexiva	Norvir	Sustiva	Complera	
Epzicom	Zerit	Prezista	Agenerase	Viramune	Fuzeon	
Retrovir	Videx	Crixivan	Aptivus	Viramune XR	Selzentry	
Truvada	Emtriva	Epivir	Invirase	Intelence	Isentress	
HIV ISTI	EI	FI	McCP	NNRTI	PI	NRTI

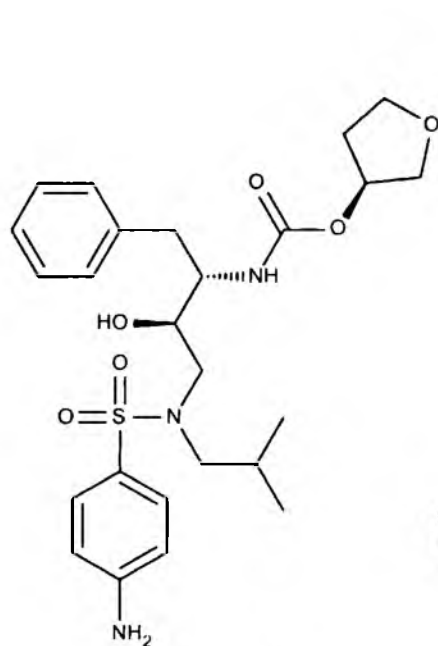
Rysunek 3.16 Leki zatwierdzone przez FDA w terapii przeciw HIV. Kolory kwadratów kodują przynależność produktu handlowego do danej klasy farmaceutyków zgodnie z mechanizmem stymulacji farmakologicznej substancji czynnej leku. NRTIs – nukleozydowe inhibitory odwrotnej transkryptazy (ang. nucleoside reverse transcriptase inhibitors); PIs – inhibitory proteazy (ang. protease inhibitors); NNRTIs – nienukleozydowe inhibitory odwrotnej transkryptazy (ang. nonnucleoside reverse transcriptase inhibitors); McCP – leki wieloskładnikowe (ang. multi-class combination products); FI – inhibitory fuzji, (ang. fusion inhibitors); EI – inhibitory wnikania wirusa do komórki gospodarza, blokery koreceptorów CCR5 (ang. entry inhibitors – CCR5 co-receptor antagonist); HIV ISTI – inhibitor integrazy (ang. HIV integrase strand transfer inhibitors).



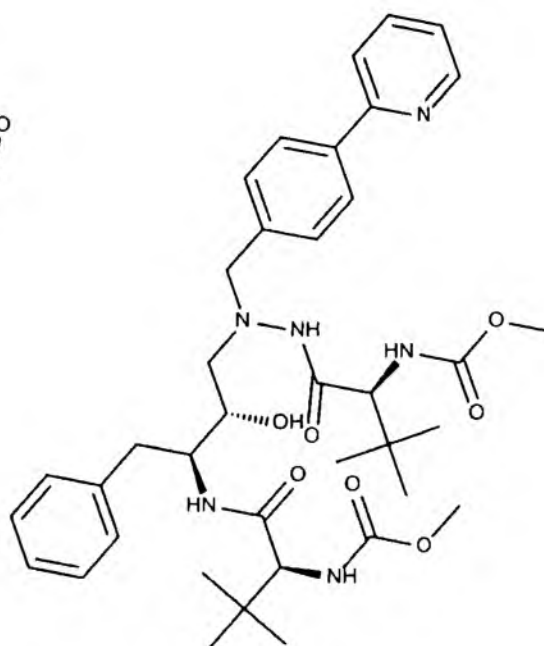
rytonawir



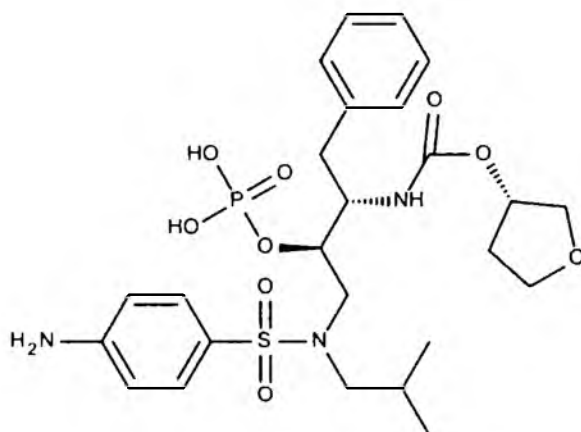
typranawir



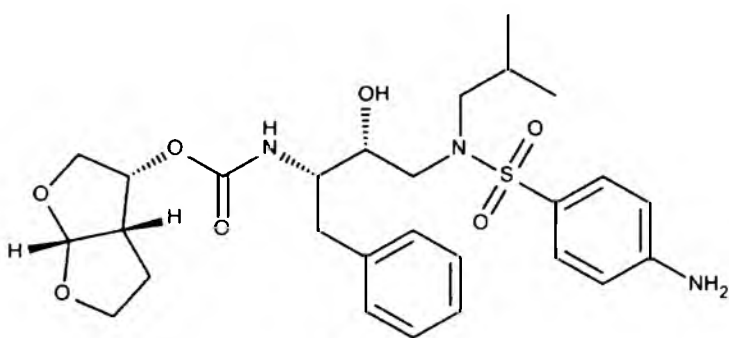
amprenawir



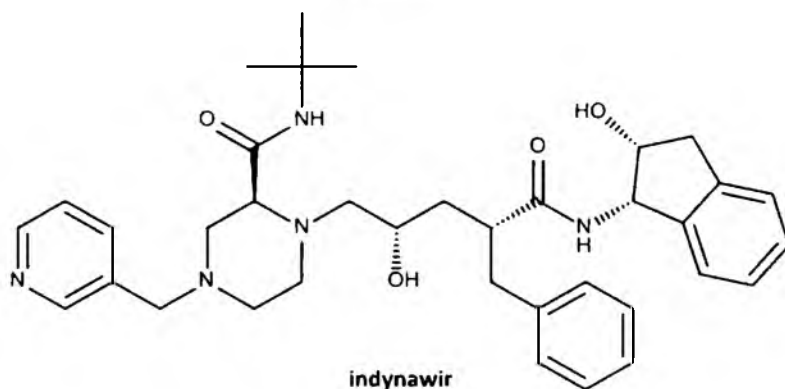
atazanawir



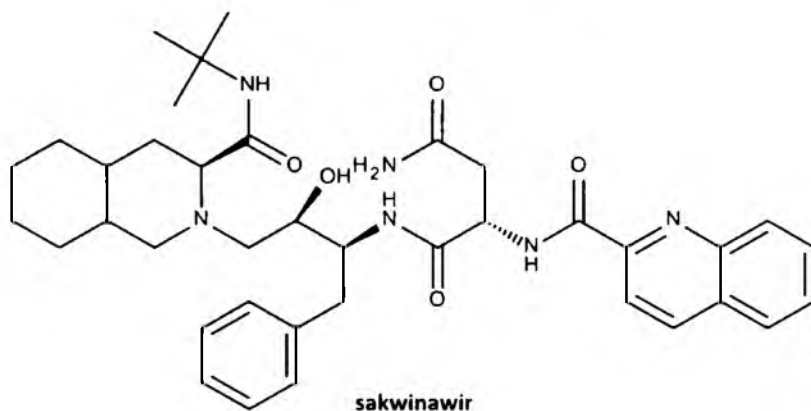
fosamprenawir



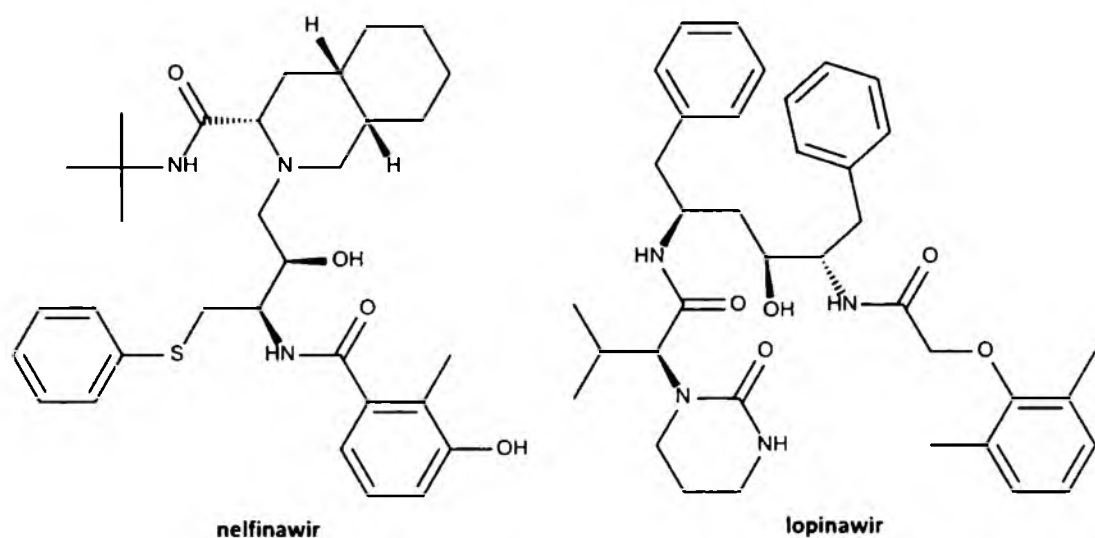
darunawir



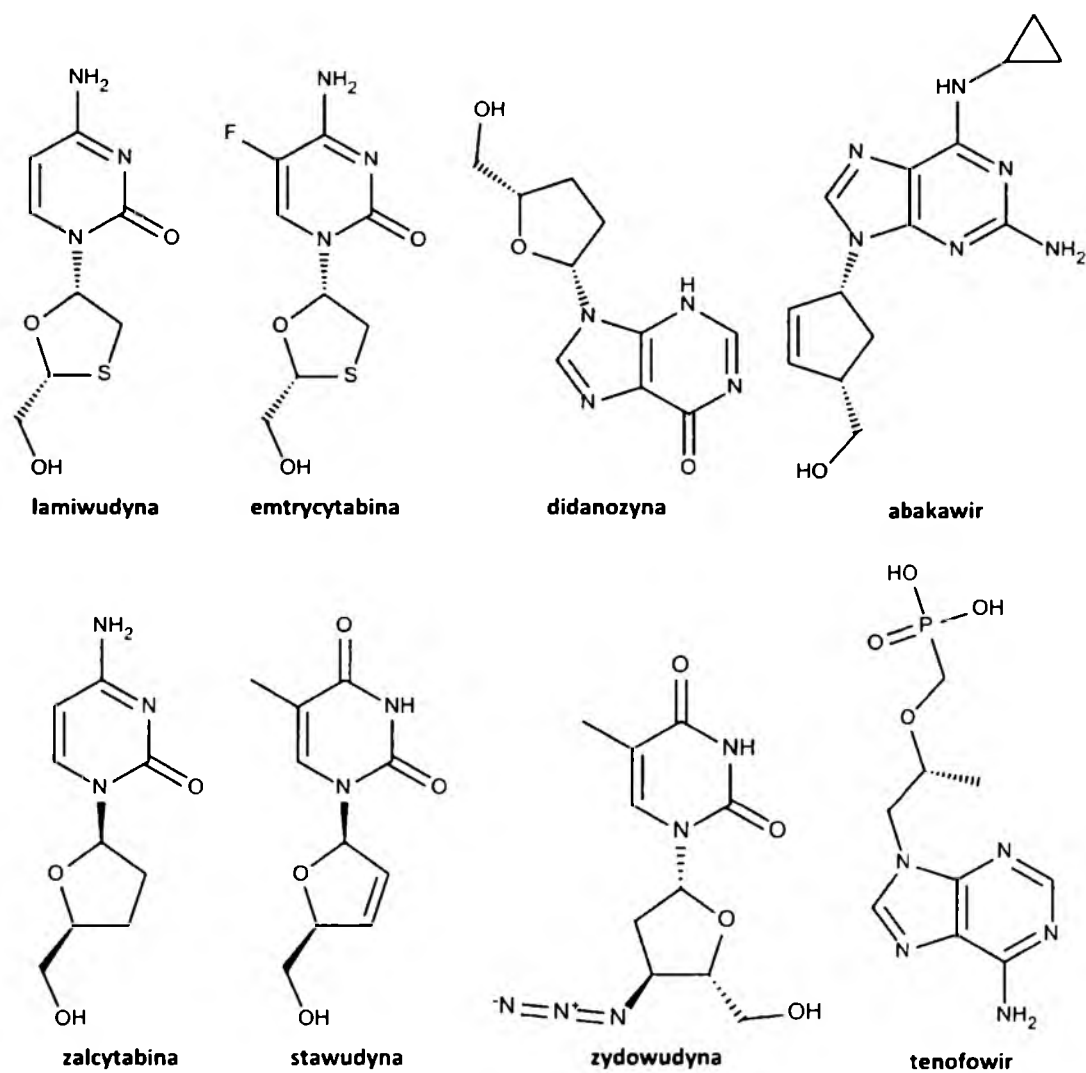
indynawir



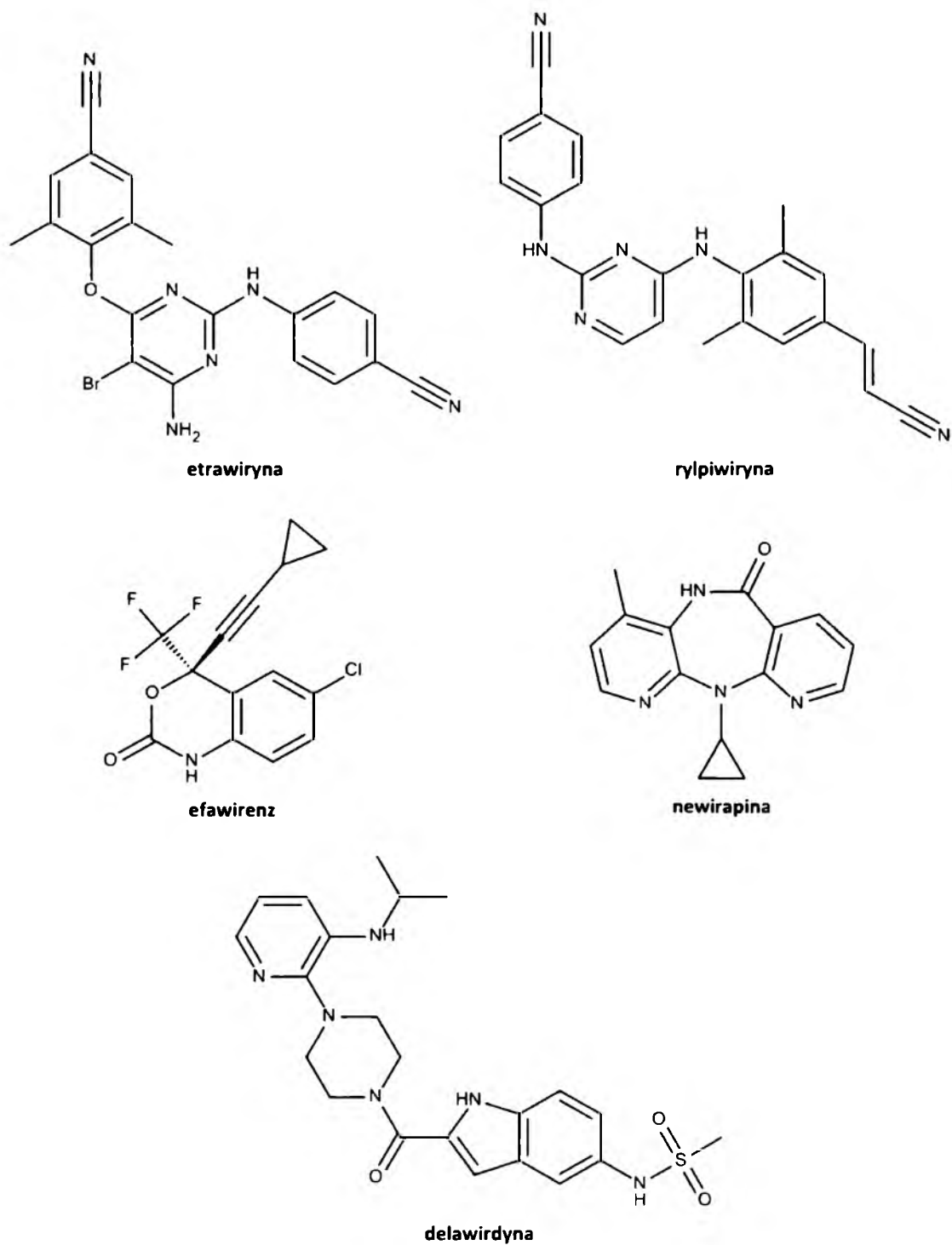
sakwinawir



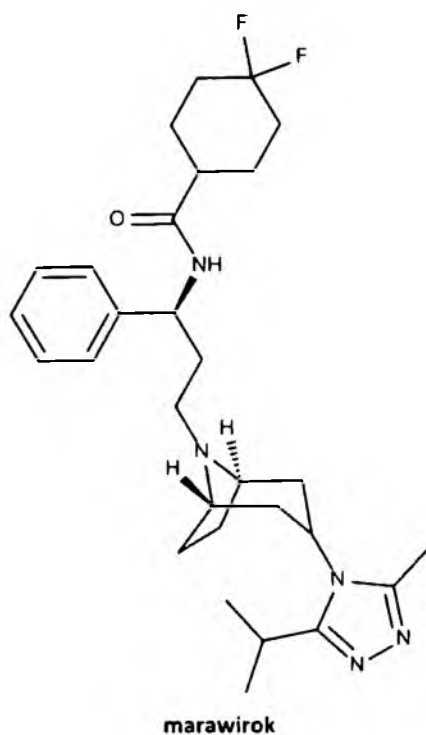
Rysunek 3.17 Inhibitory proteazy, PI.



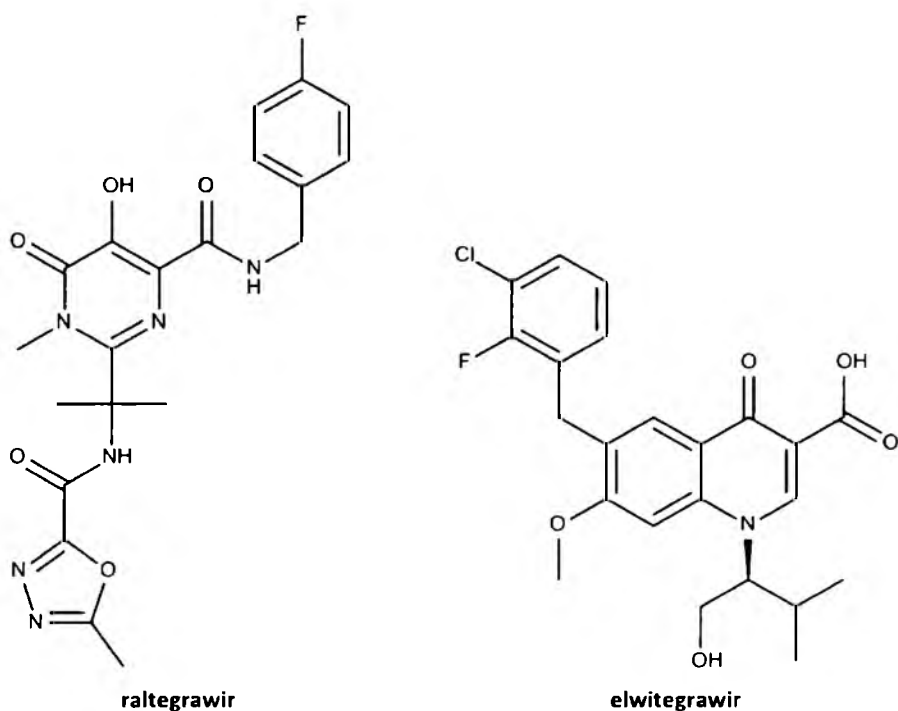
Rysunek 3.18 Nukleozydowe inhibitory odwrotnej transkryptazy, NRTI.



Rysunek 3.19 Nienukleozydowe inhibitory odwrotnej transkryptazy, NNRTI.



Rysunek 3.20 Inhibitor wnikania wirusa do komórki gospodarza – bloker koreceptorów CCR5, EI.



Rysunek 3.21 Inhibitory integrazy, HIV ISTI.

Preparat Isentress to jedyny⁵ lek stosowany w terapii przeciw HIV, który zawiera jako substancję czynną wyłącznie związek stanowiący inhibitor enzymu integrazy – raltegrawir. Raltegrawir został zatwierdzony przez FDA 12 października 2007 roku, a dwa miesiące później (20 grudnia 2007 roku) Komisja Europejska udzieliła pozwolenia firmie Merck Sharp & Dohme Limited na sprzedaż preparatu w krajach Unii Europejskiej. Enzym integrazy (IN) jest jednym z trzech kluczowych enzymów (obok proteazy i odwrotnej transkryptazy) w cyklu życiowym wirusa HIV-1. IN jest enzymem specyficznym wirusa o charakterze niespotykanym wśród enzymów gospodarza, co pozwala wiązać znaczne nadzieje z możliwością jej selektywnego blokowania, szczególnie w kontekście ostatnich doniesień na temat rozpoznania struktury trójwymiarowej IN [128–131].

Produkt leczniczy Stribild jest przeciwwirusowym lekiem wieloskładnikowym, posiadającym w swoim składzie cztery różne substancje czynne: elwitegrawir, kobicystat, emtrycytabinę i dizoproksyl tenofowiru (ang. multi-class combination products, McCP). Elwitegrawir – podobnie jak raltegrawir – jest inhibitorem IN. Kobicystat wzmacnia działanie elwitegrawiru poprzez przedłużenie czasu jego działania w organizmie. Dizoproksyl tenofowiru jest prolekiem tenofowiru. W organizmie jest on przekształcany w substancję czynną tenofowir. Tenofowir i emtrycytabina są nukleozydowymi inhibitorami odwrotnej transkryptazy. Struktury tych związków przedstawia Rysunek 3.18. Stribild został zatwierdzony przez FDA 27 sierpnia 2012 roku, a kilka miesięcy później (24 maja 2013 roku) Komisja Europejska udzieliła pozwolenia firmie Gilead Sciences International Limited na sprzedaż preparatu w krajach Unii Europejskiej.

⁵ Po sporządzeniu treści niniejszej rozprawy doktorskiej pojawiła się informacja o zatwierdzeniu przez FDA drugiego leku z zastosowaniem w terapii przeciw HIV, którego substancja czynna stanowi inhibitor enzymu integrazy. Dolutegravir (nazwa handlowa Tivicay) został zatwierdzony przez FDA 13 sierpnia 2013 roku.

4 BADANIA WŁASNE

Koncepcja lekotypii lub lekopodobieństwa (ang. drug-likeness) zakłada istnienie specyficznego zespołu cech strukturalnych i właściwości fizykochemicznych, odróżniających cząsteczki leków wśród innych cząsteczek organicznych. Reguła pięciu Lipińskiego [22], czy reguła Vebera [21] zostały zdefiniowane na podstawie badań wybranej populacji związków o potencjalnym zastosowaniu terapeutycznym. Wyniki analiz pozwoliły wskazać zakresy wartości właściwości molekularnych, które preferencyjnie spełniane są w zbiorze leków.

Pierwsza część opisu przeprowadzonych badań (Rozdział 4.2) obejmuje studium lekotypii wybranej podprzestrzeni chemicznej. Strategia badawcza ma na celu opracowanie takich metod eksploracji baz danych, które będą umożliwiały poszukiwanie wiedzy koniecznej dla znajdowania fragmentów molekularnych przydatnych do projektowania leków. Druga część opisu badań własnych (Rozdział 4.3) obejmuje badanie topologii lekotypii fragmentów molekularnych wydzielonych w strukturach cząsteczek leków stosowanych w terapii przeciw HIV. Innowacyjność przeprowadzonych badań wynika z ujęcia zagadnień dotyczących lekotypii w odniesieniu do wybranych fragmentów molekularnych. Trzecia część opisu badań (Rozdział 4.4) obejmuje charakterystykę opracowanego protokołu wirtualnego skriningu (VS), który wykorzystano do przeszukiwania baz komercyjnie dostępnych związków w celu poszukiwania nowych inhibitorów enzymu integrazy (IN) wirusa HIV.

4.1 NOTACJA ORAZ WYKAZÓW SKRÓTÓW I SYMBOLI STOSOWANYCH W OPISIE BADAŃ

Notacja stosowana w opisie badań:

f[indeks]	fragment molekularny opisany indeksem porządkowym;
ralt_[indeks]	fragment molekularny wydzielony w strukturze cząsteczki raltegrawiru w wyniku przeprowadzonej procedury fragmentacji dekrementacyjnej; indeks fragmentu koduje sekwencyjny zapis kierunku wykonanej instrukcji rozłączeń;
ralt-Q_[indeks]	fragment molekularny wydzielony w wyniku przeprowadzonej procedury fragmentacji dekrementacyjnej w uogólnionej reprezentacji struktury cząsteczki raltegrawiru, w której w miejscu atomu azotu, tlenu i fluoru dopuszczono wystąpienie dowolnego heteroatomu, co wskazuje wpisany symbol Q; indeks fragmentu koduje sekwencyjny zapis kierunku wykonanej instrukcji rozłączeń;
elwit_[indeks]	fragment molekularny wydzielony w strukturze cząsteczki elwitegrawiru w wyniku przeprowadzonej procedury fragmentacji dekrementacyjnej; indeks fragmentu koduje sekwencyjny zapis kierunku wykonanej instrukcji rozłączeń;
param_[indeks]	parametr skonstruowany na potrzeby przeprowadzanych analiz;
Q	dowolny heteroatom.

Wykaz często używanych skrótów:

IN	enzym integrazy wirusa HIV;
PS	uprzywilejowane motywy strukturalne (ang. privileged structures);

VS	skrining wirtualny (ang. virtual screening);
ML	uczenie maszynowe (ang. machine learning);
KlekFP	deskryptor daktyloskopowy Klekoty-Rotha;
FDA	U. S. Food and Drug Administration (Amerykańska Agencja ds. Żywności i Leków);
IC ₅₀	parametr szacujący stopień aktywności biologicznej poprzez określenie stężenia inhibitora, jakie jest konieczne dla zmniejszenia aktywności enzymu o 50%.

Symbole stosowane w opisie badań:

$\bar{x}$	średnia arytmetyczna;
σ	odchylenie standardowe w populacji;
<i>Mdn</i>	mediana.

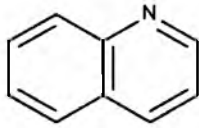
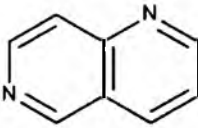
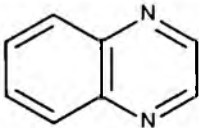
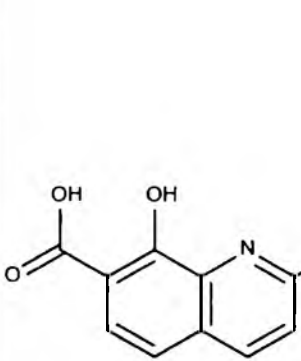
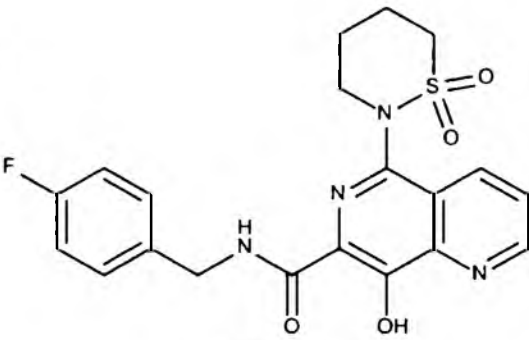
4.2 STUDIUM LEKOTYPII FRAGMENTU MONO- I DIAZANAFTALENOWEGO

Pierścień naftalenowy jest częstym fragmentem wielu ważnych związków chemicznych. Podobnie układ azanaftalenowy, na przykład chinolina, która występuje w cząsteczkach wielu leków (Rysunek 3.10 w Rozdziale 3.5). Antybiotyki których struktura zawiera fragment chinolinowy, są dobrym przykładem skuteczności tego układu w chemii medycznej [132]. Innym przykładem ewolucyjnej preferencji układu chinolinowego może być chinina, która jest aktywna w kilku układach receptorów. Podstruktura chinoliny pojawia się także w wielu innych klasach związków biologicznie aktywnych o właściwościach przeciwpierwotniakowych [133], grzybobójczych [134] i przeciwwirusowych [135]. Niektóre analogi chinolinowe wykazują również aktywność przeciwnowotworową [136].

Pochodne chinoliny stanowią również ważną grupę związków wykazujących aktywność wobec enzymu integrazy (IN) wirusa HIV [137–145]. Otrzymano szereg inhibitorów integrazy HIV opartych na układzie 1,6-naftyrydynowym [146,147] oraz 1,4-naftyrydynowym (chinoksalinowym) [148,149]. Tabela 4.1 przedstawia przykłady związków stanowiące inhibitory IN zawierające fragment mono- lub diazanaftalenowy jako element podstruktury. Związek L-870,810, zawierający ugrupowanie 8-hydroksy-1,6-naftyrydyno-7-karboksyamidowe, jest selektywnym inhibitorem etapu przeniesienia nici⁶ i wykazuje wartość IC_{50} wynoszącą 10 nM. Był on jednym z pierwszych inhibitorów IN, który osiągnął etap badań klinicznych. Jednak badania te wstrzymano ze względu na zaobserwowaną toksyczność u zwierząt [9]. Jednym z najbardziej aktywnych inhibitorów IN spośród pochodnych styrylochinolin jest związek o symbolu FZ-41, dla którego wartości IC_{50} procesu 3' terminacji oraz przeniesienia nici wynoszą odpowiednio 0,7 μ M oraz 1,7 μ M [138].

⁶ Enzym integrazy wirusa HIV katalizuje wprowadzenie materiału genetycznego wirusa do genomu gospodarza w dwóch następujących po sobie etapach. Pierwszy z nich stanowi proces 3' terminacji (ang. 3' end processing), a drugi etap przeniesienia nici (ang. strand transfer). Wyniki badania aktywności względem integrazy podawane są więc wraz ze wskazaniem odpowiedniego procesu.

Tabela 4.1 Inhibitory IN wirusa HIV oparte na układzie azanaftalenowym.

Fragment molekularny		
		
chinolina	1,6-naftyrydyna	1,4-naftyrydyna (chinoksalina)
[137–145]	[146,147]	[148,149]
Przykład związków		
		
FZ-41	L-870,810	

Studium lekotypii fragmentów molekularnych przeprowadzono dla ugrupowania mono- i diazanaftalenowego ze względu na skuteczność tego układu w chemii medycznej oraz potencjalne możliwości zastosowania do konstrukcji nowych leków stosowanych w terapii antyretrowirusowej. Dane liczbowe, stanowiące wyniki przeszukiwań chemicznych baz danych w oparciu o podstrukturę fragmentów azanaftalenowych, zebrano w tabelach w części eksperymentalnej (Rozdział 6.1). W poniższych rozdziałach (Rozdziały od 4.2.1 do 4.2.5) przedstawiono omówienie wyników.

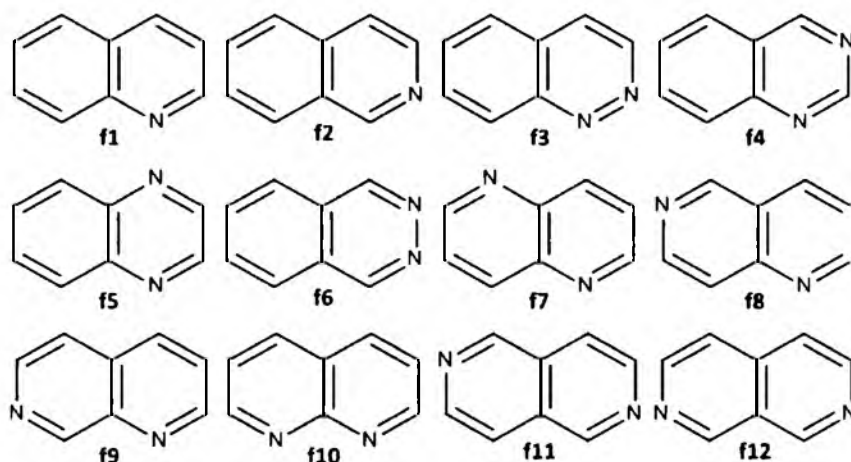
Badanie lekotypii fragmentów azanaftalenowych opisano w załączonej publikacji:

- **Załącznik 2a**

Polanski J., Kurczyk A., Bak A., Musiol R.; Privileged structures – dream or reality: preferential organization of azanaphthalene scaffold., *Curr. Med. Chem.*, 2012, 19(13), 1921-1945.

4.2.1 PRZESTRZEŃ CHEMICZNA I FARMAKOLOGICZNA AZANAFTALENÓW

Przestrzeń chemiczną azanaftalenów zdefiniowano jako pulę związków organicznych zawierających fragment mono- lub diazanaftalenowy jako podstrukturę. W ten sposób w obrębie badanej przestrzeni wskazano dwanaście zbiorów związków chemicznych posiadających jeden z fragmentów molekularnych **f1** - **f12** jako element podstruktury (Rysunek 4.1). Analizy porównawcze wyodrębnionych zbiorów cząsteczek przeprowadzono na podstawie informacji uzyskanych metodą eksploracji chemicznych baz danych.



Rysunek 4.1 Fragment mono- (**f1** i **f2**) i diazanaftalenowy (**f3** - **f12**).

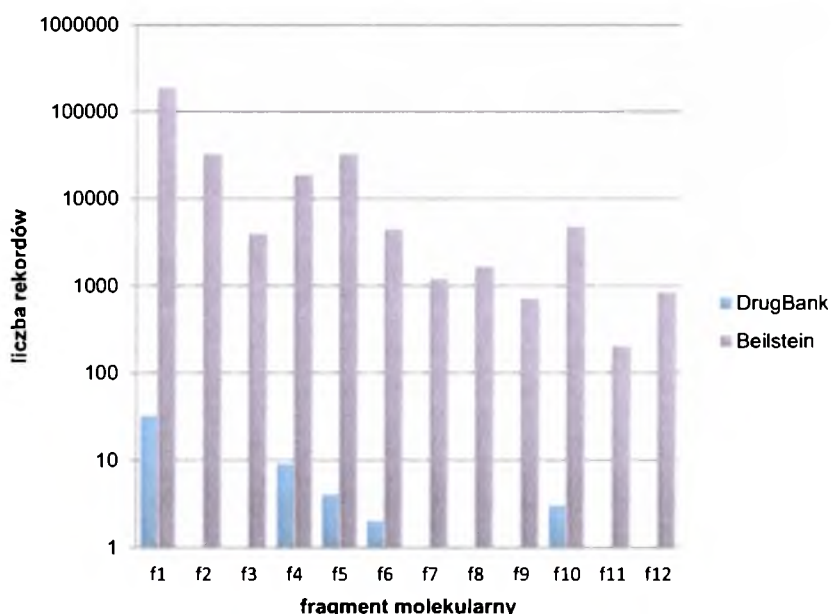
Podprzestrzeń chemiczna zajmowana przez związki zawierające jako podstrukturę fragment monoazanaftalenowy (**f1** i **f2**) jest licznie zdominowana przez pochodne chinoliny (Tabela 4.2). Wyniki przeszukiwania bazy Beilstein (B) i DrugBank (D) w oparciu o podstrukturę chinoliny (**f1**) oraz jej izomeru – izochinoliny (**f2**) wskazują, że fragment molekularny **f1** jest bardziej popularny od ugrupowania **f2**. Stosunek liczby rekordów **f1/f2** wskazuje liczebną przewagę chinolin względem izochinolin zarówno w bazie Beilstein, jak i w bazie DrugBank (Tabela 4.2, kolumna 4).

Tabela 4.2 Liczba rekordów uzyskana w wyniku przeszukiwania baz danych Beilstein oraz DrugBank w oparciu o podstrukturę chinoliny (**f1**) oraz izochinoliny (**f2**).

Fragment molekularny	Beilstein (B)	DrugBank (D)	f1/f2
chinolina (f1)	185470	32	5,77 (baza B)
izochinolina (f2)	32137	1	32 (baza D)

Dane na dzień 4 kwietnia 2010.

Także wśród związków zawierających jako podstrukturę fragment diazanaftalenowy (**f3** - **f12**) można zaobserwować wyraźne zróżnicowanie pod względem rozpowszechnienia w zależności od rozmieszczenia dwóch atomów azotu w układzie skondensowanych aromatycznych pierścieni naftalenowych. Rysunek 4.2 ilustruje statystykę występowania poszczególnych fragmentów mono- i diazanaftalenowych jako podstruktur znanych związków organicznych oraz leków. Przestrzeń chemiczna diazanaftalenów jest liczebnie zdominowana przez zbiór związków posiadających jako podstrukturę ugrupowanie oznaczone indeksem **f5**. Jednak liczebność tego zbioru jest prawie sześciokrotnie niższa od liczebności chinolin. Liczebność zbioru związków posiadających jako podstrukturę ugrupowanie **f4** i **f5** przekracza wartość 10 000, posiadających jako podstrukturę fragment molekularny **f3**, **f6**, **f8** lub **f10** mieści się w przedziale od 1000 do 10000 związków, natomiast liczebność zbiorów posiadających jako podstrukturę układ oznaczony indeksem **f9**, **f11**, **f12** nie przekracza wartości 1000. Zbiór związków z podstrukturą **f11** jest najmniej liczebny i zawiera zaledwie 198 struktur chemicznych. Liczebność tego zbioru jest ponad dziewięćset razy niższa od liczebności chinolin.

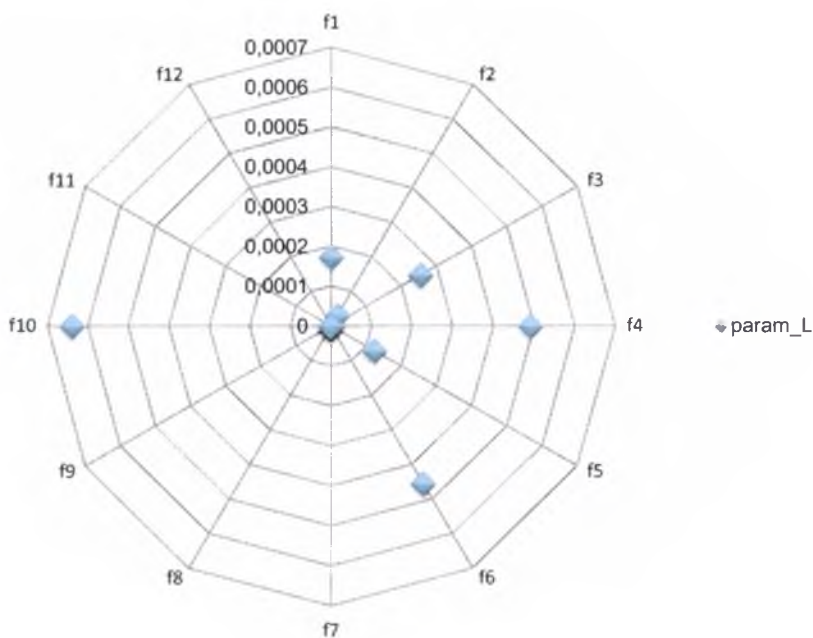


Rysunek 4.2 Statystyka rozpowszechnienia fragmentów azanaftalenowych jako podstruktur znanych związków organicznych oraz leków. Dane uzyskano metodą eksploracji bazy Beilstein oraz DrugBank. Liczba rekordów bazy DrugBank została ograniczona do leków posiadających status zatwierdzenia przynajmniej w jednym kraju. Dane na dzień 4 kwietnia 2010.

Parametr L (Równanie 4.1) pozwala oszacować udział przestrzeni farmakologicznej w odpowiadającej jej – odpowiednio zdefiniowanej – podprzestrzeni chemicznej. Parametr ten umożliwia obliczenie częstotliwości występowania fragmentów molekularnych **f1 - f12** jako podstruktur leków w odniesieniu do liczby znanych związków zawierających w swojej strukturze chemicznej dane ugrupowanie. Rysunek 4.3 przedstawia zestawienie wyliczonych wartości parametru L dla fragmentów **f1 - f12**.

$$\text{param_L} = \frac{\text{liczba leków}}{\text{liczba znanych związków organicznych}} \quad \text{Równanie 4.1}$$

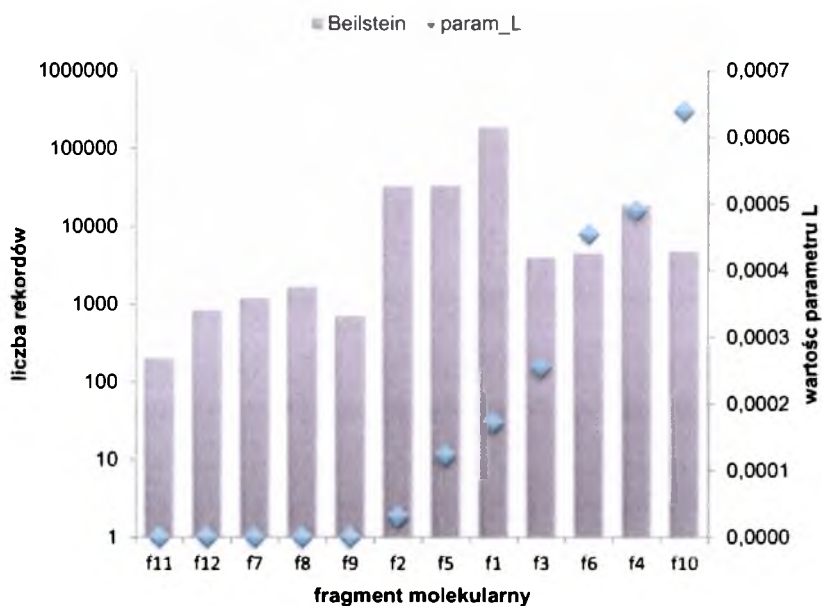
*Liczba leków i znanych związków organicznych odpowiada liczbie rekordów uzyskanych metodą eksploracji bazy DrugBank i Beilstein w oparciu o identyczne zapytanie strukturalne.



Rysunek 4.3 Wartości parametru L dla fragmentów molekularnych **f1 - f12**.

Uzyskane wyniki wskazują, że częstotliwość występowania danego fragmentu w populacji leków nie wykazuje korelacji ze stopniem jego rozpowszechnienia w opisaną dotychczas przestrzeni chemicznej. Najwyższą wartość parametru L posiada fragment oznaczony indeksem **f10** (trzeci w kolejności pod względem stopnia rozpowszechnienia w przestrzeni diazanaftalenów oraz piąty pod względem rozpowszechnienia wśród wszystkich dwunastu zbiorów). Fragment chinolinowy **f1** –

najbardziej rozpowszechniony pośród badanej puli związków – posiada wartość parametru L, która plasuje go na piątym miejscu w hierarchii dwunastu badanych zbiorów pod względem udziału przestrzeni farmakologicznej w podprzestrzeni chemicznej. Zbiory związków posiadających fragment molekularny **f7**, **f8**, **f9**, **f11**, **f12** nie tworzą przestrzeni farmakologicznej i równocześnie jest to pięć najmniej liczebnych zbiorów (zgodnie z danymi z bazy Beilstein i DrugBank). Rysunek 4.4 przedstawia zestawienie opisanych powyżej zależności.



Rysunek 4.4 Wartości parametru L wyliczone dla fragmentów molekularnych f1 - f12 w odniesieniu do liczebności zbioru.

Udział przestrzeni farmakologicznej w badanej podprzestrzeni chemicznej jest bardzo zróżnicowany w zależności od rozpatrywanego zbioru związków azanaftalenowych, ale równocześnie niezależny od liczebności zbioru. Zaobserwowane zróżnicowanie skłoniło mnie do przeprowadzenia analiz mających na celu scharakteryzowanie przestrzeni bioaktywności azanaftalenów.

4.2.2 PRZESTRZEŃ BIOAKTYWNOŚCI AZANAFTALENÓW

Przeprowadzono eksplorację bazy PubChem Compound, wyszukując związki zawierające jako podstrukturę fragment molekularny o topologii określonej indeksem

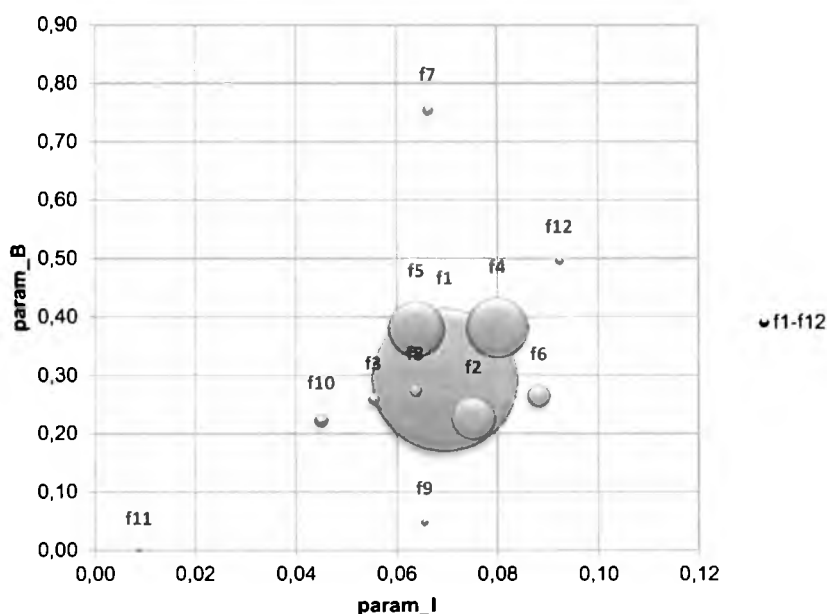
porządkowym kolejno od **f1** do **f12** (struktury fragmentów przedstawia Rysunek 4.1). W ten sposób wydzielono dwanaście zbiorów zdeponowanych w bazie struktur chemicznych. Następnie w obrębie każdego z dwunastu zestawów cząsteczek obliczono wskaźnik bioaktywności – parametr B (Równanie 4.2) oraz wskaźnik szacujący stopień „zainteresowania” daną strukturą w chemii medycznej (Równanie 4.3). Wartość parametru B wskazuje liczebny udział związków aktywnych *A* wśród związków, które zostały przebadane w testach biologicznych *T*. Natomiast wartość parametru I wskazuje jaka część związków chemicznych wyodrębnionego zbioru *W* została przebadana pod względem bioaktywności. Wysoka wartość parametru I świadczy o dużym zainteresowaniu wybraną grupą związków w chemii medycznej.

$$\text{param_B} = \frac{A}{T} \quad \text{Równanie 4.2}$$

$$\text{param_I} = \frac{T}{W} \quad \text{Równanie 4.3}$$

**A* - liczba związków aktywnych, *T* – liczba związków testowanych oraz *W* – liczba wszystkich uzyskanych rekordów.

Wartości parametru B wyliczone dla poszczególnych fragmentów molekularnych **f1-12** wskazują, że bioaktywna skuteczność danego ugrupowania strukturalnego nie wykazuje korelacji ze stopniem jego rozpowszechnienia w przestrzeni chemicznej związków mono- i diazanaftalenowych (Rysunek 4.5). Najwyższą wartość parametru B posiada fragment oznaczony indeksem **f7** (dziewiąty w kolejności pod względem stopnia rozpowszechnienia wśród wszystkich dwunastu zbiorów). Fragment chinolinowy **f1** – najbardziej rozpowszechniony pośród badanej puli związków – posiada wartość parametru B, która plasuje go na piątym miejscu w hierarchii dwunastu badanych zbiorów pod względem bioaktywnej skuteczności. Zbiór związków posiadających fragment molekularny **f11** nie tworzy przestrzeni bioaktywnej i równocześnie jest to najmniej liczebny zbiór (zgodnie z danymi z bazy PubChem).



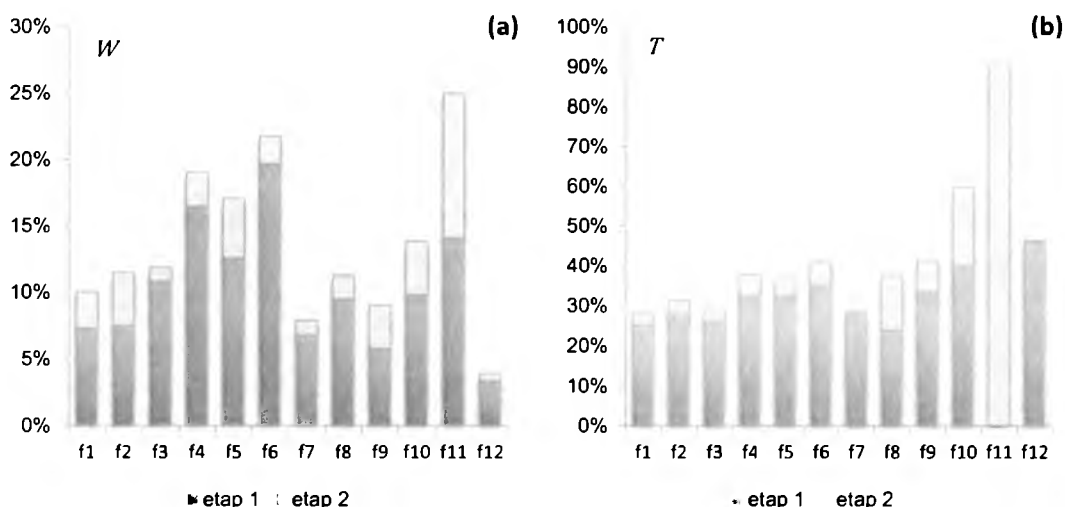
Rysunek 4.5 Przestrzeń bioaktywności fragmentu mono- i diazanaftalenowego. Parametr B oraz I zdefiniowano poprzez Równanie 4.2 oraz Równanie 4.3. Średnica kół odzwierciedla liczebność poszczególnych dwunastu zbiorów związków chemicznych posiadających jeden z fragmentów molekularnych f1 - f12 jako element podstruktury. Diagram sporządzono na podstawie danych uzyskanych metodą eksploracji bazy PubChem Compound. Dane na dzień 27 maja 2010.

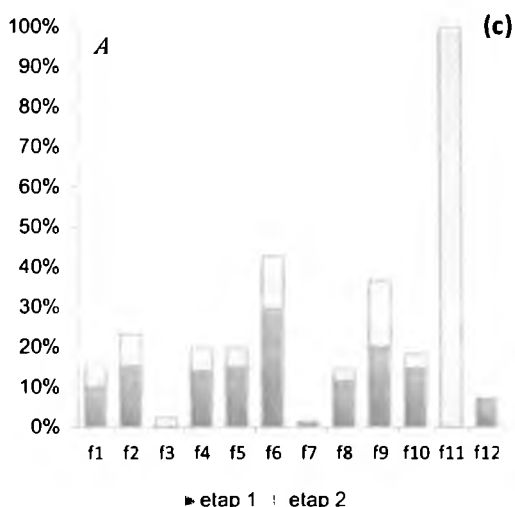
Baza PubChem jest dynamicznie rozwijana oraz systematycznie aktualizowana o nowe dane. Ciągły wzrost zdeponowanych w bazie struktur chemicznych oraz ewidencjonowanych informacji odnośnie badań i danych dotyczących aktywności biologicznej w sposób oczywisty przekłada się na wartość parametrów B oraz I. W związku z tym analizę przeprowadzano cyklicznie w okresie dwunastu miesięcy. Dzięki temu zarejestrowano zmiany wartości parametrów B oraz I, jakie następowały wraz z bieżącą aktualizacją bazy. W poniższym rozdziale zamieszczono opis dotyczący tej części badań (Rozdział 4.2.3).

4.2.3 ANALIZA DANYCH DOTYCZĄCYCH FRAGMENTU AZANAFTALENOWEGO EWIDENCJONOWANYCH W BAZIE PUBCHEM COMPOUND W FUNKCJI CZASU

Przeprowadzono dwunastomiesięczną obserwację zmienności danych zewidencjonowanych w bazie PubChem Compound. W okresie rocznej analizy

zaobserwowano swoistą ewolucję poszczególnych zbiorów związków azanaftalenowych, posiadających jeden z fragmentów molekularnych **f1** - **f12** jako element podstruktury. Wzrost liczby struktur chemicznych zdeponowanych w bazie PubChem oraz informacji odnośnie badań biologicznych przekładał się na wyliczaną wartość wskaźnika bioaktywności B (Równanie 4.2), przy czym zmiana ta następowała w różnym stopniu dla poszczególnych fragmentów **f1** - **12**. Rysunek 4.6 ilustruje zmiany w liczebności każdego ze zbiorów w okresie rocznego monitoringu. Wysokość słupków wskazuje procentowy przyrost liczebności danego zbioru po sześciomiesięcznym okresie analizy (etap 1) i następnie po kolejnych sześciu miesiącach (etap 2). Dla każdego z dwunastu badanych zbiorów wyliczono całkowity przyrost populacji związków W (Rysunek 4.6a) oraz przyrost zbioru związków przetestowanych T i aktywnych A (odpowiednio Rysunek 4.6b oraz Rysunek 4.6c).





Rysunek 4.6 Przyrost populacji badanych zbiorów. Diagram sporządzono na podstawie danych uzyskanych metodą eksploracji bazy PubChem Compound. Dane przedstawiono w dwóch seriach: po sześciomiesięcznym okresie analizy (etap 1) i następnie po kolejnych sześciu miesiącach (etap 2). *W* określa populację związków posiadających jeden z fragmentów molekularnych f1 - f12 jako element podstruktury, natomiast *T* oraz *A* wskazują zbiory związków testowanych i aktywnych wyodrębnione w każdym z dwunastu badanych zbiorów *W*.

Rysunek 4.6 ilustruje zmienność liczebności dla poszczególnych zbiorów posiadających fragment **f1 - 12** jako podstrukturę, natomiast Tabela 4.3 przedstawia zestawienie uzyskanych wartości w postaci szeregów przedziałowych. Spośród wszystkich badanych zbiorów w największym stopniu powiększył się zbiór pochodnych 2,6-naftyrydyny (Rysunek 4.6a, **f11**), drugim w kolejności pod względem liczebnej ekspansji jest zbiór pochodnych 2,3-naftyrydyny (Rysunek 4.6a, **f6**). Po okresie dwunastu miesięcy liczebność tych zbiorów wzrosła odpowiednio o 23,45% i 21,34%. Liczba związków przebadanych biologicznie (*T*) w obrębie zbioru **f11** (Rysunek 4.6b) wzrosła w tym czasie o 90,91%. W zbiorze **f6** liczba nowych przetestowanych struktur stanowiła 39,06% względem początkowego stanu. Stosunkowo większym zainteresowaniem pod względem liczby testowań cieszyły się związki oparte na trzonie molekularnym **f10** i **f12**. Liczba związków przetestowanych w obrębie tych dwóch zbiorów wzrosła odpowiednio o 51,87% oraz 46,31% względem stanu początkowego. Ekspansja zbiorów związków aktywnych (*A*) sięgnęła wartości 100% dla związków posiadających motyw strukturalny o indeksie **f11**. Jednak zbiór ten jako jedyny spośród dwunastu badanych pierwotnie nie posiadał związków aktywnych i stuprocentowy

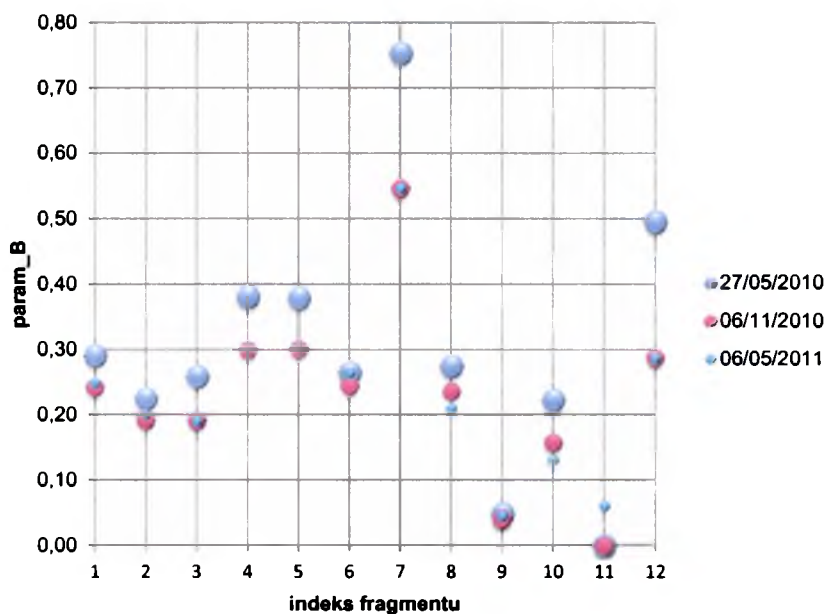
przyrost w tym przypadku oznacza pojawienie się zaledwie dwu związków posiadających aktywność biologiczną. Drugim w kolejności pod względem liczebnej ekspansji związków aktywnych jest zbiór **f6**. Podzbiór *A* dla tego motywu strukturalnego zwiększył swoją liczebność o 38,75%. Wartość ta odpowiada liczbie 174 nowych związków o potwierdzonej bioaktywności. Dane statystyczne dotyczące zmian liczebności zbiorów *W*, *T* oraz *A* wskazują, że największy wzrost danych odnotowano dla zbioru związków testowanych *T* – średnio o 41% (Tabela 4.3). Wartości odchylenia standardowego wyliczone dla wzrostu liczebności dwunastu zbiorów *W* oraz wyodrębnionych pośród każdego z nich podzbiorów *T* oraz *A* wskazują, że największe zróżnicowanie w wielkości wzrostu w zależności od rozpatrywanego fragmentu **f1 - 12** występuje dla związków aktywnych (*A*). Dane liczbowe sugerują więc, iż bioaktywna skuteczność związków azanaftalenów wykazuje silne preferencje w zależności od topologii rozpatrywanego fragmentu.

Tabela 4.3 Statystyka ewolucyjnych preferencji związków azanaftalenowych.

Wzrost liczebności zbioru	Zbiór <i>W</i>	Zbiór <i>T</i>	Zbiór <i>A</i>
poniżej 20%	f1 - f5, f7 - f10, f12		f1, f3 - f5, f7 - f8, f10, f12
od 20% do 39%	f6, f11	f1-9	f2, f6, f9
od 40% do 59%		f10, f12	
od 60% do 79%			
od 80% do 100%		f11	f11
Statystyka			
$\bar{x}$	13%	41%	24%
σ	6%	17%	25%
<i>Mdn</i>	11%	36%	19%

Opisana powyżej zmienność danych ewidencjonowanych w bazie PubChem Compound miała swoje przełożenie na zmianę wartości wyliczanego wskaźnika bioaktywności *B* (Równanie 4.2) fragmentu. Rysunek 4.7 przedstawia wykres zmian wartości parametru *B*, jakie nastąpiły w okresie rocznej analizy. Wartości wyliczono dla każdego z dwunastu zbiorów związków chemicznych posiadających jeden z fragmentów molekularnych **f1 - 12** jako element podstruktury. Wzrost parametru *B* zaobserwowano tylko dla fragmentu o topologii oznaczonej indeksem **f11**. Dla pozostałych jedenastu fragmentów nastąpił spadek wartości omawianego wskaźnika. Zmiany nastąpiły

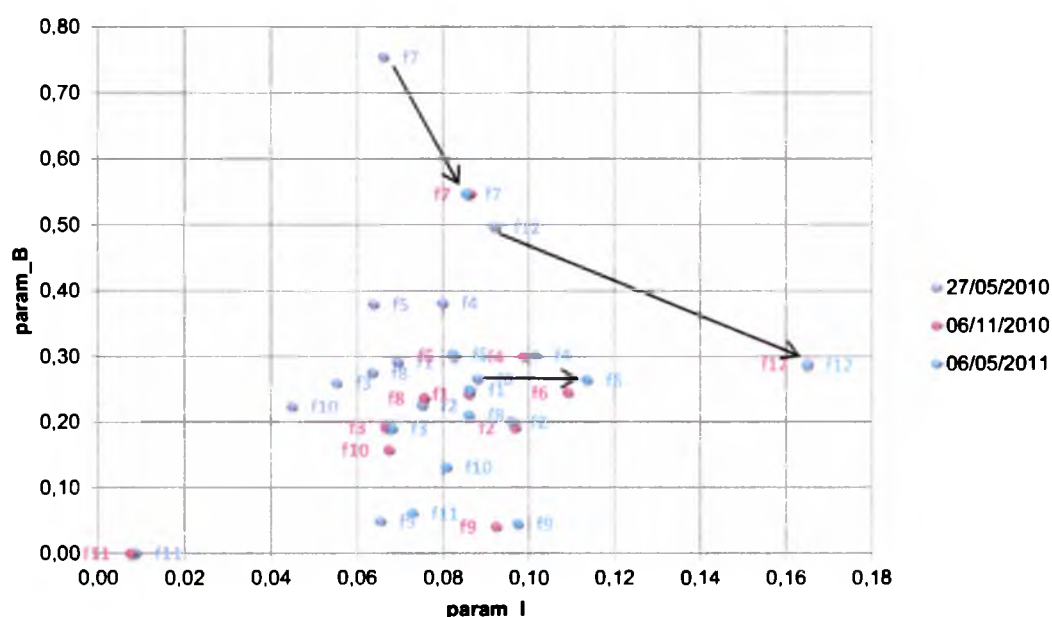
głównie w pierwszym etapie analizy (po okresie sześciu miesięcy). W drugim etapie (po kolejnych sześciu miesiącach) wartości wyliczanego wskaźnika zmieniły się w nieznacznym stopniu. Największy spadek wartości parametru B odnotowano dla fragmentu 2,7-naftyrydyny oznaczonego indeksem **f12**. Dla fragmentu 1,5-naftyrydyny (**f7**) spadek wartości omawianego wskaźnika był tylko nieznacznie mniejszy. Dla pozostałych dziewięciu badanych zbiorów związków spadek wartości parametru B był już zdecydowanie mniejszy.



Rysunek 4.7 Zmiana wartości parametru B w czasie. Diagram sporządzono na podstawie danych uzyskanych metodą eksploracji bazy PubChem Compound. Dane przedstawiono w trzech seriach kolejno na dzień 27 maja 2010, 6 listopada 2010 oraz 6 maja 2011.

Pełny obraz zmienności przestrzeni bioaktywności azanaftalenów uzyskano zestawiając zmiany parametru B w odniesieniu do wskaźnika szacującego stopień zainteresowania I (Równanie 4.3) daną grupą związków chemicznych. Rysunek 4.8 przedstawia wykres zmian wartości parametrów B oraz I, jakie nastąpiły w okresie rocznej analizy. Uzyskana swoista trajektoria punktów diagramu wskazuje, że zmiany parametrów dla poszczególnych zbiorów związków chemicznych posiadających jeden z fragmentów molekularnych **f1** - **12** jako element podstruktury następowały w bardzo różnym stopniu. Na przykład trajektoria wyznaczona dla fragmentu **f7** jest znacznie bardziej „stroma” niż trajektoria fragmentu **f12**, czy **f6**. Co oznacza, iż dla fragmentu **f7**

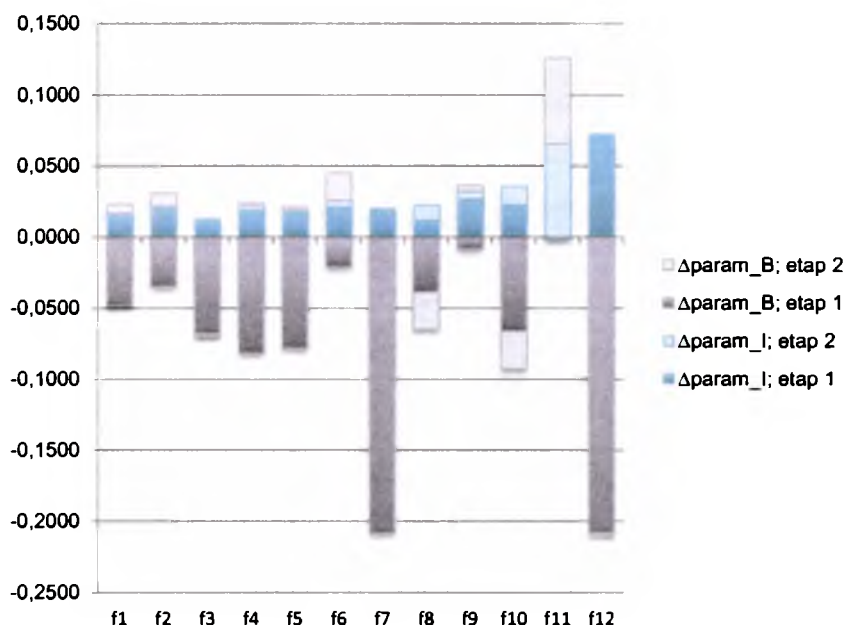
nastąpiła znacznie większa zmiana wartości parametru B względem zmiany wartości parametru I w stosunku do wymienionych fragmentów **f12** i **f6**. Innymi słowy – bioaktywna skuteczność dla fragmentu **f7** uległa znacznemu obniżeniu mimo stosunkowej niewielkiego wzrostu liczby danych odnośnie nowych testowań tej grupy związków. Wartość parametru B dla fragmentu molekularnego **f12** uległa największemu spadkowi spośród badanych zbiorów związków, jednak trajektoria zmian wyznaczona dla tego fragmentu nie jest bardzo stroma. Efekt ten jest rezultatem sporego wzrostu liczby danych odnośnie testowań związków posiadających ugrupowanie **f12** jako podstrukturę. Najmniejsza zmiana wartości parametru B w okresie rocznej analizy nastąpiła dla fragmentu **f6**.



Rysunek 4.8 Trajektoria zmian wartości parametrów B oraz I w czasie. Diagram sporządzono na podstawie danych uzyskanych metodą eksploracji bazy PubChem Compound. Dane przedstawiono w trzech seriach kolejno na dzień 27 maja 2010, 6 listopada 2010 oraz 6 maja 2011.

Rysunek 4.9 stanowi zestawienie wyników uzyskanych w trakcie analizy danych, dotyczących fragmentu azanaftalenowego ewidencjonowanych w eksplorowanej bazie w okresie rocznego monitoringu. Słupki wykresu nie wskazują bezwzględnych zmian wartości parametrów B oraz I, ale przedstawiają proporcje pomiędzy wielkością zmiany obydwu wskaźników w obrębie każdego z dwunastu badanych zbiorów. Największy spadek wartości parametru B nastąpił dla fragmentu **f12** i **f7**. Przy czym dla pierwszego

motywu strukturalnego fakt ten związany jest ze stosunkowo dużym wzrostem danych odnośnie testowań. Z kolei dla fragmentu **f7** wykonano stosunkowo mało nowych oznaczeń biologicznych, a te które wykonano nie przyniosły potwierdzenia aktywności – stąd wysoki spadek wskaźnika B przy stosunkowo niskim stopniu eksploracji podprzestrzeni chemicznej.



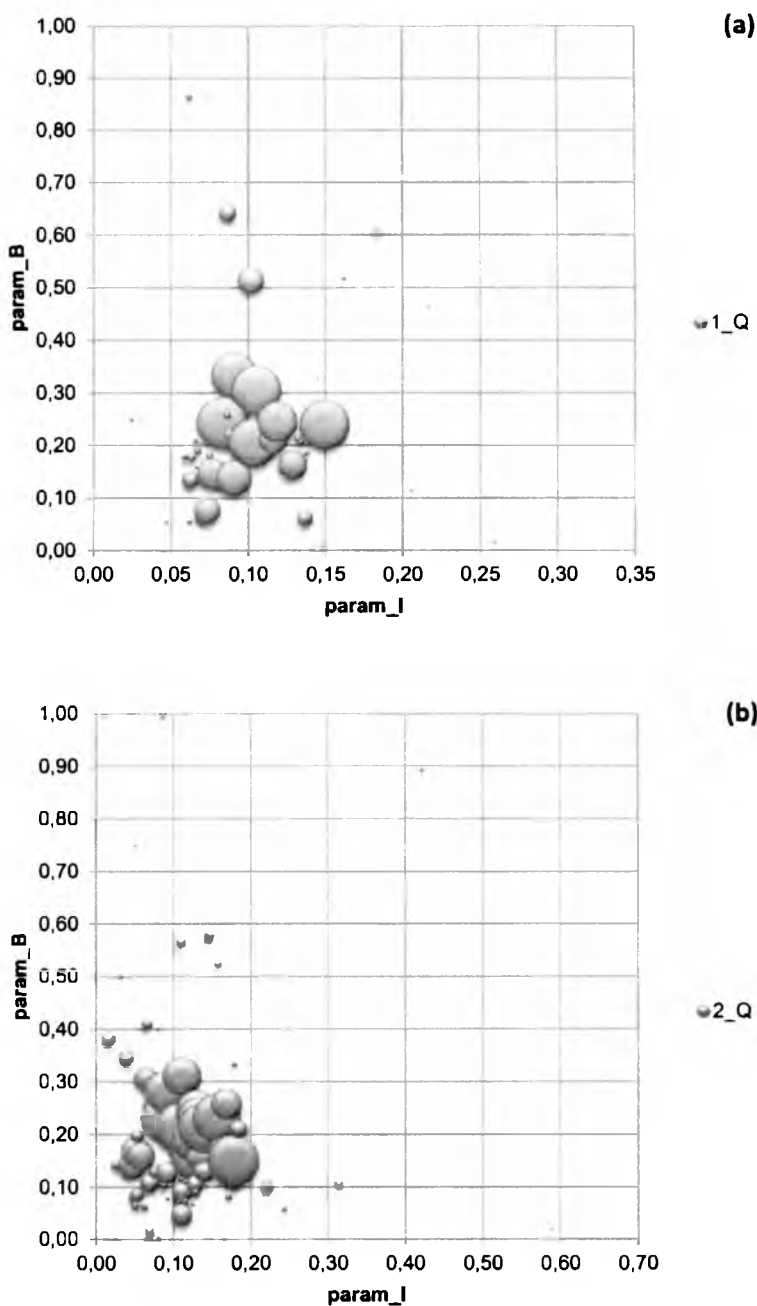
Rysunek 4.9 Zestawienie wyników uzyskanych w okresie rocznego monitoringu danych, dotyczących fragmentu azanaftalenowego, ewidencjonowanych w bazie PubChem Compound.

Zmiany następujące w obrębie każdego z dwunastu badanych zbiorów są obserwowane wskutek wzrostu liczby danych zewidencjonowanych w bazie. Jednak analiza porównawcza tych zmian dla poszczególnych zbiorów wskazuje niezmienną zaobserwowanych trendów i tendencji w czasie. Fragment 1,5-naftyrydyny zachowuje najwyższą wartość wskaźnika bioaktywnej skuteczności B w całym okresie rocznej analizy. Preferencje tworzenia przestrzeni bioaktywności fragmentu azanaftalenowego są wyraźnie zróżnicowane w zależności od topologii rozpatrywanego fragmentu. Preferencje te pozostały niezmiennie w okresie rocznej analizy danych dotyczących badanego fragmentu.

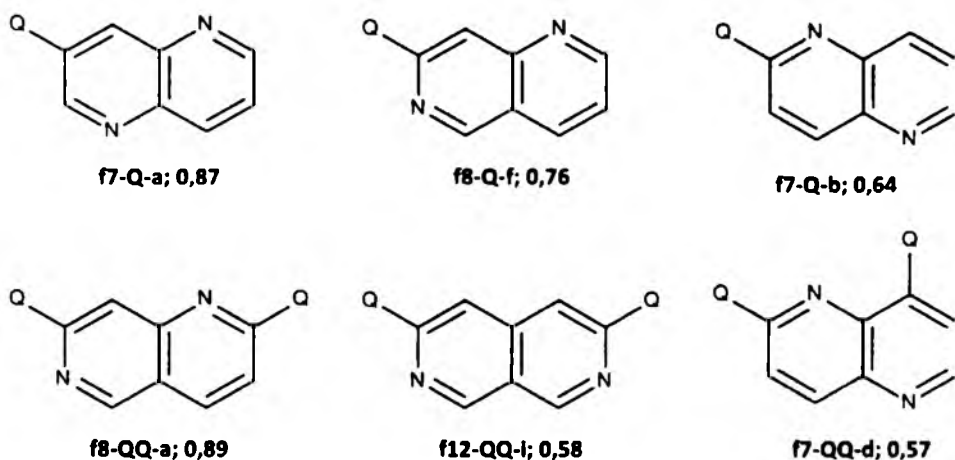
4.2.4 UPRZYWILEJOWANA ARCHITEKTURA PODSTAWIEŃ AZANAFTALENÓW

Studium bioaktywnej skuteczności przeprowadzono, formułując zapytanie nie tylko w oparciu o sam trzon molekularny **f1 - 12**, ale także analizowano uprzywilejowanie architektury podstawień dwunastu azanaftalenowych motywów.

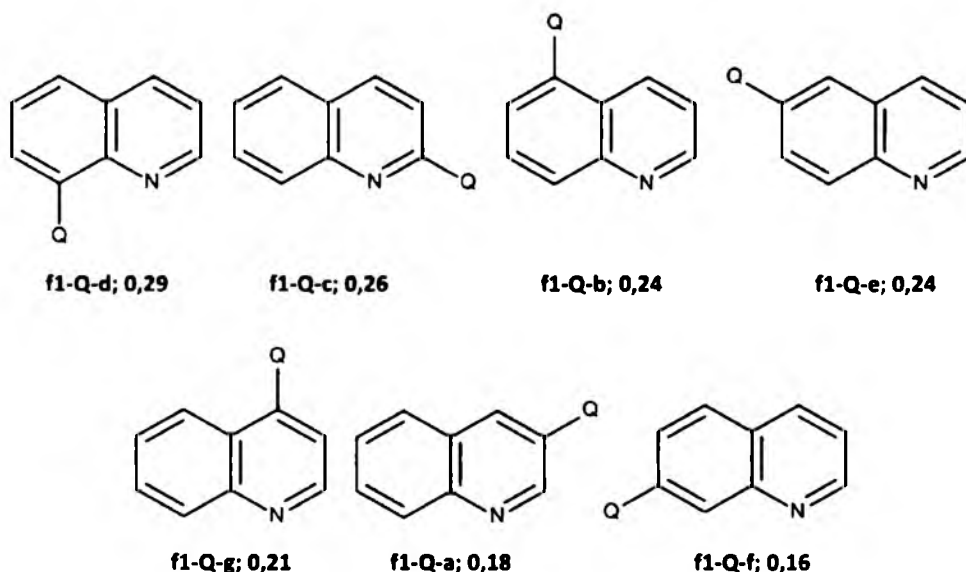
W tym celu przeprowadzono eksplorację bazy PubChem Compound, wyszukując związki zawierające jako podstrukturę fragment molekularny o topologii określonej indeksem porządkowym kolejno od **f1** do **f12** (struktury fragmentów przedstawia Rysunek 4.1) podstawiony jednym bądź dwoma heteroatomami. Istnieje 56 możliwych kombinacji podstawień dwunastu fragmentów tylko w jednym miejscu oraz 156 kombinacji podstawień równocześnie w dwóch miejscach. W ten sposób wydzielono 205 zbiorów zdeponowanych w bazie struktur chemicznych (56 monopodstawionych oraz 149 dipodstawionych zbiorów azanaftalenów; 7 spośród 156 możliwych kombinacji dipodstawień nie posiadały rekordów w bazie). Następnie w obrębie każdego z zestawów cząsteczek obliczono wskaźnik bioaktywności – parametr B (Równanie 4.2) oraz wskaźnik I szacujący stopień „zainteresowania” daną strukturą w chemii medycznej (Równanie 4.3). Rysunek 4.10 przedstawia uzyskane wyniki dla azanaftalenów podstawionych jednym (Rysunek 4.10a) bądź dwoma (Rysunek 4.10b) heteroatomami. Topologia niektórych podstawień jest mało rozpowszechniona w populacji badanych związków – co ilustrują koła o małej średnicy. Z reguły zbiory o niskiej liczebności rozkładają się na obrzeżach wykresu, przyjmując wartości graniczne parametru B równe 0 i 100. Są to zbiory, w których przetestowano zaledwie kilka związków chemicznych nie uzyskując potwierdzenia bioaktywności dla żadnego z nich (wartość 0 parametru B) bądź uzyskując dla wszystkich wskazanie aktywności biologicznej. Rysunek 4.11 przedstawia architekturę podstawień dla której uzyskano najwyższe wartości parametru B. Przy czym wyboru dokonano tylko spośród zbiorów, dla których liczba związków przetestowanych wynosiła co najmniej 50. Natomiast Rysunek 4.12 ilustruje rozkład wyliczonych wartości parametru B dla fragmentu chinolinowego **f1** w zależności od zdefiniowanej ornamentacji trzonu molekularnego.



Rysunek 4.10 Przestrzeń bioaktywności fragmentu mono- i diazanaftalenowego podstawionego jednym (a) i dwoma (b) heteroatomami. Parametr B oraz I zdefiniowano poprzez Równanie 4.2 oraz Równanie 4.3. Średnica kół odzwierciedla liczebność poszczególnych zbiorów struktur chemicznych. Diagram sporządzono na podstawie danych uzyskanych metodą eksploracji bazy PubChem Compound. Dane na dzień 5 listopada 2010.



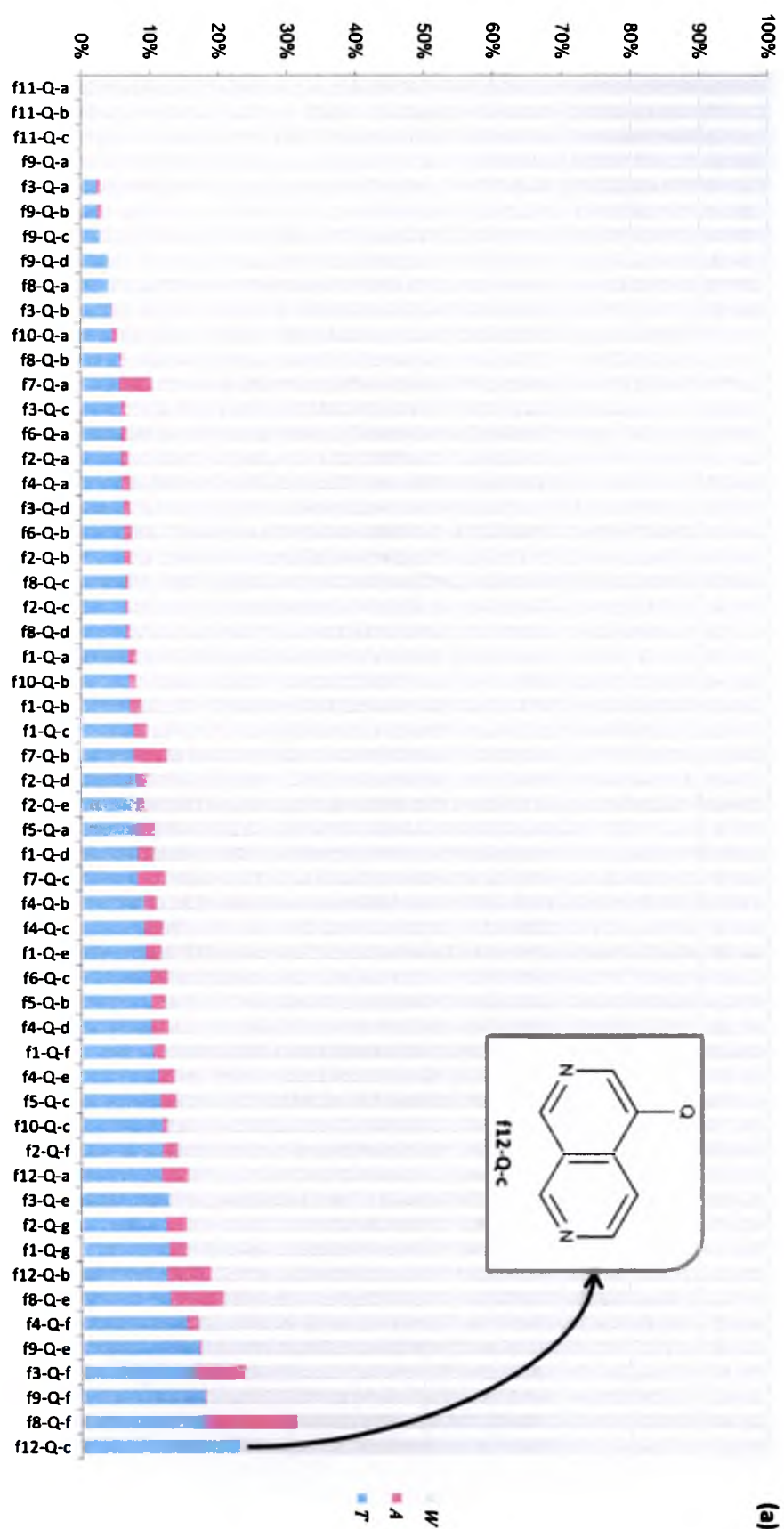
Rysunek 4.11 Uprzywilejowana architektura podstawień fragmentu azanaftalenowego. Górny rząd przedstawia trzy topologie monopodstawień, natomiast dolny rząd ilustruje trzy kompozycje dipodstawień, dla których uzyskano najwyższe wartości parametru B. Q oznacza dowolny heteroatom, a podane pod strukturami liczby oznaczają indeks fragmentu oraz wyliczoną wartość parametru B.

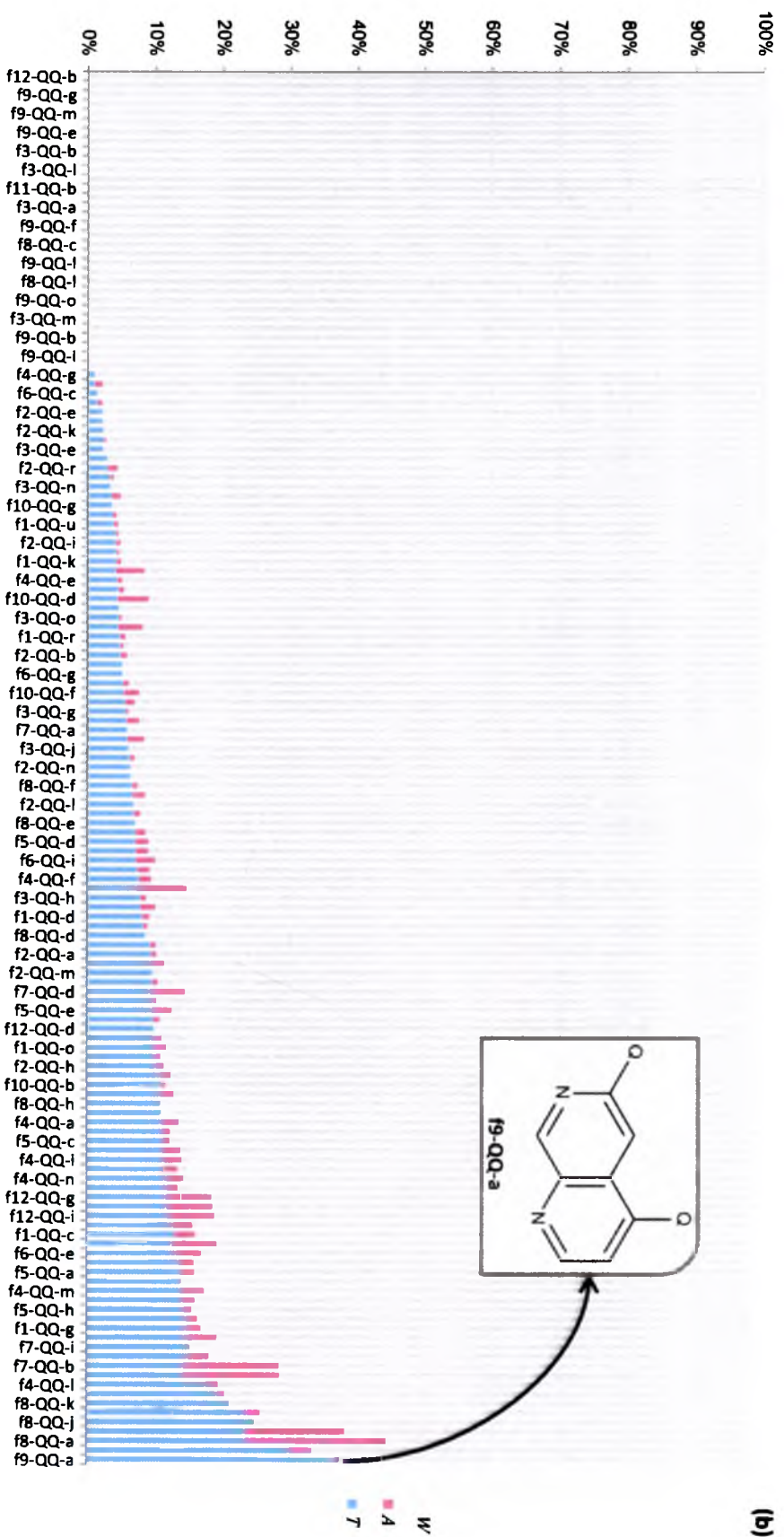


Rysunek 4.12 Przestrzeń bioaktywności fragmentu chinolinowego f1 w zależności od architektury podstawienia trzonu molekularnego. Q oznacza dowolny heteroatom, a podane pod strukturami liczby oznaczają indeks fragmentu oraz wyliczoną wartość parametru B.

Rysunek 4.13 przedstawia charakterystykę poszczególnych architektur podstawienia azanaftalenów jednym (Rysunek 4.13a) bądź dwoma heteroatomami (Rysunek 4.13b) uzyskaną metodą eksploracji bazy PubChem Compound. Słupki wykresu przedstawiają proporcje pomiędzy liczebnością zbiorów *W*, *T*, *A* w zależności od ornamentacji trzonu

molekularnego. Pośród 56 różnych kompozycji monopodstawień (Rysunek 4.13a) oraz 149 kompozycji dipodstawień (Rysunek 4.13b) można wskazać fragmenty molekularne, dla których liczebność zbioru związków aktywnych A jest stosunkowo duża względem liczebności całego zbioru W . Są to motywy strukturalne, dla których wartości parametru B były najwyższe – topologię tych ugrupowań przedstawia Rysunek 4.11. W opozycji do nich można wskazać takie architektury podstawień trzonu azanaftalenowego, dla których liczebność zbioru związków aktywnych A jest względnie niewielka pomimo dużej liczebności całkowitej populacji W i podzbioru związków testowanych T , na przykład fragment molekularny **f9-Q-f**, czy **f12-Q-c** (Rysunek 4.13a) oraz **f9-QQ-a** (Rysunek 4.13b).





Rysunek 4.13 Profil uprzywilejowania architektury podstawień fragmentu azanafalenowego jednym (a) i dwoma (b) heteroatomami. Diagram sporządzono na podstawie danych uzyskanych metodą eksploracji bazy PubChem Compound. W/ określa populację związków posiadających jeden z fragmentów molekularnych f1-12 jako element podstawki o określonej ornamentacji podstawników, natomiast T oraz A wskazują zbiory związków testowanych i aktywnych wydodrężnione w każdym z badanych zbiorów W. Dane na dzień 5 listopada 2010.

4.2.5 BADANIE WSKAŹNIKÓW DOSTĘPNOŚCI SYNTETYCZNEJ

Dostępność syntetyczna jest jednym z kluczowych wskaźników wstępnej priorytetyzacji związków w procesie projektowania leków. Wskazaniem jest wykluczenie z badanego zbioru takich struktur chemicznych, których synteza będzie w znacznym stopniu utrudniona, bądź też niewykonalna w warunkach laboratoryjnych czy przemysłowych. Istnieje koncepcyjna różnica pomiędzy pojęciem wykonalności syntezy (ang. synthetic feasibility, SF), a dostępnością syntetyczną (ang. synthetic accessibility, SA) [150]. Pierwsze z określeń używane jest w celu weryfikacji praktycznych możliwości zsyntezowania związku chemicznego, natomiast drugie definiuje stopień trudności przeprowadzenia syntezy. Innymi słowy wykonalność syntezy oznacza, iż jej przeprowadzenie jest praktycznie możliwe w określonych warunkach reakcji, natomiast dostępność syntetyczna jest metryką szacującą stopień skomplikowania syntezy.

Istnieją dwie możliwe strategie badawcze oszacowania wskaźników SF oraz SA: ekspertyza wykonana przez doświadczonych chemików syntetyków [150–154] lub symulacje *in silico* w układach modelowych. Przy czym w przypadku dużych zbiorów związków jedyną możliwością oszacowania wskaźników jest zastosowanie metod obliczeniowych. W tym celu zaproponowano wiele różnych algorytmów, które oparto na szacowaniu stopnia złożoności struktury chemicznej oraz badaniu elementów podstruktury [151], bądź bazujące na ścieżkach retrosyntetycznych rozłączeń. Przykładem tego typu rozwiązań są programy SPROUT/CAESA [155] czy też WODCA [156]. Intuicyjne szacowanie dostępności syntetycznej wykonane przez chemików również wymaga zastosowania statystycznych obliczeń w celu wyeliminowania zróżnicowania ocen wynikającego z indywidualnych doświadczeń [151].

4.2.5.1 DOSTĘPNOŚĆ SYNTETYCZNA AZANAFTALENÓW

W ramach badań własnych przeprowadzono analizę syntetycznych trendów podprzestrzeni chemicznej związków azanaftalenowych. Analiza danych masowych przechowywanych w chemicznych repozytoriach pozwala na ujawnienie relacji oraz

nowych, nieznanym wcześniej faktów. Informacja pozyskiwana tą metodą jest wolna od subiektywnych ocen i stanowi odzwierciedlenie obecnego stanu nauki w zakresie badanego zagadnienia. Niezależność od indywidualnych ocen jest szczególnie istotna w przypadku szacowania wskaźników dostępności syntetycznej.

Zdefiniowano dwa parametry: parametr SA (Równanie 4.4) oraz P (Równanie 4.5). Pierwszy z nich wylicza stosunek liczby reakcji chemicznych, w których rozważany motyw strukturalny występuje jako produkt reakcji (p) do całkowitej liczebności badanej podprzestrzeni chemicznej (s). Natomiast drugi z parametrów odnosi liczbę reakcji chemicznych (przeliczanych jako rekordy, a nie typy reakcji), w których rozważany motyw strukturalny występuje jako produkt (p) do liczby reakcji chemicznych, w których to samo ugrupowanie występuje jako substrat (r). W odniesieniu do koncepcji ewolucji chemicznej sieci wzajemnych połączeń, którą omówiono w części literaturowej pracy w Rozdziale 3.4.2.1, zdefiniowane parametry SA oraz P pozwalają na oszacowanie „popularności” wybranej grupy związków w chemii organicznej [23,55]. Wyliczane wartości parametrów SA oraz P są odzwierciedleniem preferencji cząsteczek chemicznych do tworzenia nowych połączeń.

$$\text{param_SA} = \frac{p}{s} \quad \text{Równanie 4.4}$$

$$\text{param_P} = \frac{p}{r} \quad \text{Równanie 4.5}$$

* p – liczba opisanych reakcji chemicznych, w których rozważany związek występuje jako produkt; s – liczba struktur chemicznych badanej podprzestrzeni chemicznej oraz r – liczba opisanych reakcji chemicznych, w których rozważany związek występuje jako substrat.

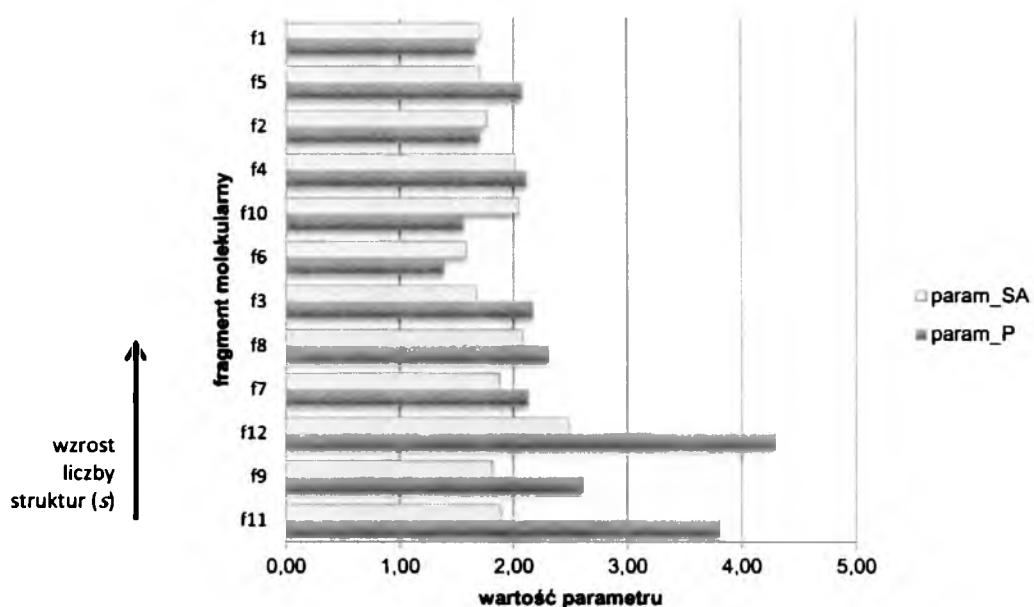
W oparciu o zdefiniowane parametry przeprowadzono analizę syntetycznych trendów w obrębie dwunastu zbiorów związków organicznych zawierających fragment mono- lub diazanaftalenowy jako podstrukturę. Przeprowadzono eksplorację bazy Beilstein, wyszukując związki zawierające jako podstrukturę fragment molekularny o topologii określonej indeksem kolejno od **f1** do **f12** (motywy strukturalne przedstawia Rysunek 4.1 w Rozdziale 4.2.1). Następnie w obrębie każdego z dwunastu zestawów cząsteczek obliczono parametr SA (Równanie 4.4) oraz parametr P (Równanie 4.5).

Tabela 4.4 podaje zestawienie wyników uzyskanych metodą eksploracji bazy Beilstein, a Rysunek 4.14 przedstawia zestawienie wyliczonych wartości wskaźników dostępności syntetycznej dla związków azanaftalenowych.

Tabela 4.4 Wskaźniki dostępności syntetycznej związków azanaftalenowych.

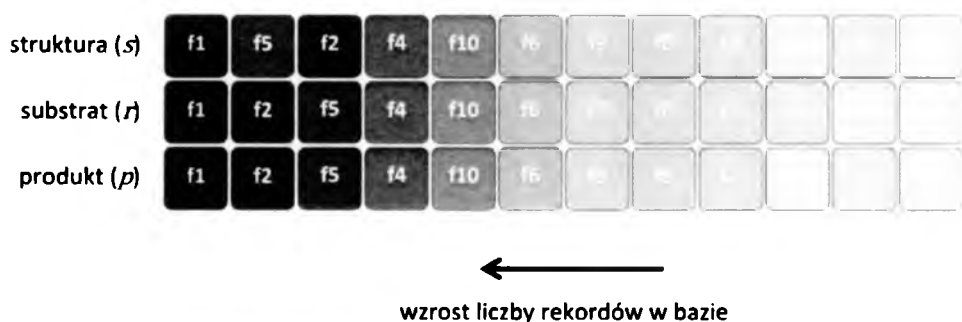
Fragment molekularny	Liczba rekordów w bazie ^a			param_SA ^b	param_P ^c
	struktura (<i>s</i>)	substrat (<i>r</i>)	produkt (<i>p</i>)		
f1	185470	190242	316593	1,71	1,66
f2	32137	33174	56637	1,76	1,71
f3	3909	3017	6551	1,68	2,17
f4	18331	17481	36905	2,01	2,11
f5	32399	26650	55269	1,71	2,07
f6	4393	5026	6961	1,58	1,38
f7	1176	1035	2204	1,87	2,13
f8	1640	1479	3413	2,08	2,31
f9	695	483	1260	1,81	2,61
f10	4694	6167	9596	2,04	1,56
f11	198	98	374	1,89	3,82
f12	827	479	2057	2,49	4,29
				Statystyka	
				$\bar{x}$	1,89 2,32
				σ	0,23 0,85
				<i>Mdn</i>	1,84 2,12
^a Dane uzyskano metodą eksploracji bazy Beilstein stosując zapytanie o podstrukturę. Dane na dzień 4 kwietnia 2010.					
^b Parametr SA zdefiniowano poprzez Równanie 4.4.					
^c Parametr P zdefiniowano poprzez Równanie 4.5.					

Wartości parametrów SA oraz P wyliczone dla poszczególnych fragmentów molekularnych f1 - f12 wykazują znaczne zróżnicowanie (Tabela 4.4 oraz Rysunek 4.14). Najwyższą wartość omawianych parametrów posiada fragment oznaczony indeksem f12. Fragment chinolinowy f1 – najbardziej rozpowszechniony pośród badanej puli związków – posiada wartość parametru SA, która plasuje go na dziesiątym miejscu w hierarchii dwunastu badanych zbiorów pod względem wartości omawianego parametru.



Rysunek 4.14 Zestawienie wartości wskaźników dostępności syntetycznej związków azanaftalenowych. Parametr SA oraz P zdefiniowano odpowiednio poprzez Równanie 4.4 oraz Równanie 4.5. Dane uzyskano metodą eksploracji bazy Beilstein stosując zapytanie o podstrukturę. Dane na dzień 4 kwietnia 2010.

Rysunek 4.14 ujawnił wpływ liczebności zbioru s , posiadającego jeden z fragmentów molekularnych **f1** - **f12** jako element podstruktury, na wartości parametrów SA oraz P. Im wyższe wartości s tym niższe wartości przyjmują parametry SA i P, co oznacza, że dla rzadkich fragmentów **f11** i **f12** obserwujemy najwyższe wartości SA oraz P. Intuicyjnie nie wydaje się, by parametry SA i P opisywać mogły w prosty sposób dostępność syntetyczną badanych zbiorów związków. Być może lepszymi miarami dostępności są bezwzględne wartości s , r lub p . Porównanie danych zebranych w tabeli powyżej (Tabela 4.4) wskazuje, że im większa liczebność zbioru s , a więc większa liczba znanych i opisanych związków chemicznych, tym też większa liczba opisanych z ich udziałem reakcji. Fakt ten jest zrozumiały, w świetle pracy chemika, którego plan realizacji syntetycznych ścieżek jest uzależniony od praktycznych możliwości pracy laboratoryjnej z wybraną grupą związków. Prawidłowości te schematycznie przedstawia Rysunek 4.15, na którym uszeregowano zbiory s , r oraz p dla poszczególnych dwunastu fragmentów azanaftalenowych uwzględniając ich liczebność. Uporządkowanie zbiorów s , r oraz p w postaci trzech zestawionych równolegle rankingów, pozwala w obrazowy sposób przedstawić opisywane zależności.



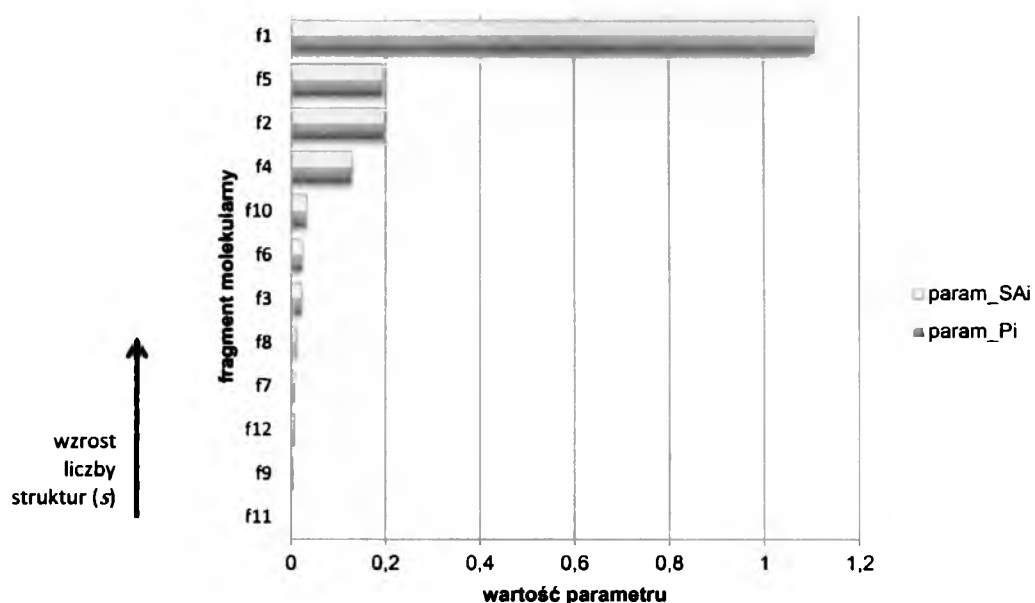
Rysunek 4.15 Ranking uwzględniający liczebność zbiorów *s*, *r* oraz *p* dla poszczególnych fragmentów azanaftaleneowych.

Równie dobrymi miarami dostępności – jak bezwzględne wartości *s*, *r* lub *p* – mogą też stanowić tożsame z nimi wartości ułamków. Rysunek 4.16 przedstawia zestawienie wyliczonych wartości wskaźników dostępności syntetycznej dla związków azanaftalenowych wyliczonych zgodnie z poniższymi równaniami.

$$\text{param_SA}_i = p_i \left(\sum_{i=1}^{12} s_i \right)^{-1} \quad \text{Równanie 4.6}$$

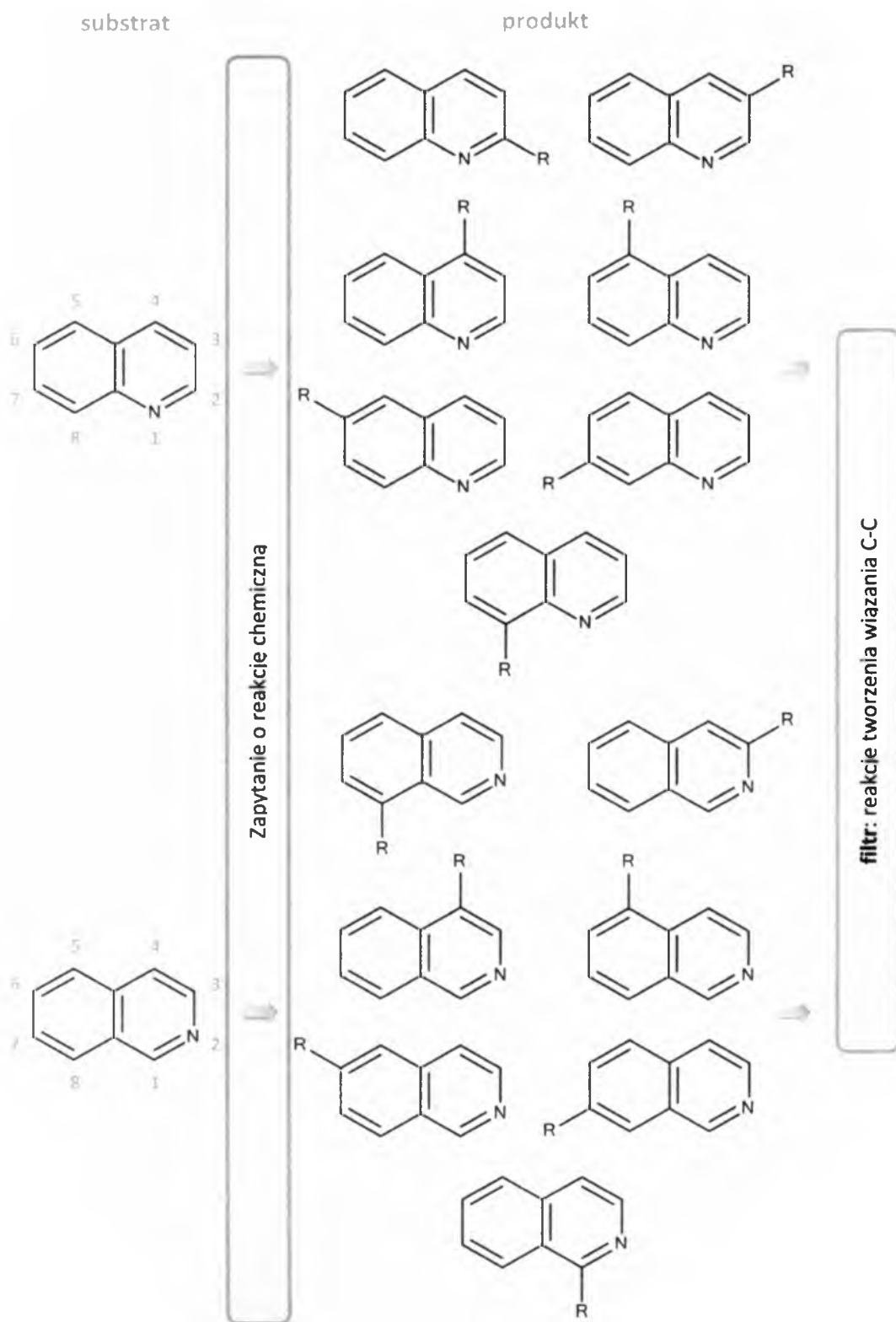
$$\text{param_P}_i = p_i \left(\sum_{i=1}^{12} r_i \right)^{-1} \quad \text{Równanie 4.7}$$

**p* – liczba opisanych reakcji chemicznych, w których rozważany związek występuje jako produkt; *s* – liczba struktur chemicznych badanej podprzestrzeni chemicznej oraz *r* – liczba opisanych reakcji chemicznych, w których rozważany związek występuje jako substrat.



Rysunek 4.16 Zestawienie wartości wskaźników dostępności syntetycznej związków azanaphtalenowych. Parametr SA_i oraz P_i zdefiniowano odpowiednio poprzez Równanie 4.6 oraz Równanie 4.7. Dane uzyskano metodą eksploracji bazy Beilstein stosując zapytanie o podstrukturę. Dane na dzień 4 kwietnia 2010.

Przeprowadzono analizę preferencyjności architektury podstawień trzonu molekularnego **f1** oraz **f2**. Wyniki uzyskano metodą eksploracji bazy Reaxys, formułując zapytanie o reakcję podstawienia pierścienia monoazanaphtalenowego **f1** i **f2** w jednej, wskazanej pozycji. Rysunek 4.17 przedstawia konstrukcję zapytania strukturalnego do bazy. Składnię zapytania zdefiniowano w taki sposób, że dozwolony jest tylko jeden, wskazany punkt podstawienia – bez możliwości podstawień przy pozostałych atomach pierścienia azanaphtalenowego. Przeszukiwanie bazy Reaxys wykonano wielokrotnie, każdorazowo badając tylko jedno z wszystkich możliwych miejsce podstawienia chinoliny i izochinoliny. W kolejnym kroku, pulę znalezionych reakcji filtrowano w celu wyodrębnienia reakcji tworzenia wiązania C-C.

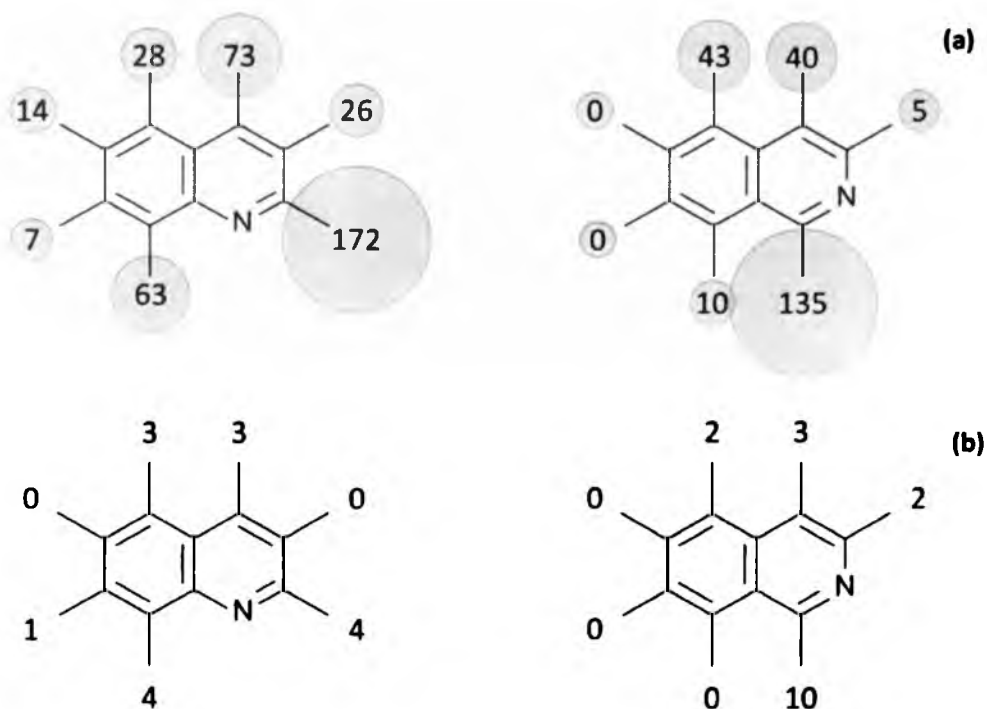


Rysunek 4.17 Konstrukcja zapytania strukturalnego do bazy. R oznacza dowolny podstawnik. Zapytanie strukturalne o reakcję chemiczną stosowano w trybie dopasowania dokładnego. Wyniki nie obejmują reakcji, w których udział biorą związki chemiczne, zawierające wskazane motywy strukturalne jako podstruktury.

Analiza uzyskanych wyników dostarcza ciekawych obserwacji na temat preferencji tworzenia dowolnego wiązania oraz wiązania C-C w obrębie dwóch badanych motywów strukturalnych. Tabela 4.5 podaje zestawienie wyników uzyskanych metodą eksploracji bazy Reaxys. W zależności od zdefiniowania miejsca podstawienia fragmentu molekularnego **f1** i **f2**, liczba znalezionych w bazie reakcji znacznie się różni, wskazując kierunkowe preferencje tworzenia podstawień dla motywu strukturalnego chinoliny i izochinoliny, co przedstawia Rysunek 4.18. Preferencje tworzenia wiązania C-C są silniej widoczne w przypadku fragmentu molekularnego **f2** niż dla motywu strukturalnego **f1** – co wskazują wyliczone wartości odchylenia standardowego (Tabela 4.5). Najwięcej reakcji podstawień fragmentu chinolinowego **f1** oraz izochinolinowego **f2**, znaleziono w przypadku podstawienia w pozycji 2 (dla chinoliny) oraz w pozycji 1 dla izochinoliny. W odniesieniu do koncepcji ewolucji chemicznej sieci wzajemnych połączeń, którą omówiono w części literaturowej rozprawy w Rozdziale 3.4.2.1, można zasugerować, iż związki posiadające tę architekturę mogą stanowić preferowane węzły (molekuły) chemicznej sieci połączeń, gdyż preferencyjność tworzenia nowych połączeń jest dla nich znacznie wyższa od pozostałych badanych miejsc podstawień fragmentu monoazanaftalenowego.

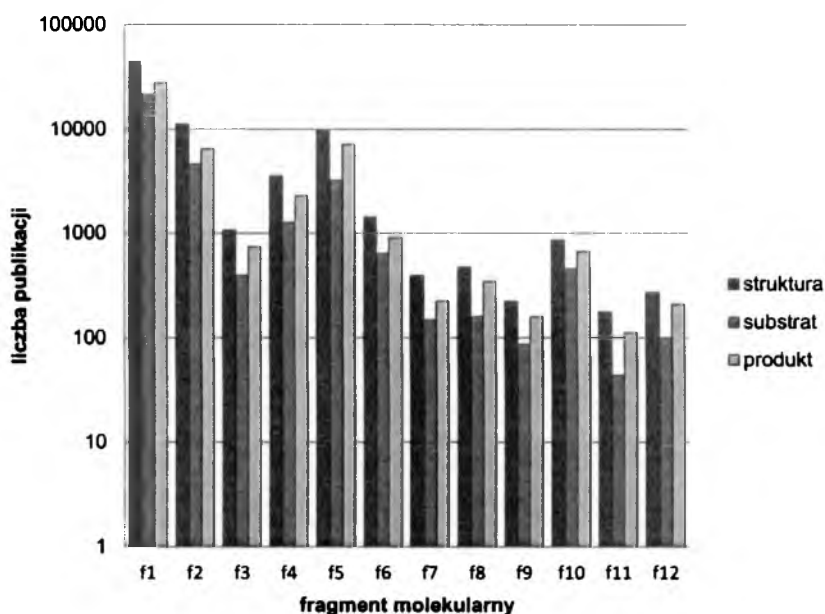
Tabela 4.5 Uprzywilejowana architektura związków monoazanaftalenowych.

Indeks atomu pierścienia ^a	Liczba reakcji ^b			
	dowolne wiązanie	wiązanie C-C	dowolne wiązanie	wiązanie C-C
	chinolina (f1)		izochinolina (f2)	
1	-	-	1	135
2	172	4	-	-
3	26	0	5	2
4	73	3	40	3
5	28	3	43	2
6	14	0	0	0
7	7	1	0	0
8	63	4	10	0
Statystyka				
$\bar{x}$	54,71	2,14	33,29	2,43
σ	52,96	1,64	44,79	3,29
Suma reakcji	383	15	233	17
^a Sposób indeksowania atomów przedstawia Rysunek 4.17.				
^b Dane na dzień 30 października 2012.				



Rysunek 4.18 Kierunkowe preferencje tworzenia dowolnego wiązania (a) oraz wiązania C-C (b) dla motywu strukturalnego chinoliny i izochinoliny. Dane uzyskano metodą eksploracji bazy Reaxys. Dane na dzień 30 października 2012.

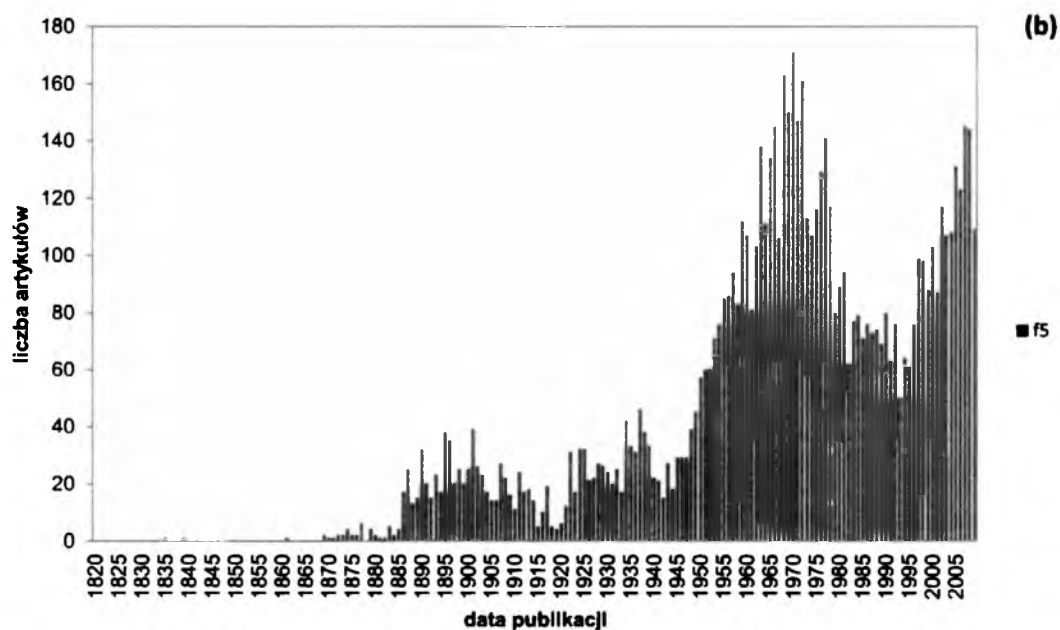
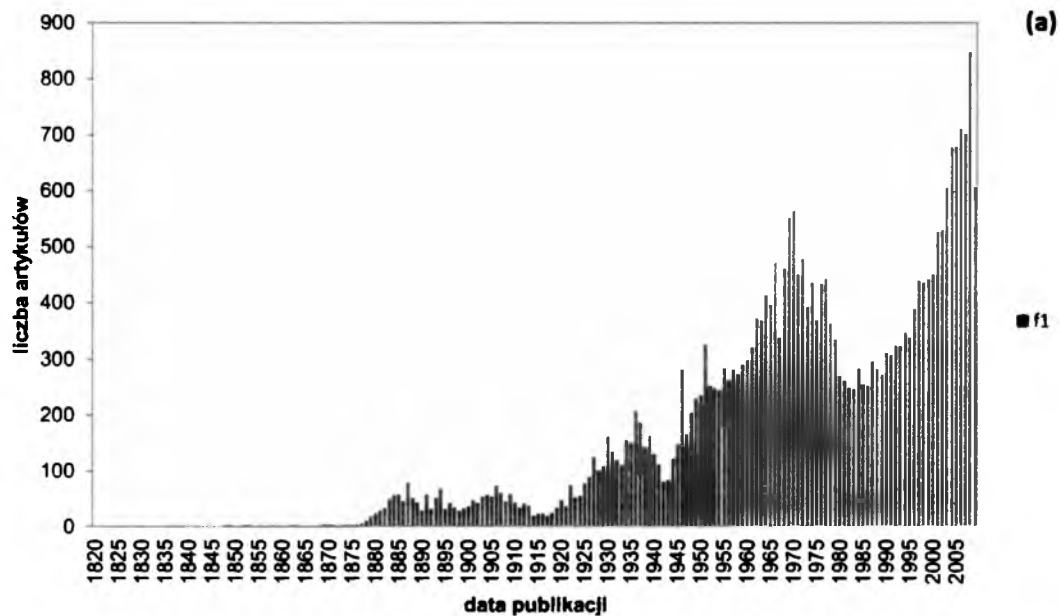
Odkryte kierunkowe preferencje tworzenia wiązania C-C, niewątpliwie związane z dostępnością syntetyczną wybranych struktur, mogą być również wynikiem szczególnego zainteresowania wybranymi grupami związków wśród chemików (Rysunek 4.19).

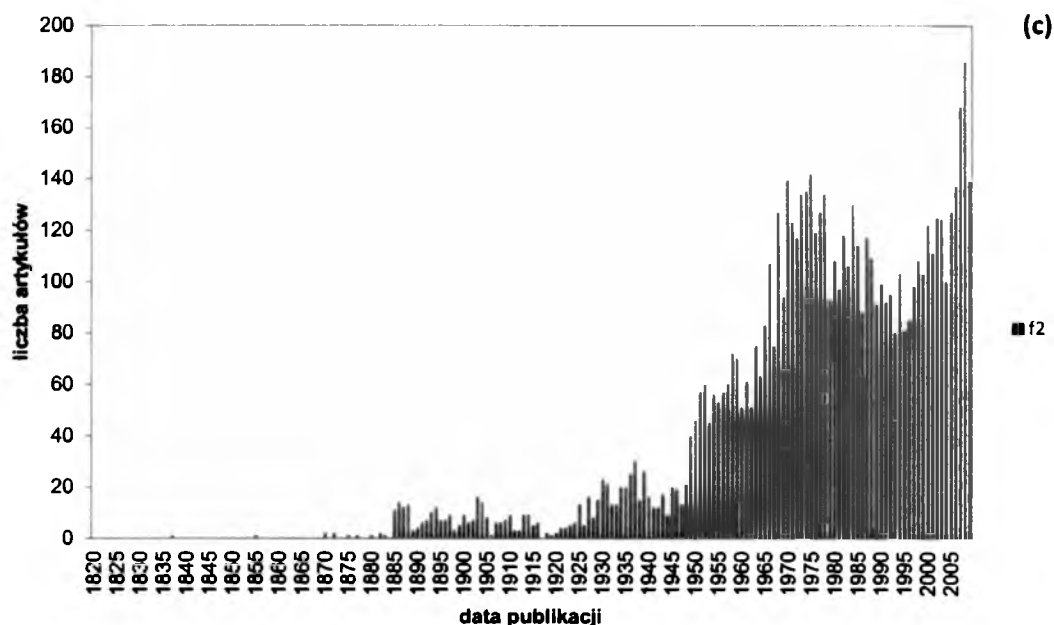


Rysunek 4.19 Liczba opublikowanych artykułów traktujących o związkach azanaftalenowych z uwzględnieniem podziału na dwanaście zbiorów struktur chemicznych posiadających jeden z fragmentów molekularnych f1 - f12 jako element podstruktury. Diagram sporządzono na podstawie danych uzyskanych metodą eksploracji bazy Beilstein. Dane na dzień 14 maja 2010.

4.2.5.2 ODKRYWANIE WIEDZY UKRYTEJ W CHEMICZNYCH BAZACH DANYCH

Analiza baz danych pod względem liczby rekordów opisujących reakcje syntezy mono- i diazanaftalenów prowadzi do dalszych ciekawych obserwacji. Eksploracja danych bibliograficznych zgromadzonych w bazie Beilstein ujawniła następujące po sobie okresy wzmożonego i zmniejszonego zainteresowania syntetycznego wśród każdej z dwunastu badanych populacji struktur chemicznych zawierających odpowiednio fragment molekularny o topologii określonej numerem porządkowym od **f1** do **f12**. Zjawisko to widoczne jest w postaci swoistego falującego profilu charakteryzującego częstość publikacji poświęconych związkom azanaftalenowym na przestrzeni ostatnich osiemdziesięciu dziewięciu lat. Rysunek 4.20 przedstawia odkryty efekt „falowania” na przykładzie trzech zbiorów związków zawierających odpowiednio fragment **f1**, **f5** oraz **f2**. Spośród przebadanych dwunastu grup, powyższe trzy charakteryzowała najwyższa liczba rekordów.





Rysunek 4.20 Tendencje i trendy syntezy organicznej na przestrzeni ostatnich osiemdziesięciu dziewięciu lat (1820-2009). Fluktuacje w liczbie pozycji literaturowych widoczne na diagramach w postaci swoistego falującego profilu przedstawiono dla trzech wybranych populacji struktur chemicznych zawierających odpowiednio fragment molekularny o topologii określonej indeksem porządkowym f1 (a), f5 (b) oraz f2 (c). Dane uzyskano metodą eksploracji bazy Beilstein. Odnośniki literaturowe referują do opisanych reakcji syntez związków zawierających jako podstrukturę rozważane fragmenty molekularne. Dane na dzień 14 maja 2010.

Liczba publikacji poświęconych pochodnym chinoliny osiągnęła wartość maksymalną w roku 2008 (Rysunek 4.20a). Pozostaje kwestią do rozstrzygnięcia które czynniki i w jakim stopniu wpływają na okresowy wzrost zainteresowań wybraną grupą związków. Jedną z przyczyn może być znany z chemii medycznej wpływ strategii „me-too”. Pojawiająca się praca rodzi zainteresowanie, które po jakimś czasie osiąga maksimum, by z kolei w etapie „znużenia” maleć do wartości minimum lokalnego. Analizy tego typu niewątpliwie nie tylko pozwalają na odkrywanie wiedzy ukrytej w informacji gromadzonej i przechowywanej w bazach danych, ale ujawniają kierunki, w których ewoluuje chemia organiczna.

4.3 FRAGMENTACJA DEKREMENTACYJNA LEKÓW AKTYWNYCH WZGLĘDEM ENZYMU INTEGRAZY WIRUSA HIV

Dekrementacja (łac. decrementum – ubywanie, ubytek) z definicji oznacza instrukcję zmniejszająca o jeden wartość zmiennej podanej jako argument i jest działaniem odwrotnym do inkrementacji.

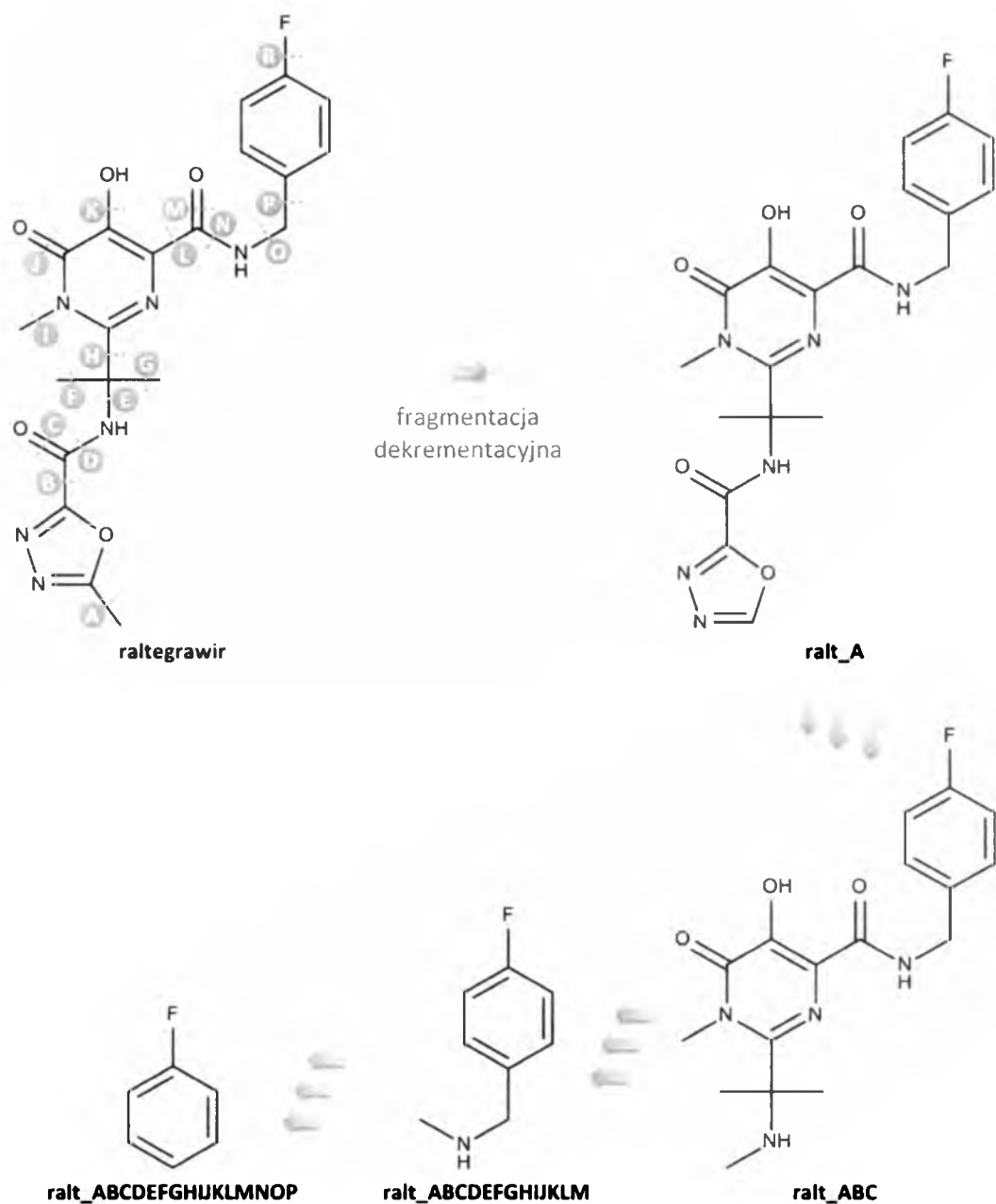
4.3.1 ALGORYTM PROCEDURY FRAGMENTACJI DEKREMENTACYJNEJ

Fragmentacja dekrementacyjna cząsteczek leków stosowanych w terapii przeciw HIV, polegała na wydzielaniu fragmentów molekularnych poprzez stopniowe zmniejszanie wyjściowej struktury chemicznej o jeden atom. Jedynie układy cykliczne odłączane były w całości, bez rozrywania tworzących je wiązań. Po każdym odcięciu jednostki strukturalnej, przeprowadzano przeszukiwanie bazy ChemDB HIV, Opportunistic Infection and Tuberculosis Therapeutics Database (ChemDB HIV/OI/TB) w oparciu o podstrukturę zminimalizowanego fragmentu molekularnego. Opis repozytorium sporządzono w Rozdziale 3.4.1.3. Stosując odpowiednią składnię zapytania strukturalnego, połączonego z filtrami zapytań tekstowych, badano statystykę rozpowszechnienia danego fragmentu wśród zdeponowanych w bazie związków o potencjalnym zastosowaniu w terapii HIV/AIDS.

W ten sposób wykonano wiele podziałów dekrementacyjnych struktury raltegrawiru i elwitegrawiru, uzyskując swoiste „ścieżki rozłączeń” w zależności od kierunku przeprowadzania instrukcji redukującej. Wyboru leków dokonano w oparciu o mechanizm stymulacji farmakologicznej, wybierając środki terapeutyczne należące do klasy inhibitorów IN. Produkty lecznicze zarejestrowane w terapii HIV/AIDS opisano w Rozdziale 3.7.1.

Algorytm tworzenia ścieżek rozłączeń w protokole fragmentacji dekrementacyjnej oraz sposób indeksowania wyodrębnianych fragmentów molekularnych przedstawia Rysunek 4.21. Składnię zapytań kierowanych do bazy oraz dane liczbowe uzyskane dla poszczególnych ścieżek fragmentacji dekrementacyjnej zebrano w trzech tabelach

części eksperymentalnej niniejszej rozprawy (Tabela 6.5, Tabela 6.6 oraz Tabela 6.7 w Rozdziale 6.2). Poniżej przedstawiono omówienie uzyskanych wyników.

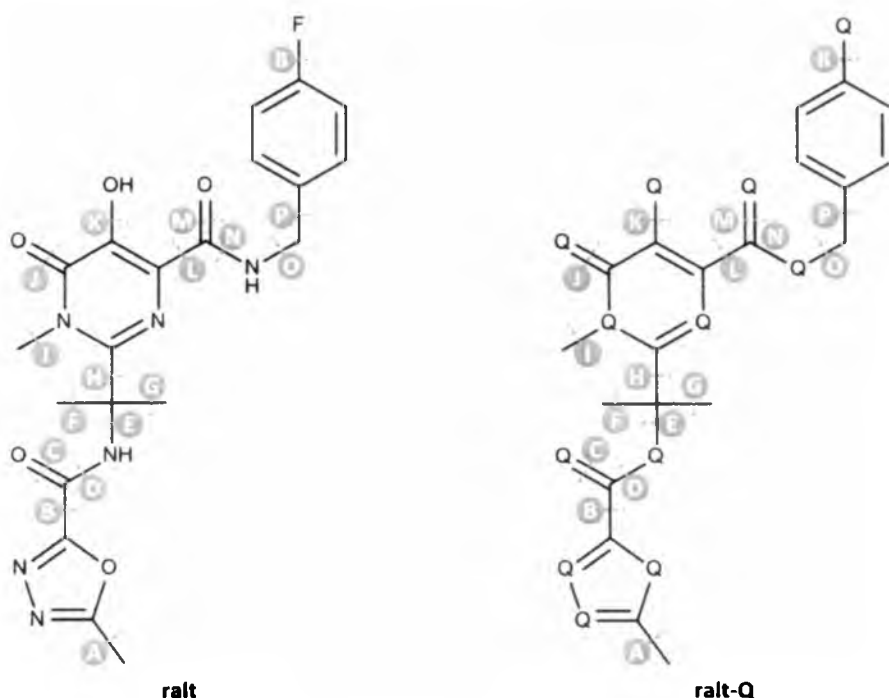


Rysunek 4.21 Algorytm tworzenia ścieżek rozłączeń w procedurze fragmentacji dekrementacyjnej oraz sposób indeksowania fragmentów molekularnych na przykładzie cząsteczki raltegrawiru.

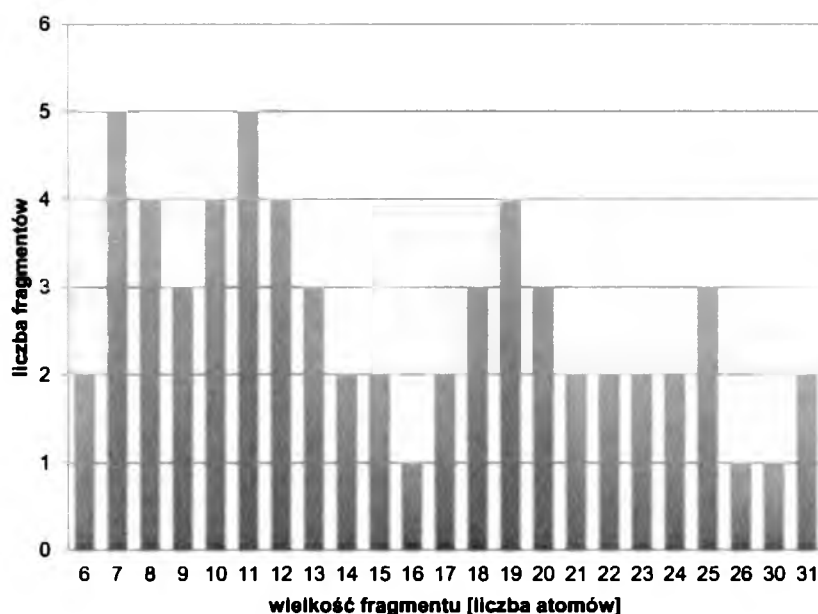
4.3.2 TOPOLOGIA LEKOTYPII FRAGMENTÓW MOLEKULARNYCH WYDZIELONYCH W STRUKTURZE CZĄSTECZKI RALTEGRAWIRU

Przeprowadzono podziały dekrementacyjne cząsteczki raltegrawiru (**ralt**) oraz uogólnionej reprezentacji jej struktury, w której w miejscu atomu azotu, tlenu i fluoru dopuszczono wystąpienie dowolnego heteroatomu. Uogólnioną reprezentację struktury cząsteczki raltegrawiru oznaczono jako **ralt-Q**. Rysunek 4.22 przedstawia sposób kodowania wiązań, umożliwiający sekwencyjny zapis kierunku przeprowadzanej ścieżki rozłączeń.

W wyniku przeprowadzonych operacji opisano pięć ścieżek fragmentacji dekrementacyjnej cząsteczki raltegrawiru oraz dwie ścieżki rozłączeń struktury **ralt-Q**. W ten sposób wydzielono 70 fragmentów molekularnych dla struktury **ralt** oraz 32 fragmenty dla reprezentacji **ralt-Q** (Tabela 6.5 oraz Tabela 6.6 w Rozdziale 6.2). Po usunięciu duplikatów, pula fragmentów uzyskanych dla struktury **ralt** została zredukowana do 62 unikatowych motywów strukturalnych. Fragmenty pogrupowano w zależności od rozmiaru. Jako kryterium wielkości fragmentu przyjęto liczbę atomów wchodzących w jego skład, przy czym nie uwzględniano atomów wodoru. Histogram przedstawiony na rysunku poniżej ilustruje liczebność fragmentów w poszczególnych grupach. Do efektywnego zarządzania informacją chemiczną w powstałej bibliotece fragmentów molekularnych wykorzystano aplikację Instant JChem grupy ChemAxon.

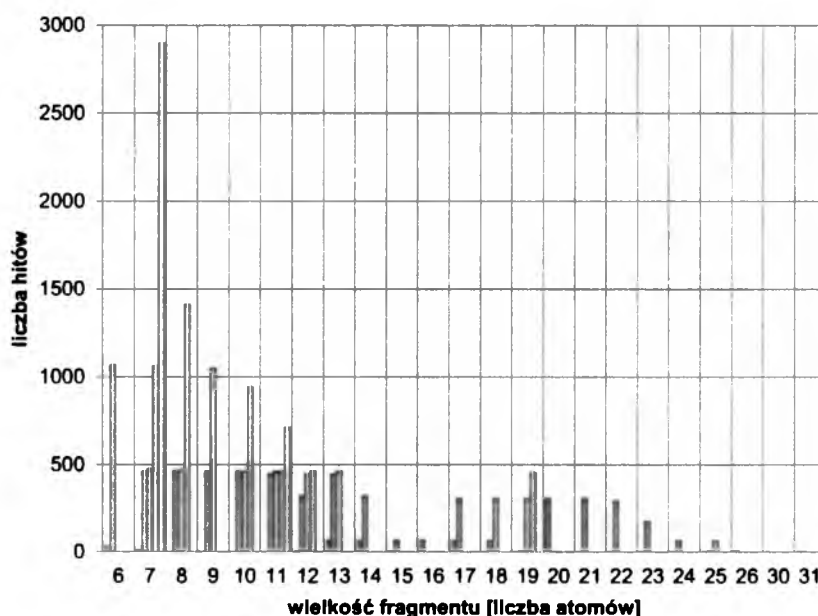


Rysunek 4.22 Sposób indeksowania wiązań w procedurze fragmentacji dekrementacyjnej dla cząsteczki raltegrawiru. Struktura ralt-Q jest uogólnioną reprezentacją struktury cząsteczki raltegrawiru (ralt), w której w miejscu atomu azotu, tlenu i fluoru dopuszczono wystąpienie dowolnego heteroatomu (symbol Q).



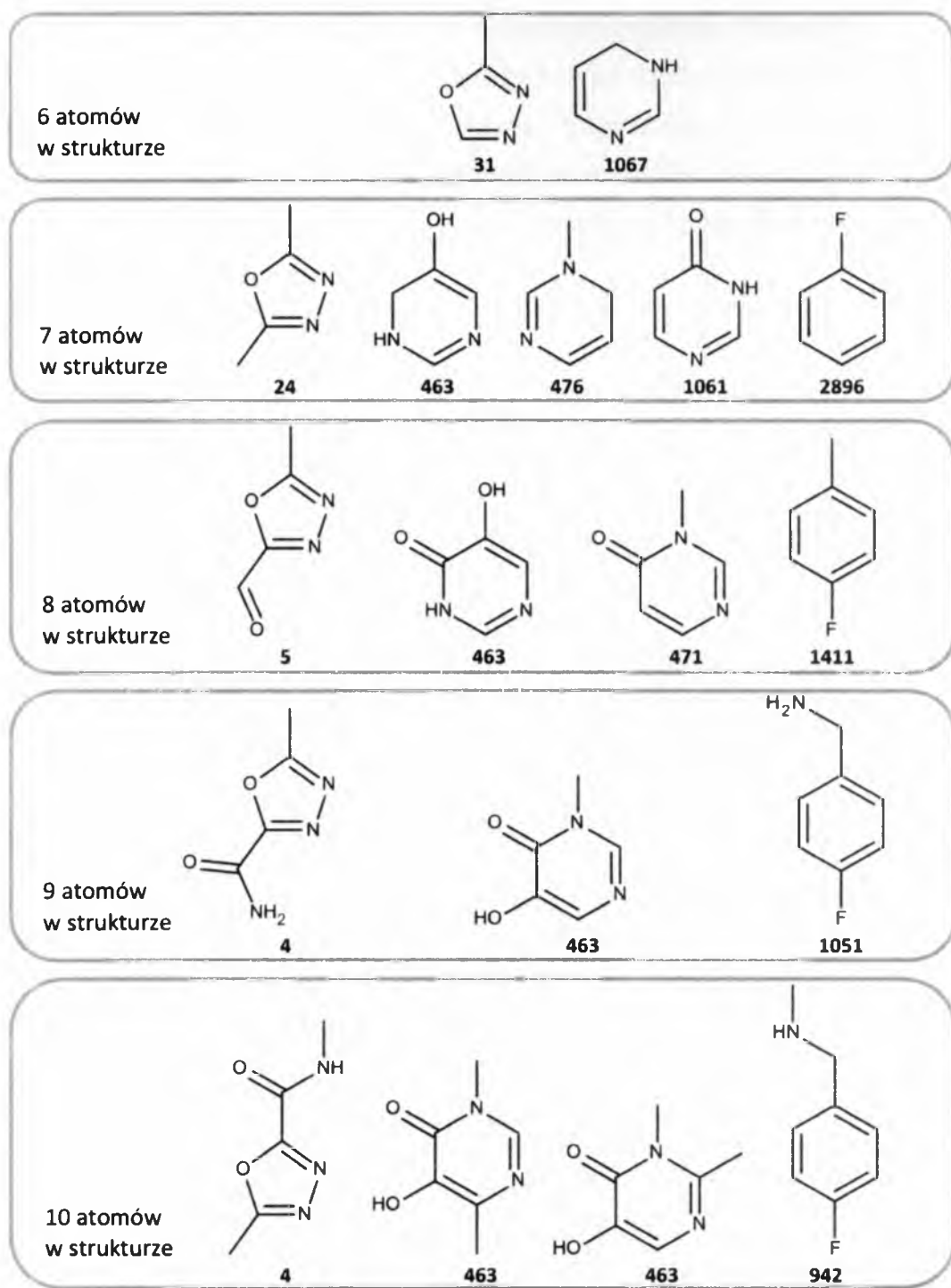
Rysunek 4.23 Liczebność fragmentów molekularnych wydzielonych na drodze fragmentacji dekrementacyjnej cząsteczki raltegrawiru (ralt) w zależności od rozmiaru wydzielonego elementu podstruktury. Wielkość fragmentu molekularnego odpowiada liczbie atomów wchodzących w jego skład (bez uwzględnienia atomów wodoru).

Analiza uzyskanych wyników dostarczyła wielu ciekawych obserwacji. Statystyka rozpowszechnienia wydzielonych fragmentów molekularnych wskazuje wyraźne zróżnicowanie w częstotliwości występowania wyodrębnionych fragmentów molekularnych w populacji związków posiadających adnotację w kierunku zastosowań w terapii przeciw HIV. Rysunek 4.24 ilustruje znalezione zależności.



Rysunek 4.24 Populacja fragmentów molekularnych wyodrębnionych na drodze fragmentacji dekrementacyjnej cząsteczki raltegraviru (ralt) w zależności od rozmiaru wydzielonego elementu podstruktury. Wielkość fragmentu molekularnego odpowiada liczbie atomów wchodzących w jego skład (bez uwzględnienia atomów wodoru). Liczba słupków dla poszczególnej kategorii związanej z wielkością fragmentu odpowiada liczbie wyodrębnionych elementów podstruktury o danej wielkości, np. wyodrębniono pięć różnych fragmentów molekularnych posiadających siedem atomów w strukturze, natomiast wyodrębniono tylko jeden fragment molekularny posiadający szesnaście atomów w strukturze. Wyniki uzyskano metodą eksploracji bazy ChemDB HIV/OI/TB.

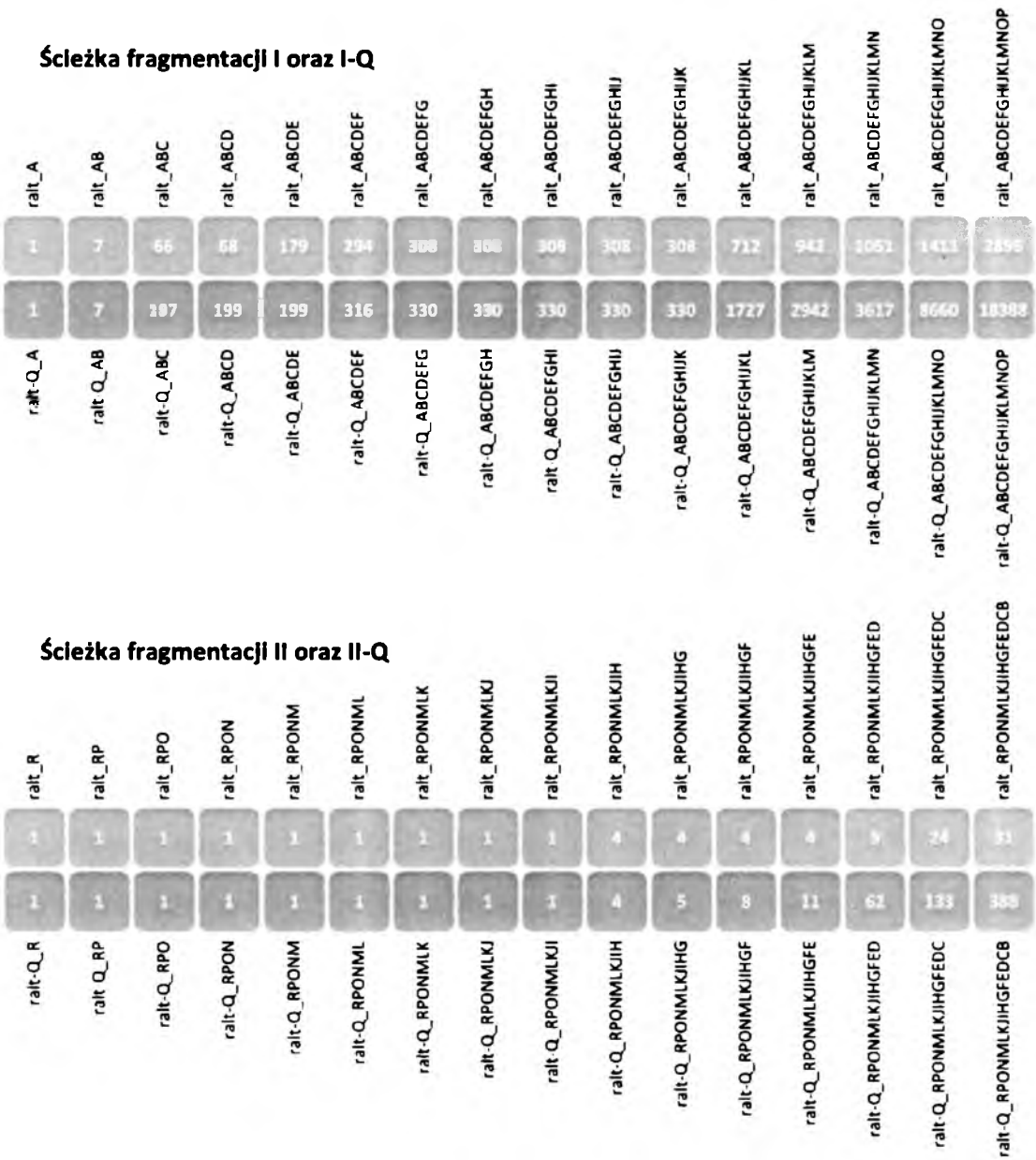
Rysunek 4.25 przedstawia najmniejsze (od sześciu do dziesięciu atomów w strukturze) wyodrębnione fragmenty wraz ze wskazaniem stopnia ich rozpowszechnienia jako elementów podstruktur związków zdeponowanych w bazie ChemDB HIV/OI/TB.



Rysunek 4.25 Struktury najmniejszych (od sześciu do dziesięciu atomów w strukturze) wyodrębnionych fragmentów molekularnych. Podano stopień rozpowszechnienia danego motywu jako elementu podstruktury związków zdeponowanych w bazie ChemDB HIV/OI/TB i posiadających adnotacje na temat zastosowania w terapii przeciw HIV.

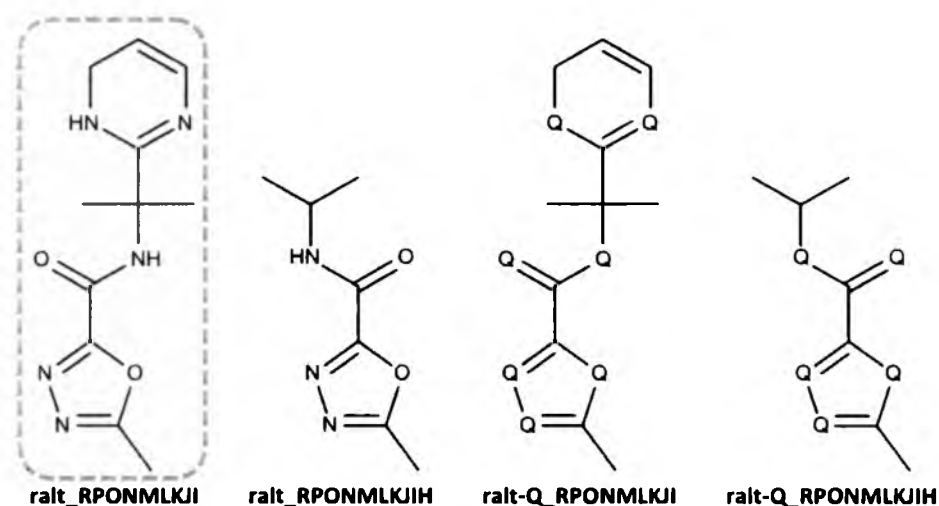
Porównanie liczby rekordów uzyskanych po każdym kroku wykonania instrukcji redukującej, dla dwóch różnych ścieżek fragmentacji dekrementacyjnej, pozwala

wskazać ugrupowania, które są swoiste dla raltegrawiru i nie występują jako podstruktury wśród związków zdeponowanych w bazie oraz takie, które z dużą częstotliwością występują w populacji związków posiadających adnotację w kierunku zastosowań w terapii przeciw HIV. Porównanie wyników uzyskanych dla dwóch różnych ścieżek fragmentacji przedstawiono na poniższym schemacie.



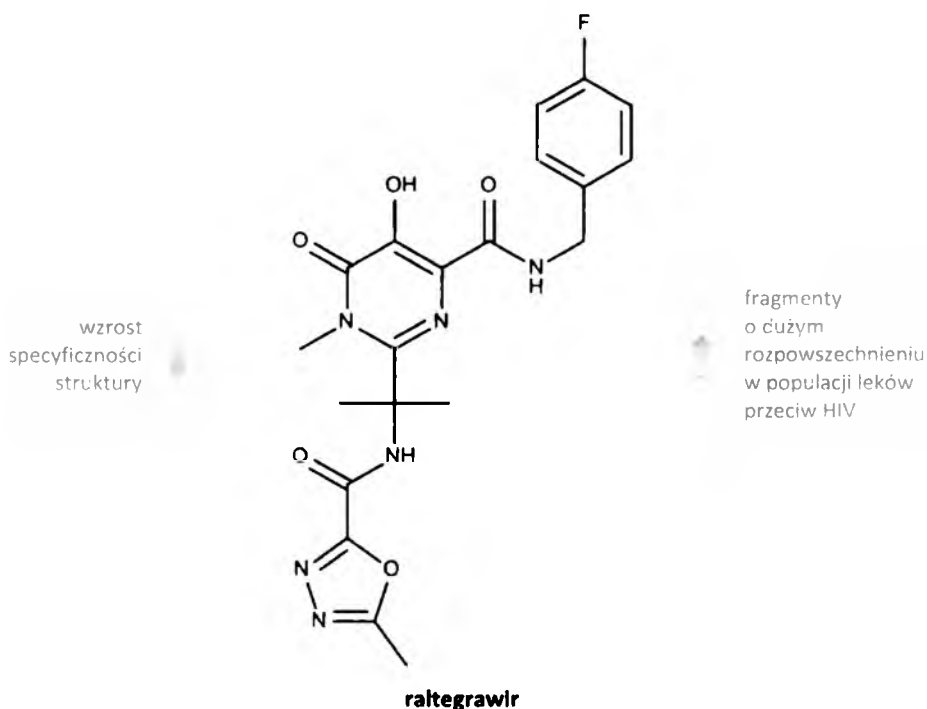
Rysunek 4.26 Porównanie dwóch ścieżek fragmentacji dekrementacyjnej cząsteczki raltegrawiru. Przedstawiono liczbę rekordów uzyskanych w bazie po każdym kroku przeprowadzenia instrukcji redukującej. Wyniki uzyskano metodą eksploracji bazy ChemDB HIV/OI/TB dla struktury ralt oraz struktury ralt-Q. Dane na dzień 19 grudnia 2010.

Fragmenty o indeksach **ralt_A**, **ralt-Q_A** oraz **ralt_R** kolejno aż do **ralt_RPONMLKJI**, a także **ralt-Q_R** kolejno aż do **ralt-Q_RPONMLKJI**) są swoiste dla raltegrawiru i nie występują jako podstruktury wśród związków zdeponowanych w bazie ChemDB HIV/OI/TB. Dwa spośród wymienionych motywów strukturalnych - fragmenty o indeksach **ralt_RPONMLKJI**, a także **ralt-Q_RPONMLKJI** – stanowią najmniejsze, swoiste elementy strukturalne cząsteczki rozważanego leku. Kolejny krok instrukcji redukującej tworzy fragmenty o indeksach **ralt_RPONMLKJIH** oraz **ralt-Q_RPONMLKJIH**, których struktury z niską częstotliwością występują już jako elementy podstruktur związków skatalogowanych w eksplorowanej bazie. Rysunek 4.27 przedstawia struktury omawianych fragmentów molekularnych.



Rysunek 4.27 Swoiste elementy strukturalne cząsteczki raltegrawiru. Linia przerywaną zaznaczono najmniejszy wyodrębniony fragment, który obecny jest tylko i wyłącznie w strukturze rozważanego leku i nie występuje jako element podstruktury innych związków zdeponowanych w bazie ChemDB HIV/OI/TB. Fragmenty o indeksach **ralt_RPONMLKJIH** oraz **ralt-Q_RPONMLKJIH** z niską częstotliwością występują jako elementy podstruktur związków skatalogowanych w bazie.

Analiza siedmiu ścieżek fragmentacji dekrementacyjnej (Tabela 6.5 oraz Tabela 6.6 w Rozdziale 6.2) pozwala na uogólnienie powyższych obserwacji i wskazanie w strukturze raltegrawiru dwóch obszarów, dla których wyodrębniane elementy podstruktury cechuje większa lub mniejsza swoistość. Zaobserwowaną zależność przedstawia Rysunek 4.28.



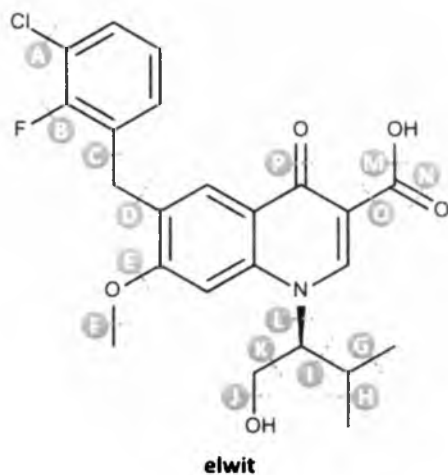
Rysunek 4.28 Obszary specyficzności strukturalnej cząsteczki raltegraviru.

4.3.3 TOPOLOGIA LEKOTYPII FRAGMENTÓW MOLEKULARNYCH WYDZIELONYCH W STRUKTURZE CZĄSTECZKI ELWITEGRAWIRU

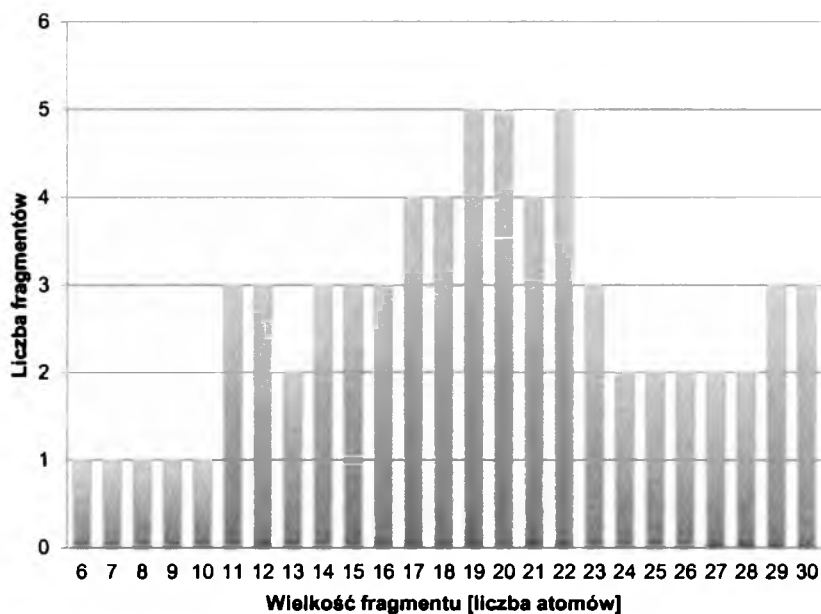
Przeprowadzono podziały dekrementacyjne cząsteczki elwitegraviru (**elwit**). Rysunek 4.22 przedstawia sposób kodowania wiązań, umożliwiający sekwencyjny zapis kierunku przeprowadzanej ścieżki rozłączeń.

W wyniku przeprowadzonych operacji opisano sześć ścieżek fragmentacji dekrementacyjnej cząsteczki elwitegraviru. W ten sposób wydzielono 88 fragmentów molekularnych dla struktury **elwit** (Tabela 6.7 w Rozdziale 6.2). Po usunięciu duplikatów, pula fragmentów została zredukowana do 68 unikatowych motywów strukturalnych. Fragmenty pogrupowano w zależności od rozmiaru. Jako kryterium wielkości fragmentu przyjęto liczbę atomów wchodzących w jego skład, przy czym nie uwzględniano atomów wodoru. Histogram przedstawiony na rysunku poniżej ilustruje liczebność fragmentów w poszczególnych grupach (Rysunek 4.30). Natomiast Rysunek 4.31 przedstawia statystykę rozpowszechnienia wydzielonych fragmentów molekularnych w populacji związków posiadających adnotację w kierunku zastosowań

w terapii przeciw HIV. Do efektywnego zarządzania informacją chemiczną w powstałej bibliotece fragmentów molekularnych wykorzystano aplikację Instant JChem grupy ChemAxon.

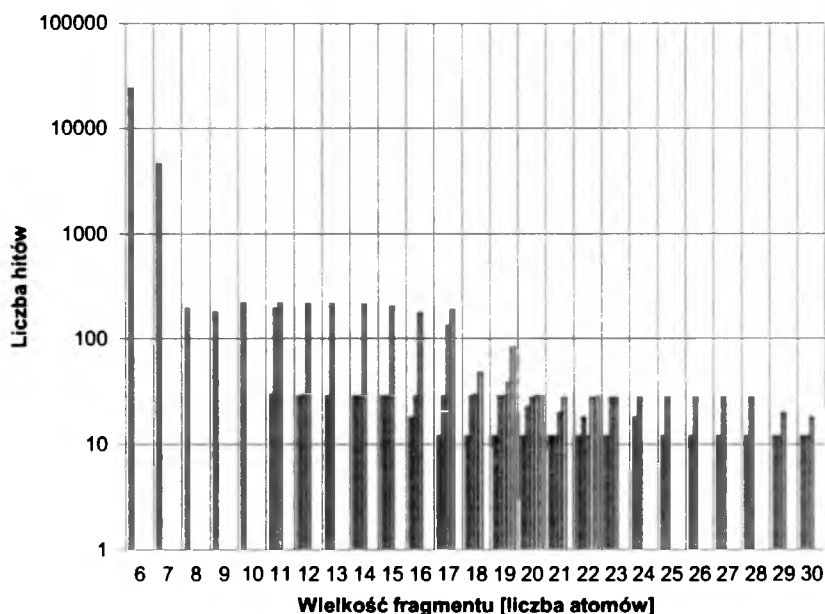


Rysunek 4.29 Sposób indeksowania wiązań w procedurze fragmentacji dekrementacyjnej dla cząsteczki elwitegrawiru.



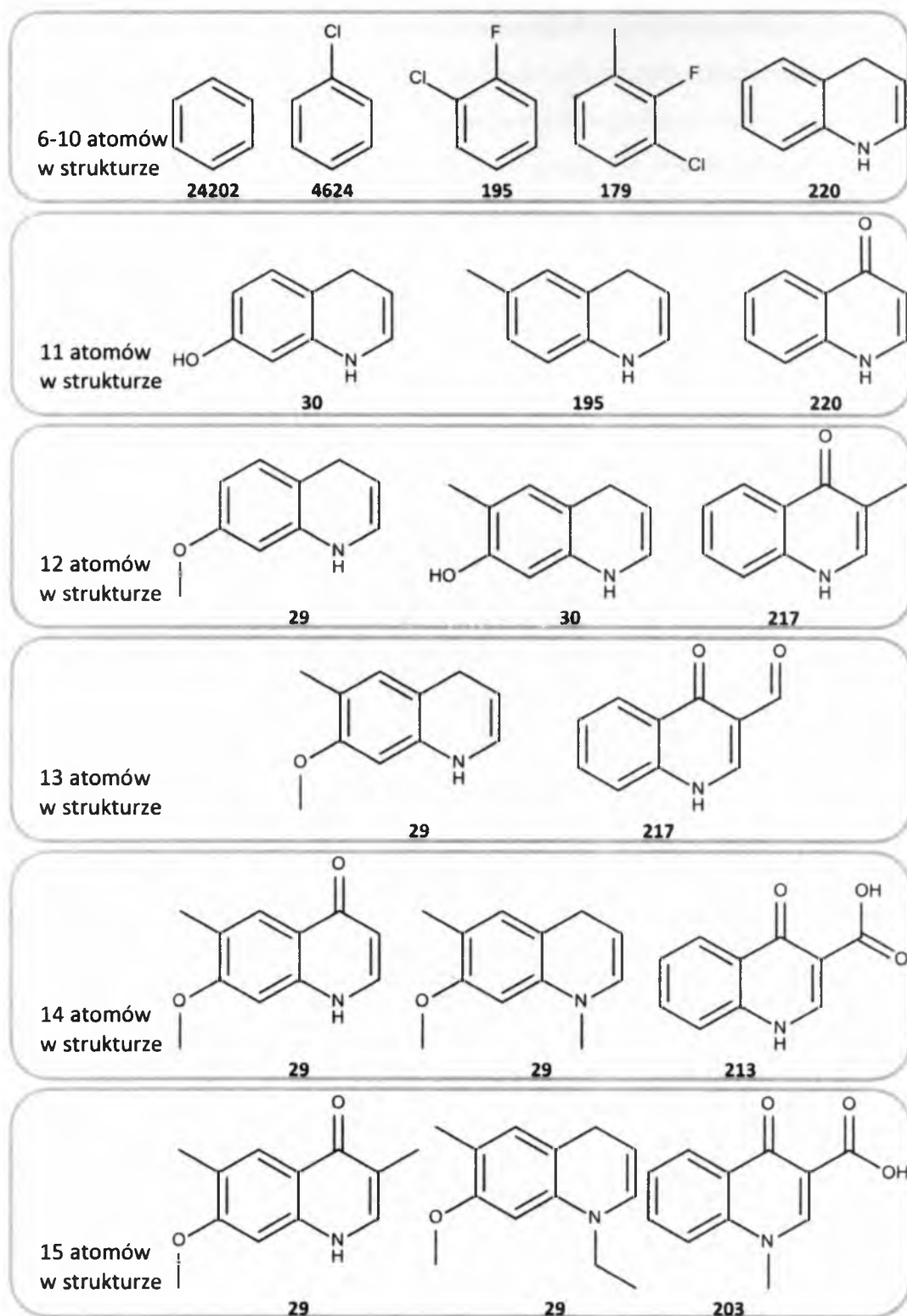
Rysunek 4.30 Liczebność fragmentów molekularnych wydzielonych na drodze fragmentacji dekrementacyjnej cząsteczki elwitegrawiru (elwit) w zależności od rozmiaru wydzielonego elementu podstruktury. Wielkość fragmentu molekularnego odpowiada liczbie atomów wchodzących w jego skład (bez uwzględnienia atomów wodoru).

Statystyka rozpowszechnienia fragmentów molekularnych, wydzielonych w strukturze cząsteczki elwitegrawiru, wskazuje wyraźne zróżnicowanie w częstotliwości występowania poszczególnych fragmentów jako elementów podstruktury w populacji związków posiadających adnotację w kierunku zastosowań w terapii przeciw HIV. Rysunek 4.31 ilustruje znalezione zależności.



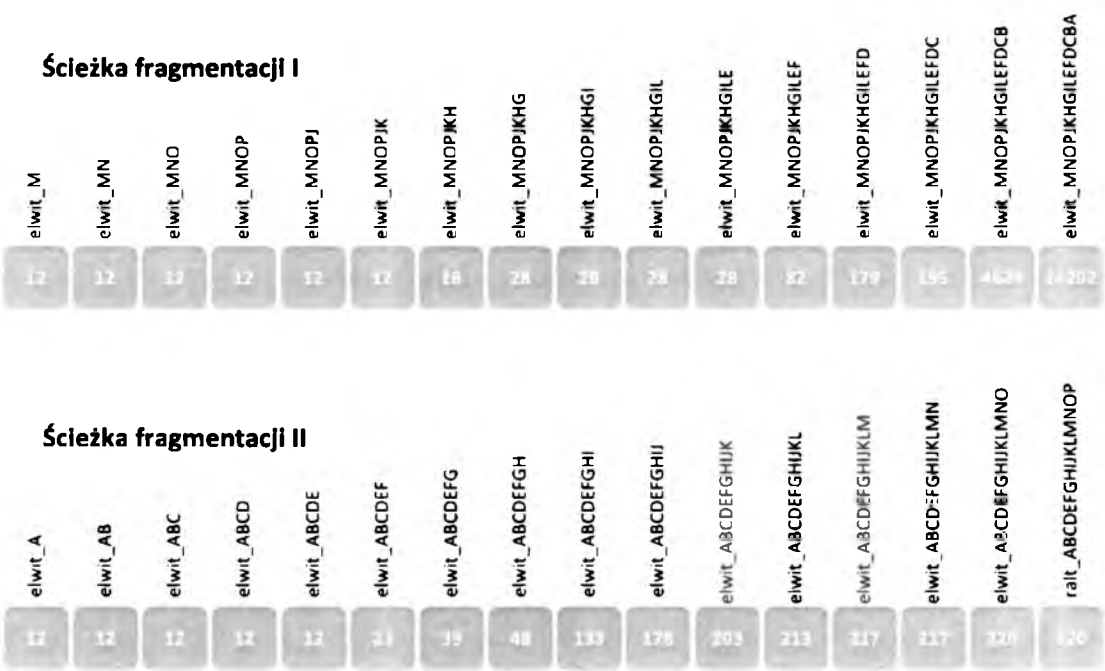
Rysunek 4.31 Populacja fragmentów molekularnych wyodrębnionych na drodze fragmentacji dekrementacyjnej cząsteczki elwitegrawiru (elwit) w zależności od rozmiaru wydzielonego elementu podstruktury. Wielkość fragmentu molekularnego odpowiada liczbie atomów wchodzących w jego skład (bez uwzględnienia atomów wodoru). Liczba słupków dla poszczególnej kategorii związanej z wielkością fragmentu odpowiada liczbie wyodrębnionych elementów podstruktury o danej wielkości, np. wyodrębniono pięć różnych fragmentów molekularnych posiadających dziewiętnaście atomów w strukturze, natomiast wyodrębniono tylko jeden fragment molekularny posiadający sześć atomów w strukturze. Wyniki uzyskano metodą eksploracji bazy ChemDB HIV/OI/TB.

Rysunek 4.32 przedstawia najmniejsze (od sześciu do piętnastu atomów w strukturze) wyodrębnione fragmenty wraz ze wskazaniem stopnia ich rozpowszechnienia jako elementów podstruktur związków zdeponowanych w bazie ChemDB HIV/OI/TB.



Rysunek 4.32 Struktury najmniejszych (od sześciu do piętnastu atomów w strukturze) wyodrębnionych fragmentów molekularnych. Podano stopień rozpowszechnienia danego motywu jako elementu podstruktury związków zdeponowanych w bazie ChemDB HIV/OI/TB i posiadających adnotacje na temat zastosowania w terapii przeciw HIV.

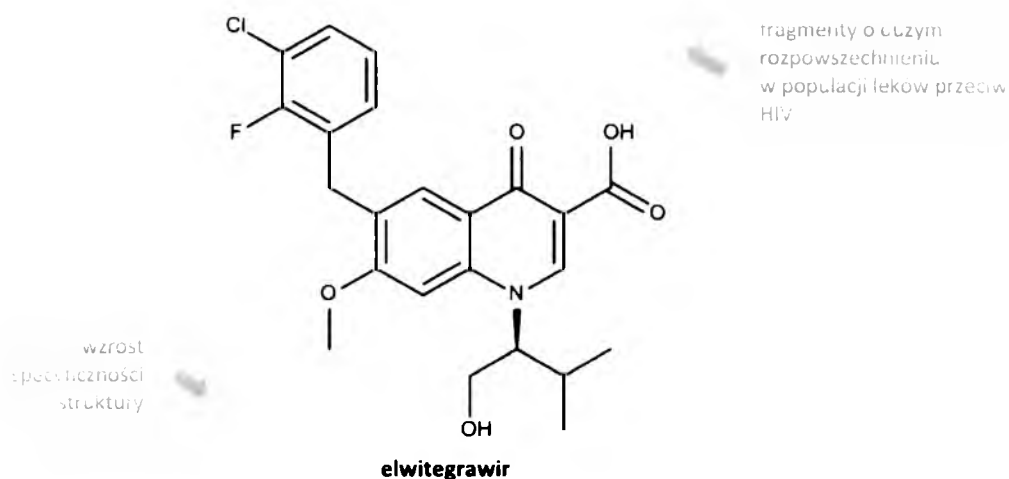
Porównanie liczby rekordów uzyskanych po każdym kroku wykonania instrukcji redukującej, dla dwóch różnych ścieżek fragmentacji dekrementacyjnej, w przypadku cząsteczki raltegrawiru, pozwoliło wskazać ugrupowania, które są swoiste dla struktury tego leku i nie występują jako podstruktury wśród związków zdeponowanych w bazie (Rysunek 4.26 i Rysunek 4.27 w Rozdziale 4.3.2). Analiza porównawcza sześciu ścieżek fragmentacji dekrementacyjnej cząsteczki elwitegrawiru wskazuje, że w strukturze tego leku nie występują ugrupowania swoiste, a każdy wyodrębniony fragment molekularny występuje z pewną częstotliwością jako element podstruktury związków zdeponowanych w bazie ChemDB HIV/OI/TB. Porównanie wyników uzyskanych dla dwóch różnych ścieżek fragmentacji dekrementacyjnej cząsteczki elwitegrawiru przedstawia Rysunek 4.33.



Rysunek 4.33 Porównanie dwóch ścieżek fragmentacji dekrementacyjnej cząsteczki elwitegrawiru. Przedstawiono liczbę rekordów uzyskanych w bazie po każdym kroku przeprowadzenia instrukcji redukującej. Wyniki uzyskano metodą eksploracji bazy ChemDB HIV/OI/TB dla struktury elwit. Dane na dzień 19 grudnia 2010.

Analiza sześciu ścieżek fragmentacji dekrementacyjnej (Tabela 6.7 w Rozdziale 6.2) pozwala na uogólnienie powyższych obserwacji i wskazanie w strukturze elwitegrawiru dwóch obszarów, dla których wyodrębniane fragmenty molekularne cechuje większy

bądź też mniejszy stopień rozpowszechnienia jako elementów podstruktury w populacji związków posiadających adnotację w kierunku zastosowań w terapii przeciw HIV. Zaobserwowaną zależność przedstawia Rysunek 4.34. Jednak w strukturze cząsteczki elwitegrawiru (przeciwnie niż w przypadku raltegrawiru) nie występują ugrupowania swoiste, a każdy wyodrębniony fragment molekularny występuje z pewną częstotliwością jako element podstruktury związków zdeponowanych w eksplorowanej bazie.



Rysunek 4.34 Obszary specyficzności strukturalnej cząsteczki elwitegrawiru.

Algorytm procedury fragmentacji dekrementacyjnej pozwala badać stopień rozpowszechnienia danego fragmentu w odpowiednio zdefiniowanej przestrzeni chemicznej. Tego typu badanie umożliwia przeprowadzenie analizy topologicznej badanej grupy związków w szerokim aspekcie problemów polifarmakologii leków. Analiza ścieżek fragmentacji dekrementacyjnej pozwala na wskazanie w strukturze badanej cząsteczki obszarów, dla których wyodrębniane elementy podstruktury cechuje większa lub mniejsza swoistość. Identyfikowane w ten sposób fragmenty mogą stanowić punkt wyjściowy w procesie projektowaniu nowych związków biologicznie aktywnych.

Procedury dekrementacyjnego podziału struktury chemicznej innego typu (ang. fragmental topology-activity landscapes, FRAGTAL) oraz ich wykorzystanie we fragonomice leków przeciw HIV opisano w załączonej publikacji:

- **Załącznik 2a**

Bak A., Magdziarz T., Kurczyk A., Serafin K., Polanski J.; Probing a Chemical Space for Fragmental Topology-Activity Landscapes (FRAGTAL): Application for Diketo Acid and Catechol HIV Integrase Inhibitor Offspring Fragments., *Comb. Chem. High Throughput Screen.*, 2013, 16(4), 274-287.

4.4 OPRACOWANIE I WERYFIKACJA PROTOKOŁU WIRTUALNEGO SKRININGU INHIBITORÓW ENZYMU INTEGRAZY WIRUSA HIV-1⁷

Zakażenie wirusem HIV stanowi główny obszar terapeutyczny przeprowadzonych badań. Badania w szczególności obejmowały inhibitory enzymu integrazy (IN) wirusa HIV-1. Wykonano różnicowanie zbiorów związków aktywnych i nieaktywnych względem wirusa HIV w kontekście uprzywilejowanych fragmentów molekularnych. W oparciu o metody bazujące na teorii informacji opracowano badawczą (obliczeniową) strategię umożliwiającą wskazanie fragmentów molekularnych, które z dużym prawdopodobieństwem mogą stanowić uprzywilejowane motywy strukturalne (PS) badanej przestrzeni farmakologicznej. Wyliczane wskaźniki PS w połączeniu z metodami uczenia maszynowego (ML) posłużyły do opracowania protokołu wirtualnego skrinigu (VS) inhibitorów IN.

Efektywność opracowanej metodyki VS zweryfikowano w układzie rzeczywistym. Nowe, potencjalne inhibitory IN wirusa HIV-1 wyłonione na drodze VS, zakupiono i skierowano do wykonania testów biologicznych *in vitro* w celu weryfikacji prognozowanej bioaktywności tych struktur.⁸ Wykonano oznaczenie cytotoksyczności i aktywności przeciw HIV preparatów w hodowlach komórek ustalonej linii CEM-T4 inokulowanej laboratoryjnym izolatem wirusa HIV-1. Wyniki oznaczeń biologicznych potwierdzają efektywność skonstruowanego protokołu VS. Dwa preparaty wykazały istotną aktywność antywirusową, jednak towarzysząca im cytotoksyczność powoduje konieczność optymalizacji ich struktur chemicznych w celu poprawienia profilu farmakokinetycznego. Związki te mogą więc stanowić struktury wiodące w dalszych badaniach nad poszukiwaniem nowych leków aktywnych względem wirusa HIV.

Rozdziały od 4.4.1 do 4.4.6 zawierają opis badań i czynności związanych z projektowaniem i tworzeniem poszczególnych modułów protokołu VS. Opis

⁷ Projekt badawczy opracowania protokołu wirtualnego skrinigu inhibitorów enzymu integrazy wirusa HIV-1 powstał we współpracy z zespołem badawczym Zakładu Chemii Leków Instytutu Farmakologii Polskiej Akademii Nauk w Krakowie. W szczególności pragnę podziękować Panu mgr Dawidowi Warszyskiemu – pracownikowi Zakładu Chemii Leków Instytutu Farmakologii PAN – za zaangażowanie i współpracę w realizacji poszczególnych etapów obliczeń. Część analiz wykonano z wykorzystaniem skryptów własnych Zakładu Chemii Leków IF PAN.

⁸ Badania biologiczne przeprowadzono w Samodzielnej Pracowni Wirusologii Narodowego Instytutu Leków w Warszawie pod kierownictwem Pana dr Andrzeja Piasek.

Rozdziały od 4.4.1 do 4.4.6 zawierają opis badań i czynności związanych z projektowaniem i tworzeniem poszczególnych modułów protokołu VS. Opis konstrukcji pełnego protokołu VS zamieszczono w Rozdziale 4.4.7, natomiast jego praktyczne zastosowanie w skryningu baz komercyjnie dostępnych związków w celu poszukiwania nowych inhibitorów IN zamieszczono w Rozdziale 4.4.8. Rozdział 4.4.9 zawiera wyniki testów biologicznych nowych, potencjalnych inhibitorów IN wyłonionych z wykorzystaniem opracowanego protokołu VS.

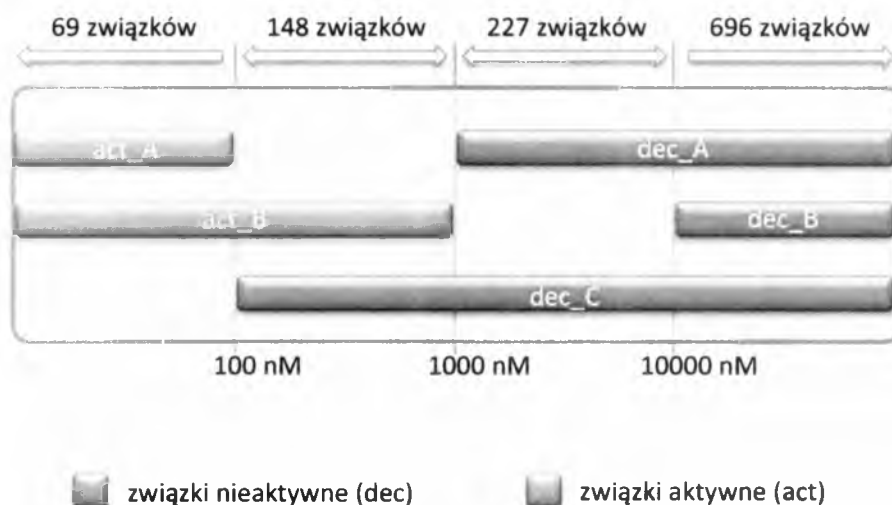
4.4.1 BIBLIOTEKA INHIBITORÓW ENZYMU INTEGRAZY WIRUSA HIV-1

Wykonano eksplorację bazy danych chemicznych ChEMBL (<https://www.ebi.ac.uk/chembl/>, wersja 12) [47,48] w wyniku której wyekstrahowano 1140 struktur posiadających udokumentowaną aktywność w kierunku hamowania enzymu IN wirusa HIV-1. Struktury wyłonił, stosując zapytanie tekstowe w oparciu o pełną nazwę celu molekularnego zgodnie z nomenklaturą stosowaną w bazie: „*Human immunodeficiency virus type 1 integrase*”. Zaimportowane dane (struktury chemiczne, wartości wybranych deskryptorów molekularnych oraz adnotacje na temat aktywności farmakologicznej wraz z referencjami literaturowymi) zapisano na dysku lokalnym i uporządkowano. Sprawdzone poprawność chemiczną zaimportowanych struktur (aplikacja Standardizer grupy ChemAxon) oraz usunięto duplikaty (funkcja Overlap Analysis aplikacji Instant JChem). W ten sposób utworzono bibliotekę inhibitorów IN wirusa HIV-1. Do efektywnego zarządzania informacją chemiczną w powstałej bibliotece związków wykorzystano aplikację Instant JChem grupy ChemAxon.

4.4.2 KOMPOZYCJA I KONSTRUKCJA ZBIORÓW UCZĄCYCH

Ze względu na stopień aktywności (wartość parametru IC_{50}) związki wyjściowej biblioteki podzielono na podzbiory oraz określono umowne zakresy wartości parametru IC_{50} , dla których zdefiniowano „aktywność” bądź „nieaktywność” w kierunku hamowania IN. Rysunek 4.35 przedstawia sposób definiowania

poszczególnych zbiorów związków aktywnych i nieaktywnych w zależności od przyjętego zakresu wartości parametru IC_{50} . W ten sposób wyodrębniono dwa zbiory związków aktywnych: bardzo aktywne (act_A: $IC_{50} \leq 100$ nM) i słabo aktywne (act_B: $IC_{50} \leq 1000$ nM).



Rysunek 4.35 Schemat podziału biblioteki inhibitorów IN na zbiory związków aktywnych/nieaktywnych w zależności od zdefiniowanych zakresów wartości parametru IC_{50} .

Związki aktywne (act_A oraz act_B) zestawiono z nieaktywnymi, przy czym nieaktywność definiowano na trzy różne sposoby:

1. pierwszą grupę stanowiły związki, których brak inhibicji enzymu IN stwierdzono eksperymentalnie, a dla uzyskanych wyników istnieją odnośniki literaturowe (ang. decoys, dec); zdefiniowano trzy zbiory związków nieaktywnych: (i) zbiór dec_A, (ii) zbiór dec_B oraz (iii) zbiór dec_C (sposób definiowania poszczególnych zbiorów związków aktywnych/nieaktywnych w zależności od przyjętego zakresu wartości parametru IC_{50} przedstawia Rysunek 4.35);
2. drugą grupę stanowiła populacja leków o kierunkach aktywności farmakologicznej innych niż hamowanie IN (ang. assumed inactives, ass inact); dane uzyskano metodą eksploracji bazy DrugStore (<https://www.ebi.ac.uk/chembl/db/drugstore>);
3. trzecia grupa związków nieaktywnych wybierana była w oparciu o algorytm DUD (ang. directory of useful decoys, DUD) [157]. Opis metody pozwalającej na

konstrukcję zbioru związków nieaktywnych z wykorzystaniem algorytmu DUD, zawarto w Rozdziale 5.2.

Zbiory związków aktywnych (act_A oraz act_B) zestawiano z nieaktywnymi, stosując różne kombinacje połączeń ze względu na zdefiniowany rodzaj nieaktywności (definicje nieaktywności opisane trzema punktami w pierwszym akapicie rozdziału) oraz ustalono różny stosunek liczebności związków aktywnych i nieaktywnych. W ten sposób utworzono trzynaście zbiorów o różnej kompozycji związków aktywnych i nieaktywnych w kierunku hamowania IN. Kompozycje zbiorów przedstawia Tabela 4.6.

Tabela 4.6 Kompozycja zbiorów.

Indeks zbioru	Związki aktywne	Związki nieaktywne	Stosunek aktywne : nieaktywne
I	act_A	dec_A	~1:13
II	act_B	dec_B	~1:3
III	act_A	dec_C	~1:15,5
IV	act_B	dec_A	~1:4
V	act_A	DUD	1:9
VI	act_A	DUD	1:36
VII	act_B	DUD	1:9
VIII	act_B	DUD	1:36
IX	act_A	leki (ass inact)	1:9
X	act_A	leki (ass inact)	1:5
XI	act_B	leki (ass inact)	1:5
XII	act_A	leki (ass inact)	1:1
XIII	act_B	leki (ass inact)	1:1

Dla każdego z trzynastu skomponowanych zbiorów wydzielono dwa rozłączne zbiory: zbiór uczący i testowy. Liczebność zbioru uczącego ustalono jako 1/3 liczebności zbioru wyjściowego. Pozostała część utworzyła zbiór testowy. Przy czym zbiór uczący wybierano dla każdego z trzynastu zbiorów wyjściowych na cztery różne sposoby:

1. wybór losowy;
2. zbiór o maksymalnej różnorodności;
3. zbiór reprezentatywny;
4. zbiór tworzony metodą LSCO (ang. leave several clusters out).

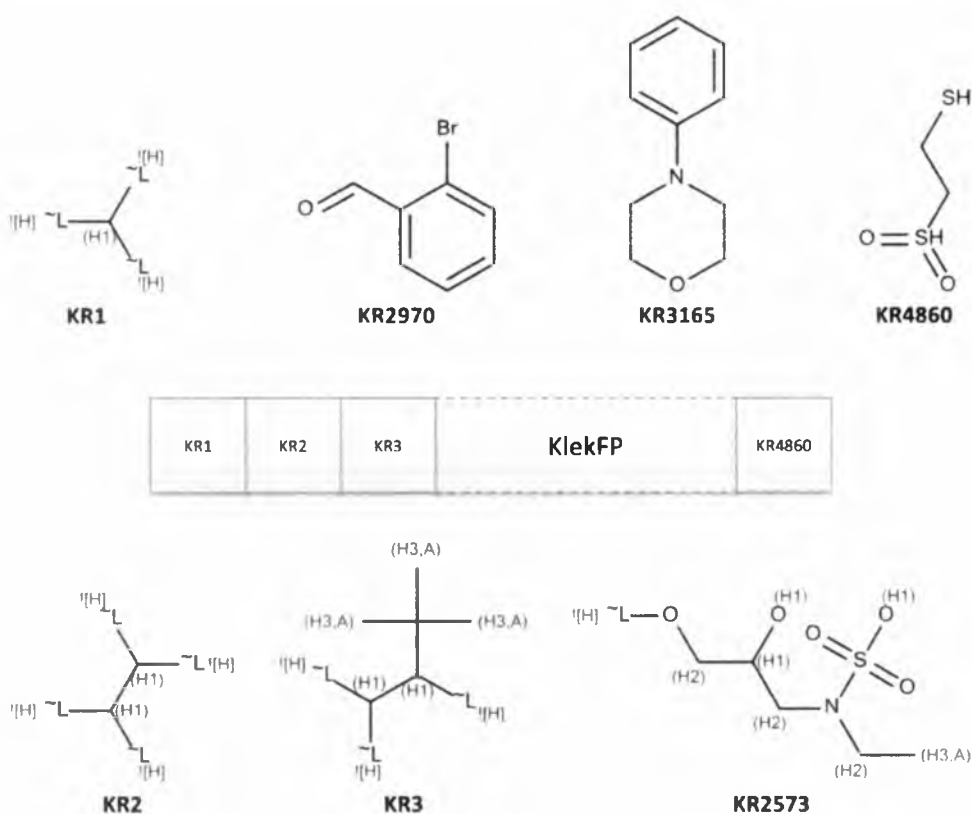
W celu utworzenia zbioru reprezentatywnego oraz tworzonego metodą LSCO wykonano hierarchiczne klastrowanie strukturalne przy użyciu deskryptora daktyloskopowego Molprint2D [158,159] oraz metryki Tanimoto (Tc). Przykłady związków wybierane z poszczególnych klastrów posłużyły do utworzenia reprezentatywnego zbioru uczącego. W metodzie LSCO tworzonego zbioru uczącego poprzez usuwanie jednorazowo kilku klastrów i pozostawienie związków należących tylko do kilku wybranych – sąsiadujących – klastrów. Miejsce odcięcia wybierano czterokrotnie, dlatego w metodzie LSCO dla każdej kompozycji zbioru wyjściowego oznaczonego indeksem od I do XIII utworzono po cztery różne zbiory uczące. W celu rozróżnienia poszczególnych zbiorów, dodano określenie „alfa”, „beta”, „gamma”, „delta”. W dalszym opisie stosowano notację zawierającą indeks zbioru wyjściowego wraz ze wskazaniem na jeden z czterech punktów odcięcia, na przykład „zbiór I alfa”, „zbiór III beta”, „zbiór VII gamma”, „zbiór XIII delta” itd. W metodzie LSCO konieczne było odwrócenie ustalonego powyżej stosunku liczebności zbioru uczącego do testowego, dlatego liczebność zbioru uczącego ustalono jako 2/3 liczebności zbioru wyjściowego. Pozostała część utworzyła zbiór testowy. Czynności związane z konstrukcją zbiorów uczących i testowych wykonano z wykorzystaniem programu Canvas pakietu Schrödinger.

4.4.3 REPREZENTACJA NUMERYCZNA CZĄSTECZEK CHEMICZNYCH

Jako deskryptor reprezentujący struktury chemiczne wybrano deskryptor daktyloskopowy Klekoty-Rotha (KlekFP) [160]. Deskryptor ten składa się z 4860 bitów, gdzie każde pole koduje obecność/brak odpowiedniego fragmentu molekularnego w strukturze rozpatrywanej cząsteczki. Schemat deskryptora daktyloskopowego KlekFP przedstawia Rysunek 4.36. Wyboru deskryptora reprezentującego badane cząsteczki inhibitorów IN dokonano ze względu na fakt, iż fragmenty molekularne jakie koduje deskryptor KlekFP opisano jako elementy podstruktur, które są istotne dla bioaktywności związków chemicznych [160].

Autorzy badań [160] wykonali skrining chemicznych baz danych pod kątem aktywności względem dwudziestu czterech różnych kierunków stymulacji farmakologicznej.

Związek uznawano za aktywny jeżeli dawał pozytywny wynik w przynajmniej jednym z dwudziestu czterech wybranych oznaczeń biologicznych. W kolejnym kroku każdy związek, który został wskazany jako aktywny, poddano fragmentacji. Zastosowano sześć różnych algorytmów fragmentacji, między innymi protokół RECAP [110]. W ostatnim etapie utworzono bibliotekę wygenerowanych fragmentów i spośród wszystkich wybrano te, które powielały się przynajmniej pięciokrotnie – co oznacza, że wybrano fragmenty, które stanowiły elementy podstruktur przynajmniej pięciu związków aktywnych. W ten sposób wyodrębniono 4860 unikatowych podstruktur, istotnych z punktu widzenia bioaktywności związków chemicznych.



Rysunek 4.36 Deskryptor daktyloskopowy Klekoty-Rotha (KlekFP) i przykłady fragmentów molekularnych, kodowanych w wybranych bitach.

Cząsteczki chemiczne wchodzące w skład poszczególnych zbiorów uczących i testowych, których konstrukcję opisano w Rozdziale 4.4.2, przedstawiono w reprezentacji deskryptora daktyloskopowego KlekFP. Obliczenia wykonano z wykorzystaniem oprogramowania PaDEL-descriptor [161].

4.4.4 BADANIE EFEKTYWNOŚCI METOD UCZENIA MASZYNOWEGO I WYBÓR
KLASYFIKATORA O NAJWYŻSZYCH PARAMETRACH PREDYKCJI

Wybrano siedem różnych metod uczenia maszynowego (ML), których skuteczność przetestowano na wszystkich utworzonych kompozycjach zbiorów, opisanych w Rozdziale 4.4.2. Wyboru metod dokonano ze względu na algorytm uczenia, tak aby badane klasyfikatory stanowiły reprezentantów wywodzących się z różnych grup ML. Tabela 4.7 zawiera wykaz metod wraz ze wskazaniem na algorytm uczenia i krótkim opisem poszczególnych grup. Obliczenia wykonano z wykorzystaniem oprogramowania WEKA Data Mining Software [162].

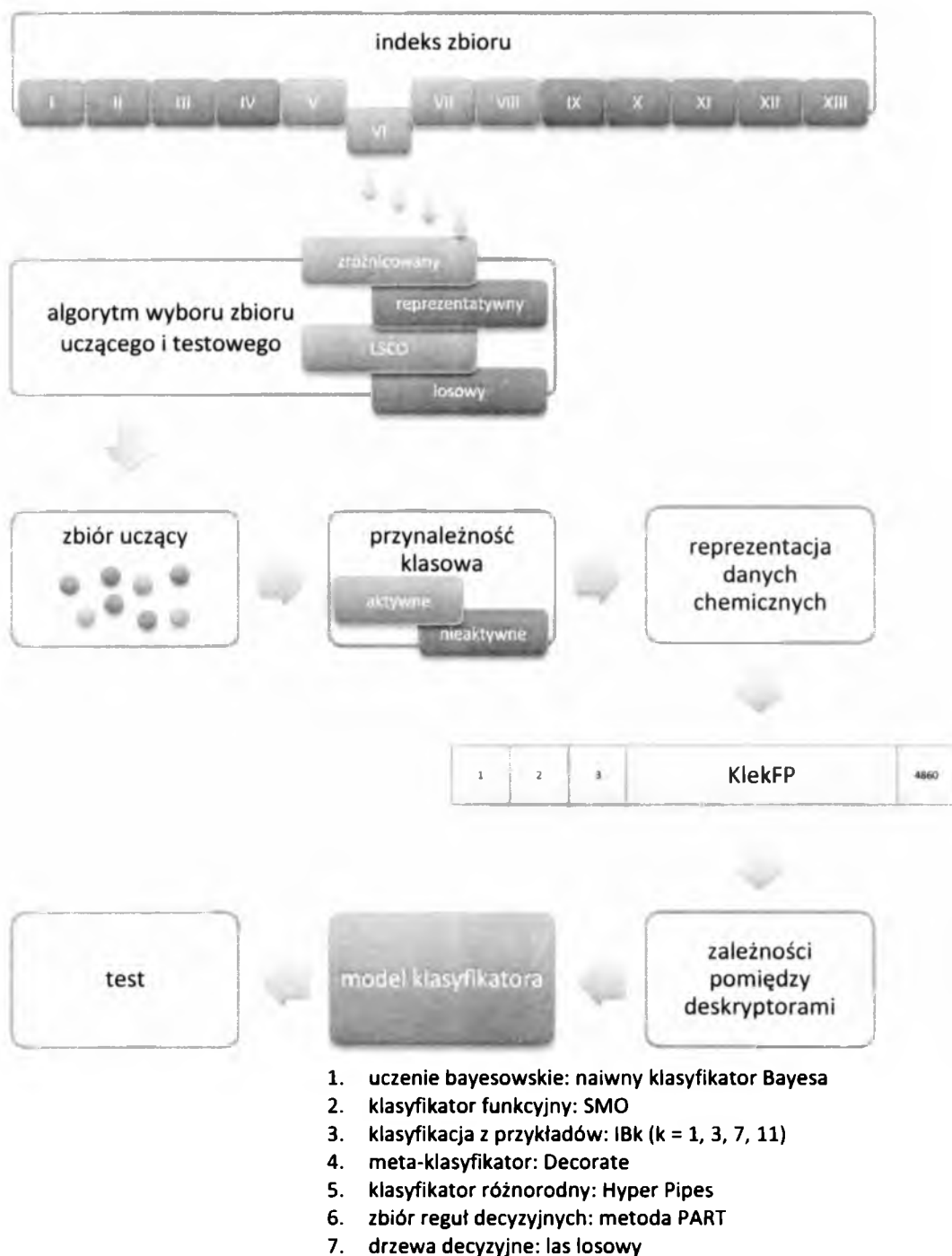
Tabela 4.7 Wykaz testowanych metod uczenia maszynowego (ML) wraz z krótką charakterystyką algorytmu.

Metoda ML ^a	Algorytm uczenia	Rodzina metod	Referencje
Naive Bayes (bayes)	naiwny klasyfikator Bayesa	Klasyfikatory probabilistyczne, oparte na twierdzeniu Bayesa, wiążącym prawdopodobieństwa warunkowe zdarzeń.	[163]
SMO (functions);	algorytm minimalnej optymalizacji sekwencyjnej	Klasyfikator funkcyjny.	[164–166]
IBk (lazy);	metoda <i>k</i> -najbliższych sąsiadów	Zbiór metod określanych jako „uczenie leniwe”. Klasyfikacja następuje poprzez zapamiętanie wszystkich przykładów i przyrównanie nowych obiektów do najbardziej podobnych instancji zbioru uczącego.	[167]
Decorate (meta);	meta-klasyfikator	Rodzina metod, których algorytm pozwala na doskonalenie procesu uczenia poprzez ocenę działania powstającego modelu, wykonaną na zbiorze uczącym i następnie dokonaniu takich zmian, które pozwolą zwiększyć skuteczność klasyfikacji jeszcze przed zbudowaniem finalnego modelu.	[168,169]
Hyper Pipes (misc);	Hyper Pipes to bardzo prosty, szybki algorytm, pozwalający na efektywną klasyfikację nawet w przypadku bardzo dużej liczby atrybutów.	Klasyfikatory różnorodne.	[170]

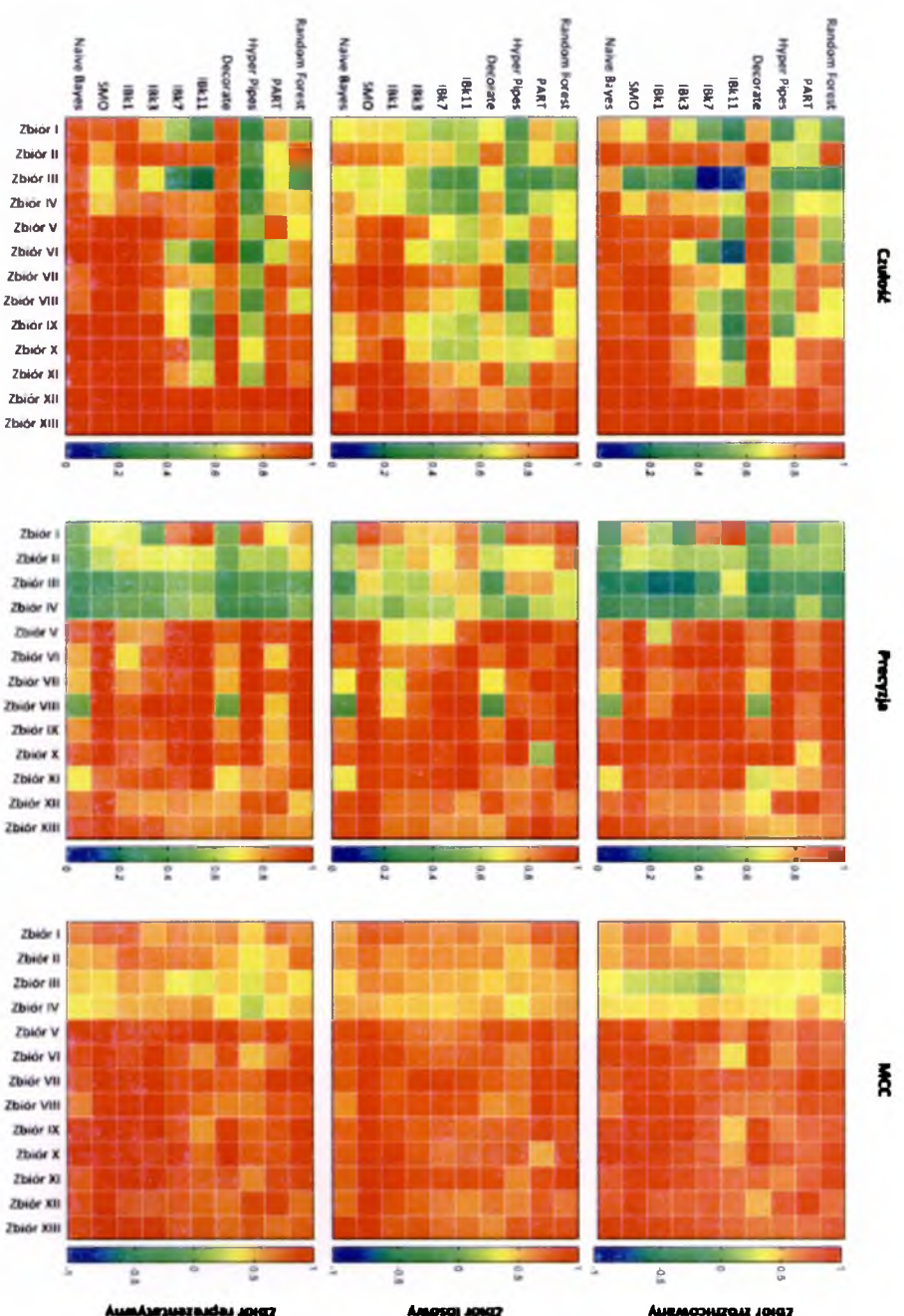
PART (rules);	zbiór reguł decyzyjnych	Reguły decyzyjne opisują związki między atrybutami warunkowymi, a atrybutem decyzyjnym z wykorzystaniem implikacji. Warunki wyrażone formułą logiczną pozwalają wskazać wartość atrybutu decyzyjnego.	[171]
Random Forest (trees)	las losowy	Drzewa decyzyjne to graficzna metoda wspomagania procesu decyzyjnego. Klasyfikacja przy pomocy lasu odbywa się na zasadzie głosowania drzew – obserwacja zostanie zaliczona do najliczniej reprezentowanej klasy na podstawie analizy wyników uzyskanych od wszystkich drzew.	[172]
^a Nomenklatura zgodnie ze specyfikacją oprogramowania WEKA Data Mining Software [162].			

W oparciu o siedem różnych metod ML (wymienionych w tabeli powyżej) oraz skonstruowane zbiory uczące/testowe (Rozdział 4.4.2) utworzono łącznie 1950 klasyfikatorów binarnych. Schemat budowy klasyfikatora przedstawia Rysunek 4.37.

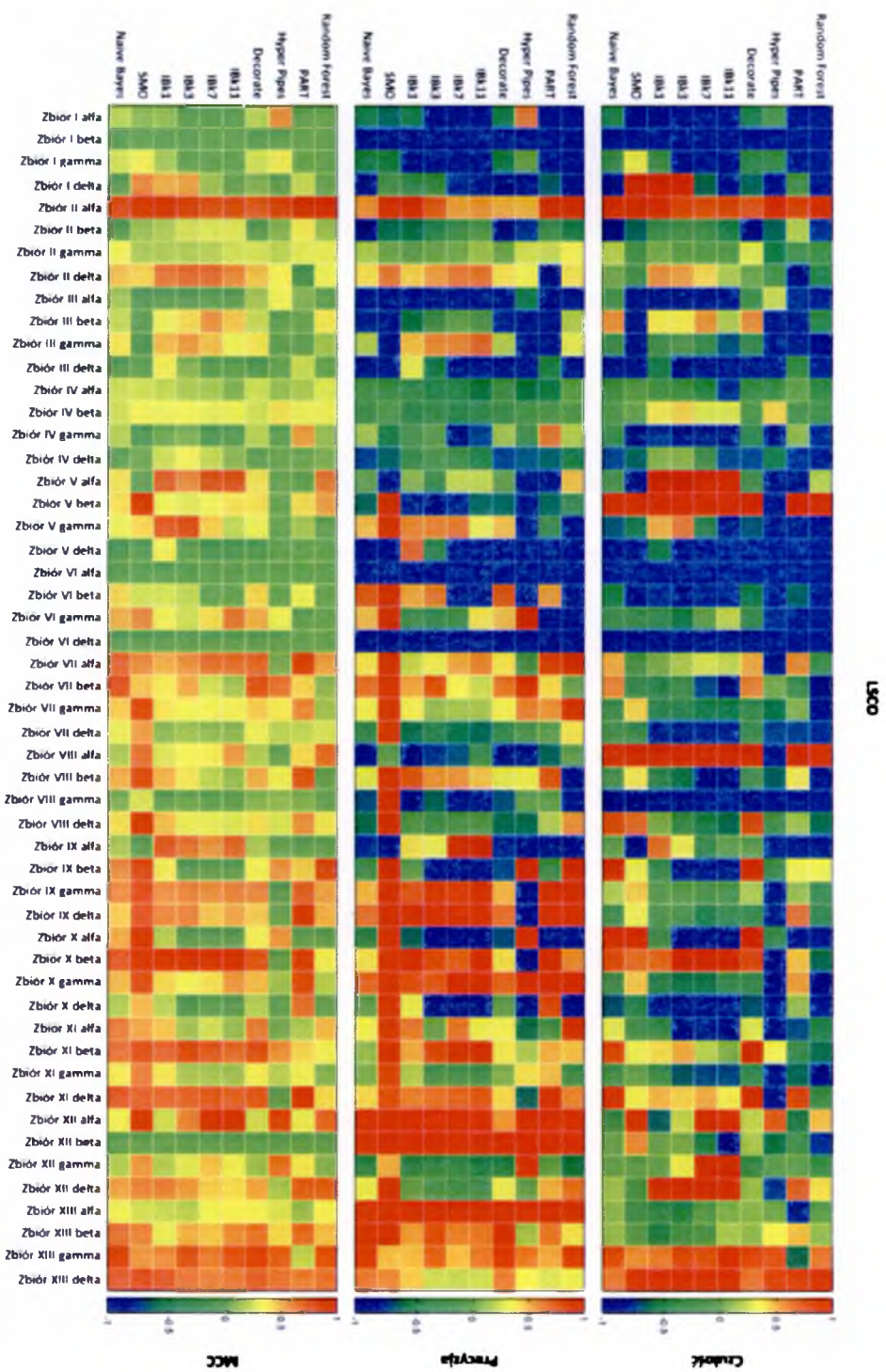
Do oceny efektywności procesu klasyfikacji wykorzystano trzy parametry: czułość, precyzję i parametr korelacyjny Matthews’a (MCC). Zestawienie wyników przedstawiono w postaci macierzy, w której wyliczone wartości poszczególnych parametrów zakodowano posługując się paletą kolorów. Kolor niebieski oznacza wartości niskie, kolor zielony i żółty wskazuje przechodzenie do obszarów o wartościach coraz wyższych i kolor czerwony koduje najwyższe wartości parametrów. Wyniki zebrano i przedstawiono w dwóch panelach (Rysunek 4.38 oraz Rysunek 4.39).



Rysunek 4.37 Schemat klasyfikatora.



Rysunek 4.38 Panel wyników efektywności metod ML, testowanych na zbiorze inhibitorów IN wirusa HIV-1 z zastosowaniem reprezentacji numerycznej cząsteczek chemicznych w postaci deskryptora daktyloskopowego KlekFP. Zestawiono wartości trzech metryk służących ocenie klasyfikatorów: czułości, precyzji oraz parametru korelacyjnego Matthews (MCC). Uwzględniono rozróżnienie ze względu na algorytm wyboru zbioru uczącego. Panel obejmuje wyniki uzyskane dla klasyfikatorów uczonych na zbiorze treningowym generowanym w oparciu o algorytm różnicowany, losowy oraz reprezentatywny.



Rysunek 4.39 Panel wyników efektywności metod ML, testowanych na zbiorze inhibitorów IN wirusa HIV-1 z zastosowaniem reprezentacji numerycznej cząsteczek chemicznych w postaci deskryptora daktyloskopowego KlefP. Zestawiono wartości trzech metryk służących ocenie klasyfikatorów: czułości, precyzji oraz parametru korelacyjnego Matthews (MCC). Panel obejmuje wyniki uzyskane dla klasyfikatorów uczonych na zbiorze treningowym generowanym w oparciu o metodę LSCO.

W celu wyodrębnienia najskuteczniejszego klasyfikatora utworzono ranking metod w oparciu o metrykę MCC (Tabela 4.8). MCC to uniwersalna metryka stosowana w ocenie klasyfikatorów binarnych, która rozpatruje zależności pomiędzy *TP*, *FP* i *N* (Równanie 5.4 w Rozdziale 5.3), a jej niewątpliwą zaletą jest stosowalność nawet w przypadku zbiorów o dużej dysproporcji klas. MCC przyjmuje wartości z przedziału od -1 do 1, gdzie 1 oznacza bezbłędną predykcję, 0 predykcję nie lepszą niż losowa, a -1 oznacza błędną predykcję. Policzono wartość średnią parametru MCC przez wszystkie trzynaście zbiorów w obrębie każdego z czterech różnych algorytmów wyboru zbioru uczącego. Wyniki wskazały klasyfikator funkcyjny SMO jako metodę o największej efektywności procesu klasyfikacji badanej grupy związków. Tylko nieznacznie gorsze wyniki klasyfikacji uzyskuje się stosując klasyfikator IBk (dla $k=1$).

Tabela 4.8 Ranking metod ML w oparciu o metrykę MCC z rozróżnieniem na jeden z czterech algorytmów wyboru zbioru uczącego. Miejsce w rankingu oznaczone jako „1” wskazuje klasyfikator o najlepszych parametrach predykcji.

Metoda ML ^a	Ranking			
	Zbiór zróżnicowany	Zbiór losowy	Zbiór reprezentatywny	Metoda LSCD
SMO	1	1	1	2
IBk (dla $k = 1$)	2	2	2	1
NaiveBayes	3	8	7	10
IBk (dla $k = 3$)	4	5	3	7
PART	5	4	6	9
Random Forest	6	3	4	4
IBk (dla $k = 7$)	7	7	5	8
Decorate	8	9	8	6
Hyper Pipes	9	10	10	3
IBk (dla $k = 11$)	10	6	9	5
^a Nomenklatura zgodnie ze specyfikacją oprogramowania WEKA Data Mining Software [162].				

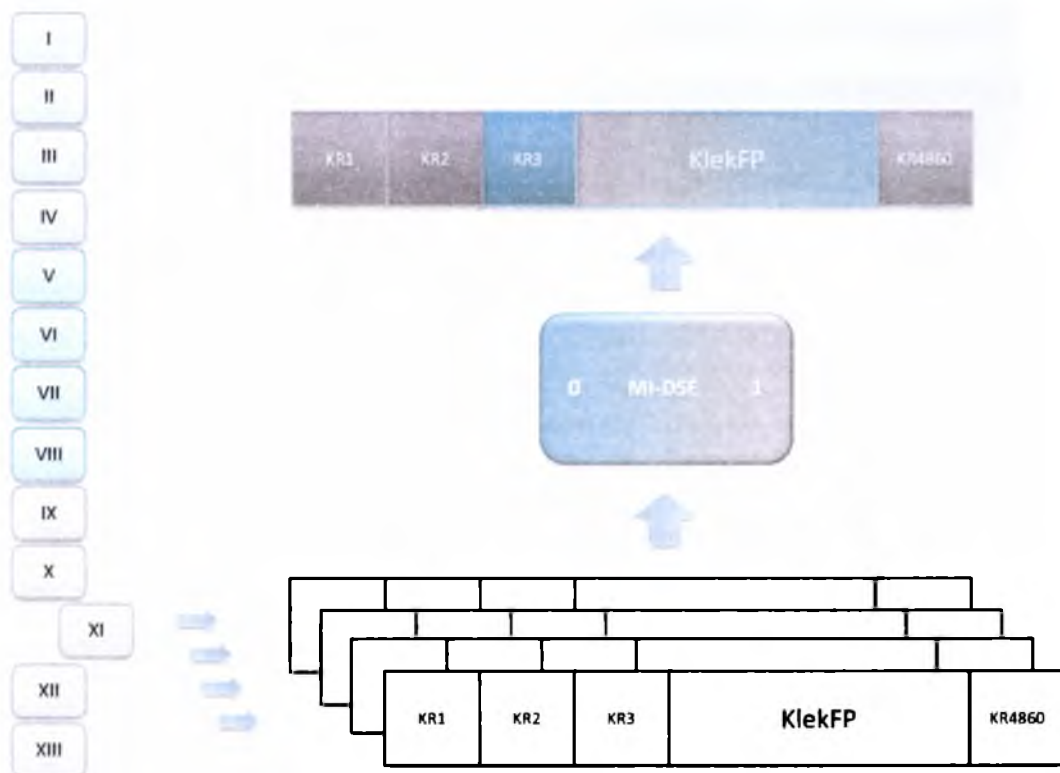
4.4.5 BADANIE MOŻLIWOŚCI IMPLEMENTACJI METODY MI-DSE W OKREŚLANIU FRAGMENTÓW MOLEKULARNYCH RÓŻNICUJĄCYCH ZBIORY ZWIĄZKÓW AKTYWNYCH I NIEAKTYWNYCH BIOLOGICZNIE

Parametr MI-DSE [173] umożliwia porównanie rozkładu wartości wybranego deskryptora molekularnego w dwóch różnych zbiorach cząsteczek. Parametr ten

przyjmuje wartości z przedziału od 0 do 1. Wartość parametru równa 1 charakteryzuje deskryptor idealnie dyskryminujący dwa rozważane zbiory związków chemicznych. Metoda ta – oparta na teorii informacji – jak dotychczas nie została praktycznie zaimplementowana w celu różnicowania zbiorów związków opisanych deskryptorami strukturalnymi.

Przetestowanie możliwości implementacji metody MI-DSE [173] w określaniu fragmentów molekularnych różnicujących zbiory związków aktywnych i nieaktywnych biologicznie wymagało obliczenia wartości wskaźnika MI-DSE dla poszczególnych strukturalnych deskryptorów fragmentarycznych badanej populacji inhibitorów IN wirusa HIV-1. Analizę wykonano w oparciu o numeryczną reprezentację cząsteczek chemicznych utworzoną z zastosowaniem deskryptora substrukturalnego KlekFP [160]. Deskryptor daktyloskopowy KlekFP omówiono w Rozdziale 4.4.3. Reprezentacje numeryczne wszystkich cząsteczek chemicznych tworzących badane zbiory wygenerowano w wykorzystaniem oprogramowania PaDEL-descriptor [161].

Parametr MI-DSE wyliczono dla każdego bitu tworzącego deskryptor daktyloskopowy KlekFP, aby następnie sprawdzić czy rozkład uzyskanych wartości pozwoli wskazać bity (fragmenty molekularne), które w znaczący sposób różnicują zbiory związków aktywnych i nieaktywnych. Analizę wykonano w obrębie każdej z trzynastu kompozycji zbiorów związków aktywnych i nieaktywnych. Kompozycje zbiorów przedstawia Tabela 4.6 w Rozdziale 4.4.2. Schemat postępowania ilustruje Rysunek 4.40.



Rysunek 4.40 Schemat obliczenia wartości wskaźnika MI-DSE dla poszczególnych bitów deskryptora daktyloskopowego KlekFP w trzynastu wyodrębnionych zbiorach inhibitorów IN wirusa HIV-1.

Zbadano wpływ konstrukcji i kompozycji zestawień zbiorów związków aktywnych z nieaktywnymi na wyliczane wartości parametru MI-DSE. Uzyskane wyniki zebrano i opisano w trzech grupach uwzględniających podział ze względu na przyjętą definicję nieaktywności. Pierwszą grupę obejmowały zbiory o indeksach I-IV (związki nieaktywne: dec), drugą grupę stanowiły zbiory o indeksach V-VIII (związki nieaktywne: DUD), natomiast trzecia grupa to zbiory o indeksach IX-XIII (związki nieaktywne: ass inact). Interpretację uzyskanych wyników zestawiono w trzech poniższych punktach.

1. Pierwsza grupa (kompozycje zbiorów I-IV).

Wartości MI-DSE maleją, gdy granica ustalona na podstawie wartości parametru IC_{50} pomiędzy związkami aktywnymi, a nieaktywnymi przechodzi z nieostrej na ostrą. Jest to zrozumiałe, ponieważ przy ostrej granicy odróżnienie aktywnego związku od zdefiniowanego jako nieaktywny jest

szczególne trudne. Zaobserwowano ponadto, że na wartości MI-DSE zasadniczo nie wpływa stosunek liczby związków aktywnych do nieaktywnych.

2. Druga grupa (kompozycje zbiorów V-VIII).

Zwiększenie wartości MI-DSE obserwowane jest tylko dla wąskiego zakresu wartości parametru IC_{50} opisującego zbiór związków aktywnych oraz przy równoczesnym zwiększeniu liczby związków nieaktywnych (DUD). Zwiększenie zakresu związków aktywnych powoduje zmniejszenie skuteczności metody MI-DSE nawet przy równoczesnym zwiększeniu liczby związków nieaktywnych (DUD). Obserwacja ta pozwala wyciągnąć wniosek, iż najaktywniejsze związki są tak unikalne, że wyróżniają się wśród związków zdefiniowanych jako nieaktywne, bez względu na liczebność populacji nieaktywnych.

3. Trzecia grupa (kompozycje zbiorów IX-XIII).

Najwyższe wartości MI-DSE uzyskujemy gdy liczebność związków zdefiniowanych jako nieaktywne jest możliwie duża. Metoda jest bardziej selektywna dla wąskiego zakresu wartości parametru IC_{50} , opisującego zbiór związków aktywnych.

Analizę wyliczonych wartości parametru MI-DSE przeprowadzono dwukierunkowo. W pierwszym aspekcie zbadano wpływ konstrukcji i kompozycji zbiorów na skuteczność metody MI-DSE w różnicowaniu zbiorów związków aktywnych i nieaktywnych (opis powyżej). Aspekt drugi zakładał próbę wskazania takich fragmentów molekularnych (zakodowanych w poszczególnych bitach deskryptora daktyloskopowego KlekFP), które z dużym prawdopodobieństwem mogą stanowić uprzywilejowane motywy strukturalne badanej populacji inhibitorów IN. Opis uzyskanych wyników przedstawiono w rozdziale poniżej.

4.4.6 PRÓBA WSKAZANIA NOWYCH UPRZYWILEJOWANYCH MOTYWÓW STRUKTURALNYCH W ZBIORZE INHIBITORÓW ENZYMU INTEGRAZY WIRUSA HIV-1

W celu wyselekcjonowania potencjalnie uprzywilejowanych fragmentów molekularnych (ang. privileged structures, PS) wśród badanej populacji inhibitorów IN, wyliczone wartości parametru MI-DSE (Rozdział 4.4.5) przeanalizowano dla poszczególnych bitów tworzących deskryptor daktyloskopowy KlekFP.

Dla każdego zbioru związków (kompozycje zbiorów przedstawia Tabela 4.6 w Rozdziale 4.4.2) uporządkowano bity deskryptora daktyloskopowego KlekFP według malejącej wartości wskaźnika MI-DSE. Dla każdej klasyfikacji wybrano dziesięć fragmentów o najwyższych wartościach MI-DSE. Tabela 4.9 przedstawia bit deskryptora KlekFP wraz z kodowaną strukturą fragmentu oraz liczbę określającą ile razy dany fragment molekularny pojawiał się w czołowej dziesiątce. Częstotliwość występowania fragmentu molekularnego uznano za wskaźnik uprzywilejowania (wskaźnik PS), a jego wartość wykorzystano w funkcji wagi w dalszych obliczeniach.

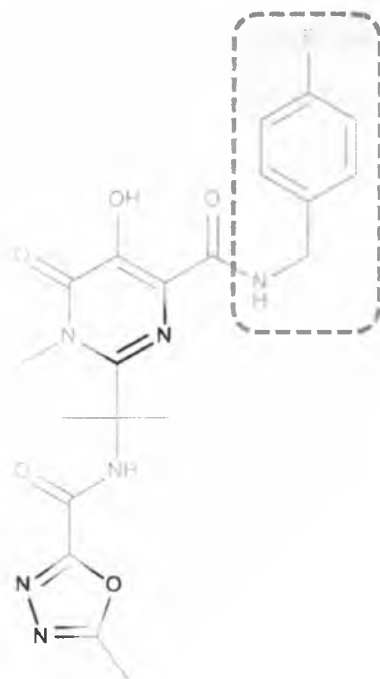
Tabela 4.9 Fragmenty molekularne wskazane metodą MI-DSE jako uprzywilejowane elementy podstruktury związków chemicznych stanowiących inhibitory IN wirusa HIV-1. Fragmenty molekularne przedstawiono wykorzystując notację liniowego zapisu struktur chemicznych – SMARTS.

Lp.	Bit deskryptora KlekFP	Fragment molekularny [notacja SMARTS]	Wartość wskaźnika uprzywilejowania PS
1	KR590	[!#1][CH2]c1[cH][cH]c(F)[cH][cH]1	13
2	KR1798	[!#1]c1[cH][cH]c(F)[cH][cH]1	11
3	KR1148	[!#1][OH]	10
4	KR2949	[OH]	9
5	KR3548	Cc1ccc(F)cc1	9
6	KR4067	Fc1ccccc1	9
7	KR582	[!#1][CH2]c1[cH][cH]c(!#1)[cH][cH]1	8
8	KR2260	[!#1]F	7
9	KR4032	F	7
10	KR1634	[!#1]c1[cH][cH][cH]nc1[!#1]	6
11	KR297	[!#1][CH2][!#1]	5
12	KR669	[!#1][CH3]	3
13	KR1193	[!#1]C(=O)[!#1]	3
14	KR1418	[!#1]C(=O)C(=O)[!#1]	3
15	KR1642	[!#1]c1[cH][cH]c(!#1)[cH][cH]1	3
16	KR3025	C=O	3

4. Badania własne

17	KR3706	CCCCC	3
18	KR4472	O=CC=O	3
19	KR1	[!#1][CH](!!#1)!!#1	2
20	KR677	[!#1][NH]!!#1	2
21	KR1406	[!#1]C(=O)[OH]	2
22	KR2983	C(NCc1cccc1)c2cccc2	2
23	KR3750	CCN	2
24	KR4772	Oc1ccnc1	2
25	KR298	[!#1][CH2][CH](!!#1)!!#1	1
26	KR3943	CNc1cccc1	1
27	KR4080	N	1
28	KR4763	Oc1cccc1C=O	1
29	KR4770	Oc1cccc1O	1

Zastosowanie metody MI-DSE z odpowiednio skomponowanym zbiorem związków przedstawionych w postaci reprezentacji numerycznej przy użyciu deskryptora daktyloskopowego KlekFP, umożliwia wskazanie fragmentów molekularnych, które z dużym prawdopodobieństwem mogą stanowić uprzywilejowane motywy strukturalne (PS) badanej przestrzeni chemicznej. Wskazano dwadzieścia dziewięć fragmentów molekularnych, które mogą stanowić uprzywilejowane elementy strukturalne inhibitorów IN wirusa HIV-1. Ciekawostką stanowi fakt, iż fragmenty molekularne wskazane jako uprzywilejowane, nałożone na strukturę raltegrawiru pokrywają go w znacznym stopniu, co przedstawia Rysunek 4.41. Fakt ten sugeruje, iż obecność tych fragmentów w strukturze cząsteczki, może mieć kluczowe znaczenie dla projektowania związków wykazujących hamowanie enzymu IN. Z dużym prawdopodobieństwem fragmenty te będą wpływały na uprzywilejowanie architektury cząsteczki w kierunku zastosowań w terapii przeciw HIV.



Rysunek 4.41 Stopień pokrycia struktury raltegrawiru fragmentami wskazanymi metodą MI-DSE jako uprzywilejowane dla inhibitorów IN wirusa HIV-1. Kolorem niebieskim zilustrowano wspólne motywy strukturalne cząsteczki leku i dwudziestu dziewięciu fragmentów molekularnych wskazanych metodą MI-DSE. Linia przerywaną zakreślono fragment o najwyższej wartości wskaźnika uprzywilejowania PS – bit KR590.

Przeprowadzono również analizę reguł decyzyjnych zbudowanych przez metodę PART. Systemy klasyfikujące oparte na regułach zawierają zwykle wiele – a czasami nawet tysiące – reguł decyzyjnych, z których każda może być oparta na innym zestawie atrybutów warunkowych. Rysunek 4.42 przedstawia fragment pliku tekstowego, zawierającego opis konstrukcji modelu metodą PART dla wybranej kompozycji zbioru uczącego. Przedstawiony fragment pliku zawiera zestawienie czterech reguł decyzyjnych utworzonych przez metodę PART na etapie uczenia modelu w oparciu o zbiór treningowy, generowany z użyciem algorytmu zróżnicowania i o kompozycji określonej indeksem XII. Analiza reguł decyzyjnych zbudowanych przez trzydzieści dziewięć klasyfikatorów PART uczonych na trzynastu kompozycjach zbiorów z rozróżnieniem na trzy różne algorytmy generowania zbioru uczącego (reprezentatywny, zróżnicowany, losowy) dostarczyła ciekawych obserwacji. Reguły decyzyjne przypisujące aktywność związku (przynależność do klasy +1) w wielu przypadkach oparte były na atrybutach warunkowych, wskazujących obecność danego fragmentu molekularnego ($att > 0$), który stanowił fragment wspólny z motywem

strukturalnym wskazanym przez metodę MI-DSE. Analiza zliczeń wszystkich wystąpień takich przypadków, wykonana w plikach wyjściowych klasyfikatorów PART wskazała osiem fragmentów molekularnych wspólnych z fragmentami molekularnymi wskazanymi metodą MI-DSE.

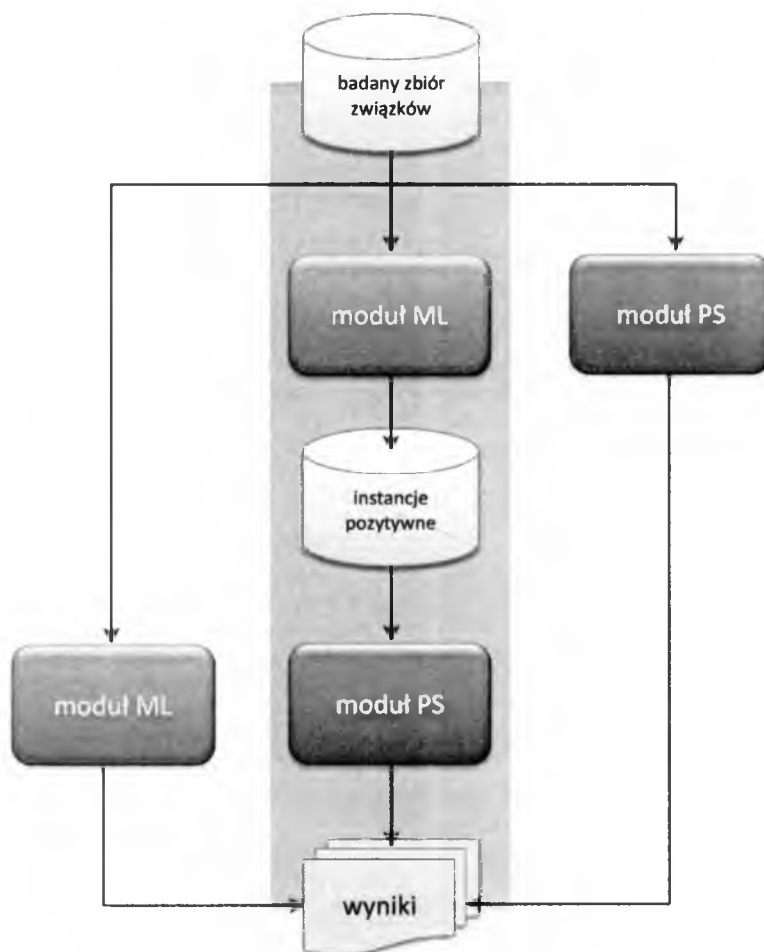
```
Set_XII_diverse_rulesPART
att_1148 > 0 AND
att_4826 <= 0 AND
att_1193 > 0 AND
att_17 <= 0 AND
att_3602 <= 0: +1 (50.0/3.0)
att_3534 <= 0 AND
att_3708 <= 0: -1 (52.0/3.0)
att_1566 <= 0: +1 (4.0)
: -1 (2.0)
Number of Rules : 4
```

Rysunek 4.42 Wycinek pliku tekstowego zawierającego opis konstrukcji klasyfikatora binarnego w oparciu o zbiór reguł tworzonych metodą PART. Kolorem niebieskim zaznaczono atrybuty warunkowe wskazujące obecność danego fragmentu molekularnego ($att > 0$). Atrybuty te stanowią bity wspólne z bitami deskryptora daktyloskopowego KlekFP wskazanymi przez metodę MI-DSE.

4.4.7 KONSTRUKCJA I WERYFIKACJA PROTOKOŁU WIRTUALNEGO SKRINIGU INHIBITORÓW ENZYMU INTEGRAZY WIRUSA HIV-1

Wyliczone wskaźniki uprzywilejowania (Tabela 4.9 w Rozdziale 4.4.6) w połączeniu z metodami ML (wybór klasyfikatora opisano w Rozdziale 4.4.4) posłużyły do opracowania protokołu wirtualnego skrinigu (VS) inhibitorów IN wirusa HIV-1. Protokół obejmuje dwa moduły: (i) moduł PS wykorzystujący informacje na temat uprzywilejowania wybranych fragmentów molekularnych oraz (ii) moduł ML, w którym następuje klasyfikacja związków. Moduły mogą stanowić samodzielne jednostki służące priorytetyzacji (moduł PS) lub klasyfikacji (moduł ML) cząsteczek chemicznych. Możliwe jest także połączenie modułów w kaskadę, w której wyniki pierwszego etapu

(z wykorzystaniem modułu PS) przekazywane są do następującego po nim etapu klasyfikacji. Rysunek 4.43 przedstawia schemat opracowanego protokołu VS.



Rysunek 4.43 Schemat opracowanego protokołu VS.

Skuteczność opracowanej procedury VS zweryfikowano wykonując skringing bazy MDL Drug Data Report (MDDR) Poniższy rozdział stanowi opis uzyskanych wyników.

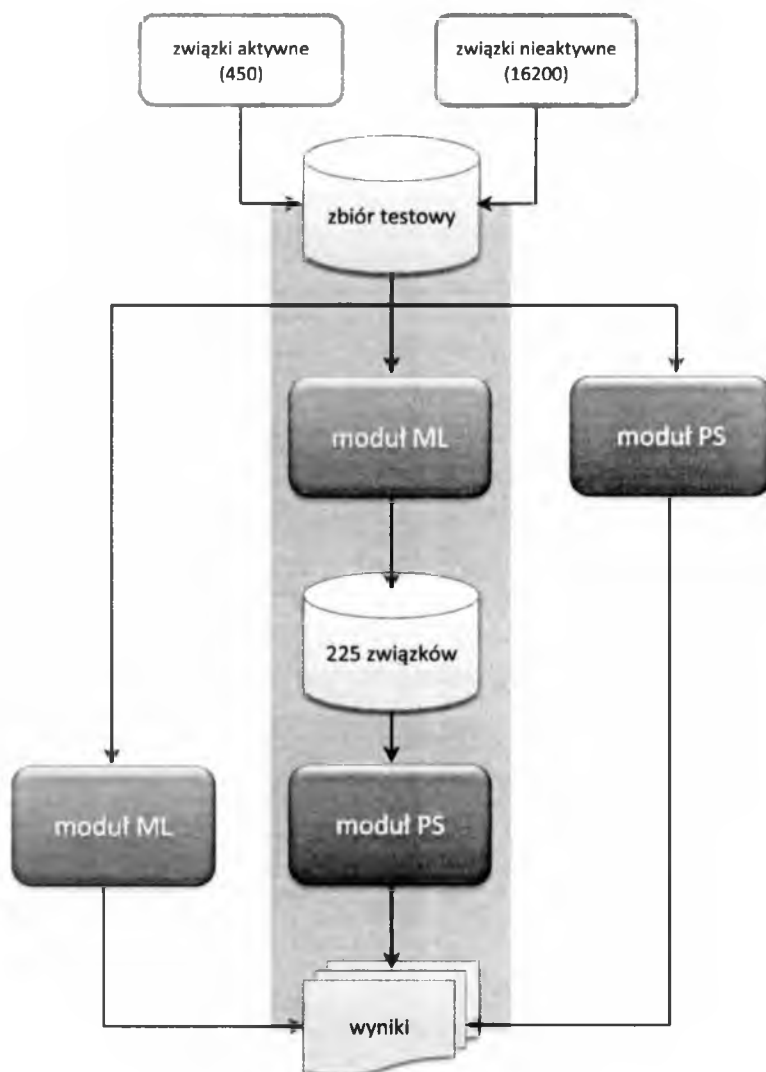
4.4.7.1 WERYFIKACJA SKUTECZNOŚCI OPRACOWANEGO PROTOKOŁU WIRTUALNEGO SKRININGU

W celu weryfikacji skuteczności opracowanego protokołu VS przeprowadzono skringing wirtualny inhibitorów IN wirusa HIV-1, których struktury wyekstrahowano z chemicznej bazy danych MDL Drug Data Report (MDDR). Baza ta dystrybuowana jest odpłatnie przez Accelrys w porozumieniu z Thomson Reuters. Repozytorium zawiera informacje

strukturalne oraz adnotacje na temat aktywności biologicznej i farmakologicznej cząsteczek organicznych. Wersja 2013 bazy obejmuje ponad 150 000 związków chemicznych. Dane zawarte w bazie pochodzą z publikacji naukowych, dokumentacji patentowej, a także materiałów konferencyjnych. Baza posiada własny system klasyfikacji związków pod względem zastosowań terapeutycznych i nie kataloguje wartości liczbowych charakteryzujących bioaktywność substancji.

Wykonano eksplorację bazy danych chemicznych MDL Drug Data Report (MDDR) w wyniku której wyekstrahowano 450 struktur posiadających udokumentowaną aktywność względem enzymu IN wirusa HIV-1. Struktury wyłoniono stosując zapytanie w oparciu o indeks celu molekularnego zgodnie z klasyfikacją stosowaną w bazie. Indeks badanego celu molekularnego w bazie MDDR to „71527”. Z wyekstrahowanego zbioru inhibitorów IN wirusa HIV-1 usunięto struktury wspólne z biblioteką wyjściową (biblioteka inhibitorów IN utworzona w wyniku eksploracji bazy ChEMBL, której konstrukcję opisano w Rozdziale 4.4.1), która służyła do konstrukcji trzynastu kompozycji zbiorów uczących (Rozdział 4.4.2) w trakcie opracowywania protokołu VS.

Utworzono zbiór testowy, który obejmował 450 inhibitorów IN wirusa HIV-1 zestawionych z 16200 związkami nieaktywnymi, które wybrano z zastosowaniem algorytmu DUD [157]. Stosunek związków aktywnych do nieaktywnych wynosił 1:36. Zbiór ten wykorzystano w celu weryfikacji skuteczności opracowanego protokołu VS w trzech różnych konfiguracjach: (i) moduł ML zastosowany samodzielnie, (ii) moduł PS zastosowany samodzielnie oraz (iii) kaskada połączonych modułów PS i ML. Rysunek 4.44 przedstawia schemat przeprowadzonej procedury weryfikującej skuteczność protokołu VS, natomiast w Rozdziałach 4.4.7.1.1 - 4.4.7.1.3 omówiono uzyskane wyniki.



Rysunek 4.44 Skryning wirtualny bazy danych MDL Drug Data Report (MDDR).

4.4.7.1.1 KLASYFIKACJA INHIBITORÓW ENZYMU INTEGRAZY Z WYKORZYSTANIEM METOD UCZENIA MASZYNOWEGO – MODUŁ ML

Moduł ML zbudowano w oparciu o dwa najskuteczniejsze klasyfikatory. Wybór klasyfikatorów opisano w Rozdziale 4.4.4. Zbiory uczące stanowiły zbiory cząsteczek o kompozycji opisanej indeksem VI oraz IX (kompozycje zbiorów przedstawia Tabela 4.6 w Rozdziale 4.4.2) odpowiednio dla klasyfikatora SMO oraz IBk. Wyboru zbiorów uczących dokonano zgodnie z wynikami uzyskanymi w Rozdziale 4.4.4 podczas tworzenia rankingów metod klasyfikacyjnych. Zbiór testowy obejmował 450 inhibitorów IN wirusa HIV-1 zestawionych z 16200 związkami nieaktywnymi, które wybrano

z zastosowaniem algorytmu DUD [157]. Konstrukcję zbioru testowego przedstawiono w Rozdziale 4.4.7.1.

Cząsteczki chemiczne wchodzące w skład zbioru uczącego i testowego przedstawiono w reprezentacji deskryptora daktyloskopowego KlekFP [160]. Reprezentacje numeryczne cząsteczek chemicznych wygenerowano w wykorzystaniem oprogramowania PaDEL-descriptor [161]. Tabela 4.10 zawiera wyniki klasyfikacji uzyskane z zastosowaniem modułu ML. Obliczenia wykonano z wykorzystaniem oprogramowania WEKA Data Mining Software [162]. Czułość, precyzję, MCC oraz metrykę F wyliczono zgodnie z wzorami przedstawionymi w Rozdziale 5.3 – „Metryki oceny efektywności klasyfikatorów binarnych”.

Tabela 4.10 Wyniki klasyfikacji uzyskane z zastosowaniem modułu ML.

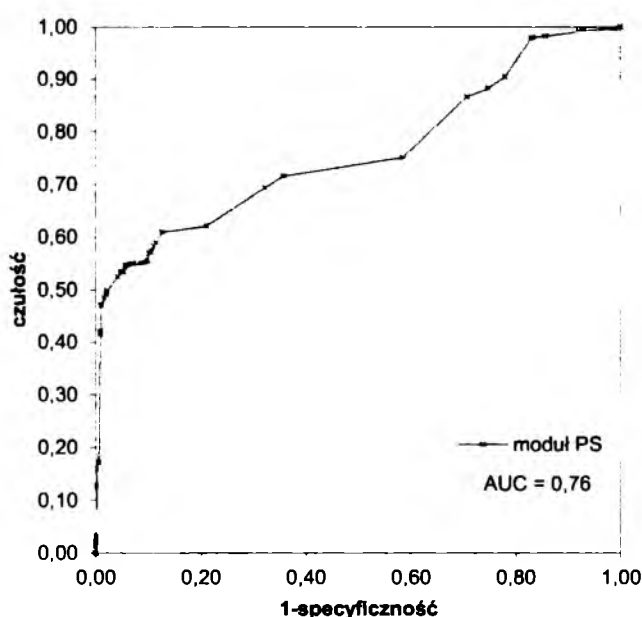
Klasyfikator	TP	FP	FN	TN	Czułość	Precyzja	MCC	metryka F
IBk	147	78	303	16122	0,33	0,65	0,45	0,44
SMO	5	0	445	16200	0,01	1,00	0,10	0,02

Klasyfikator SMO wskazał poprawnie 5 aktywnych związków. Brak związków błędnie wskazanych jako aktywne ($FP = 0$) wskazuje na imponującą precyzję metody. W przypadku metody IBk liczba wskazanych poprawnie związków aktywnych wyniosła 147, ale przy $FP=78$.

4.4.7.1.2 MODUŁ PRIORYTETYZACJI ZWIĄZKÓW CHEMICZNYCH W OPARCIU O WSKAŹNIK UPRIZYWILEJOWANIA – MODUŁ PS

Moduł PS zbudowano w oparciu o zdefiniowany w Rozdziale 4.4.6 wskaźnik uprzywilejowania PS (Tabela 4.9, kolumna 4). Wartość wskaźnika PS wykorzystano w funkcji wagi. Zbiór testowy modułu PS obejmował 450 inhibitorów IN wirusa HIV-1 zestawionych z 16200 związkami nieaktywnymi, które wybierano w oparciu o procedurę DUD [157]. Konstrukcję zbioru testowego przedstawiono w Rozdziale 4.4.7.1. Cząsteczki chemiczne wchodzące w skład zbioru testowego przedstawiono w reprezentacji deskryptora daktyloskopowego KlekFP [160]. Reprezentacje numeryczne cząsteczek chemicznych wygenerowano w wykorzystaniem

oprogramowania PaDEL-descriptor [161]. W kolejnym kroku przeprowadzono analizę zliczeń fragmentów uprzywilejowanych (fragmenty molekularne przedstawia Tabela 4.9 w Rozdziale 4.4.6) występujących jako podstruktury w poszczególnych cząsteczkach zbioru testowego. W ten sposób dla każdej cząsteczki zbioru testowego wyliczono ocenę punktową, którą stanowiła suma wartości wskaźników PS wszystkich fragmentów molekularnych (Tabela 4.9 w Rozdziale 4.4.6) obecnych jako podstruktury rozpatrywanej cząsteczki. Rysunek 4.45 przedstawia krzywą ROC opisywanej metody modułu PS. Metodę konstrukcji krzywych ROC omówiono w Rozdziale 5.3 – „Metryki oceny efektywności klasyfikatorów binarnych”.

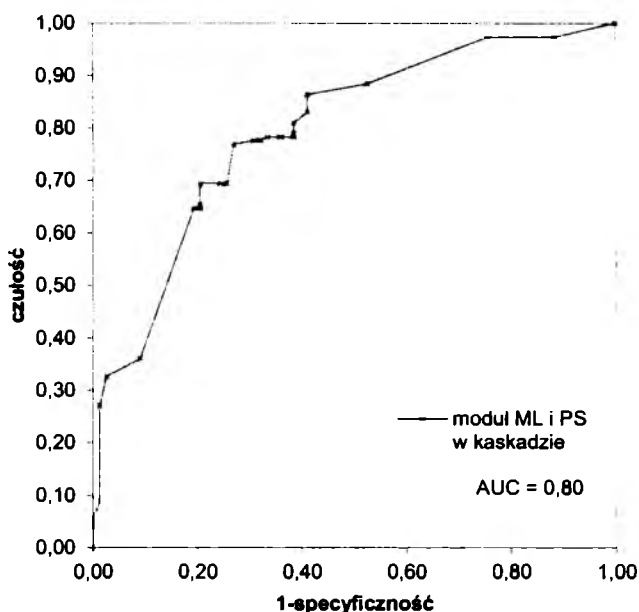


Rysunek 4.45 Wykres ROC dla modułu PS.

4.4.7.1.3 KASKADOWE POŁĄCZENIE MODUŁÓW KLASYFIKACJI ORAZ PRIORYTETYZACJI

Skuteczność opracowanego protokołu VS okazała się najwyższa w przypadku zastosowania kaskady połączonych modułów ML oraz PS. Taka konfiguracja zakłada w pierwszym etapie przeprowadzenie klasyfikacji z użyciem modułu ML. Do drugiego etapu – z wykorzystaniem priorytetyzacji związków w oparciu o punktację PS – zostają przekazane tylko te związki chemiczne, które zostały sklasyfikowane w pierwszym etapie jako *TP*, a więc zostały poprawnie sklasyfikowane jako aktywne. Rysunek 4.46

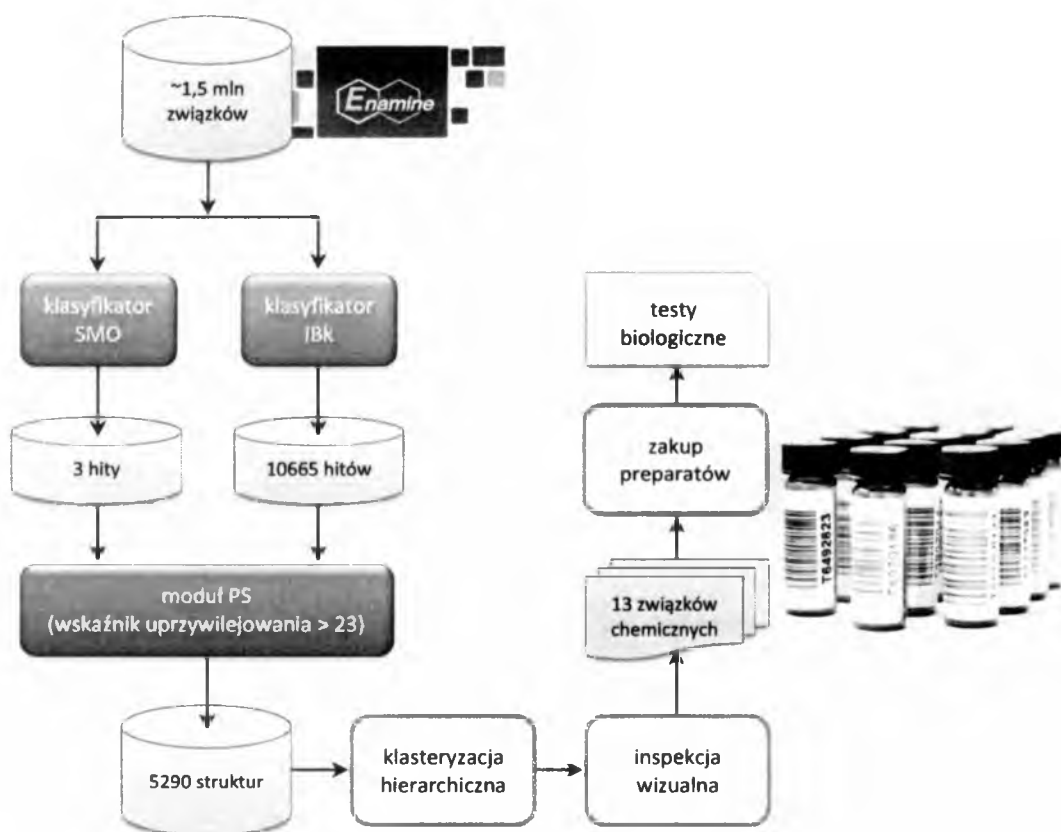
przedstawia krzywą ROC opisywanej metody kaskadowej konfiguracji modułów ML oraz PS. Metodę konstrukcji krzywych ROC omówiono w Rozdziale 5.3 – „Metryki oceny efektywności klasyfikatorów binarnych”.



Rysunek 4.46 Wykres ROC dla modułów ML i PS połączonych w konfiguracji kaskadowej.

4.4.8 SKRINING BAZ KOMERCYJNIE DOSTĘPNYCH ZWIĄZKÓW W OPARCIU O SKONSTRUOWANY PROTOKÓŁ VS

Wykonano skringing bazy komercyjnie dostępnych związków – Enamine (<http://www.enamine.net/>) w celu poszukiwania nowych inhibitorów IN wirusa HIV-1. W roku 2012 firma Enamine dysponowała bazą liczącą ponad 1,5 mln związków chemicznych. Skringing wykonano z zastosowaniem kaskady połączonych modułów ML oraz PS. Opis modułów zamieszczono w Rozdziale 4.4.7.1.1 oraz w Rozdziale 4.4.7.1.2.



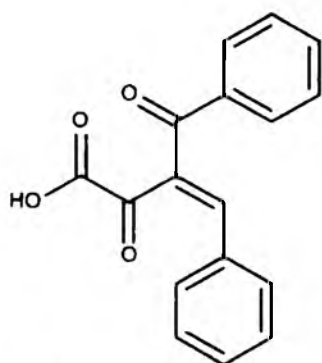
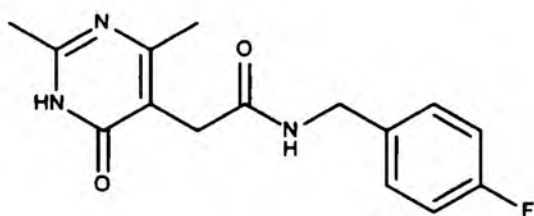
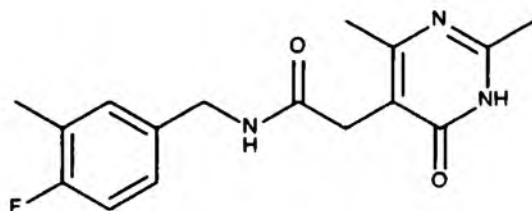
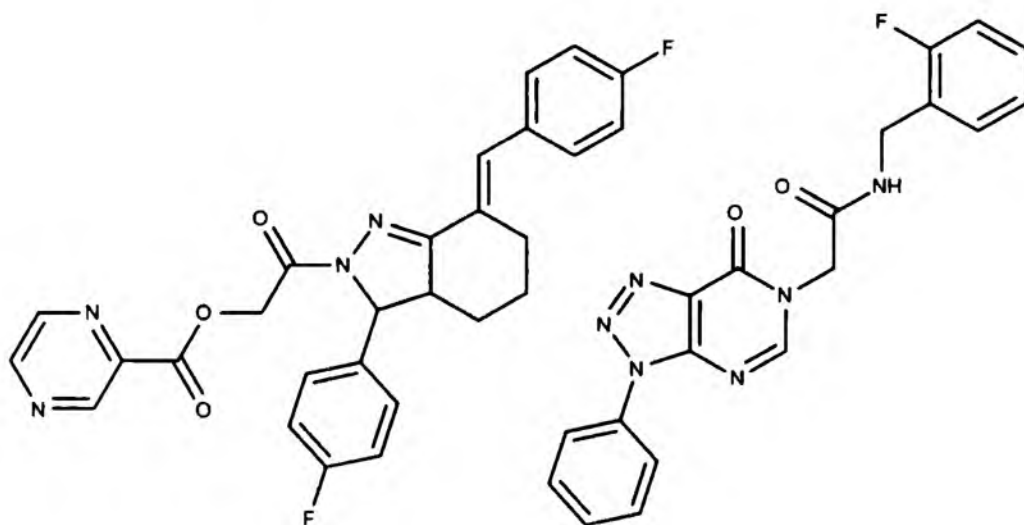
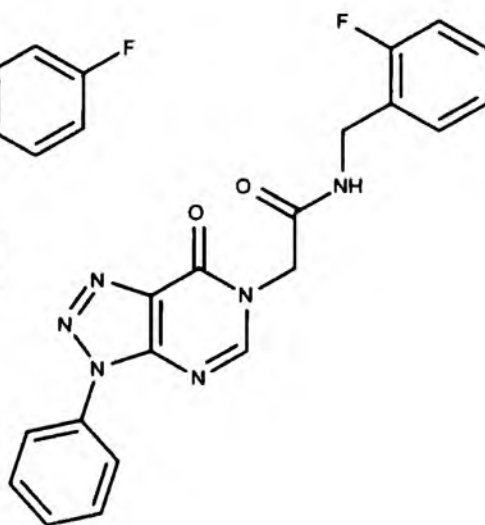
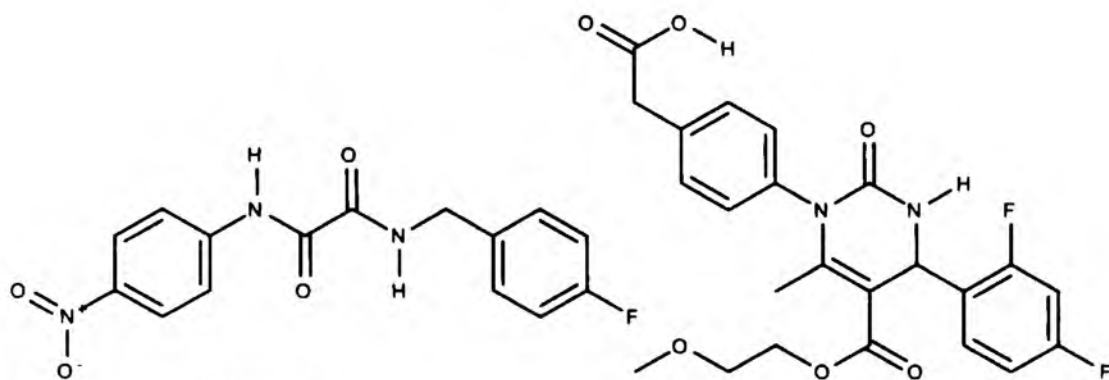
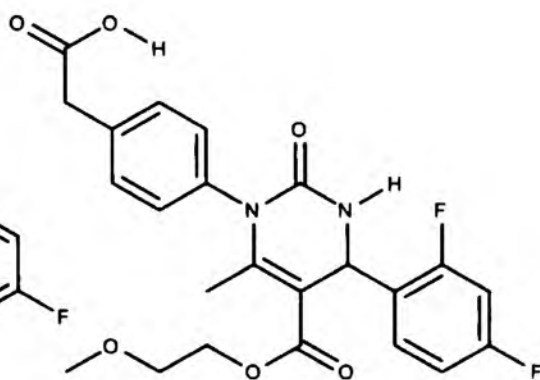
Rysunek 4.47 Schemat opracowanego protokołu VS zastosowanego do przeszukiwania komercyjnej bazy związków chemicznych w celu identyfikacji nowych inhibitorów IN wirusa HIV-1.

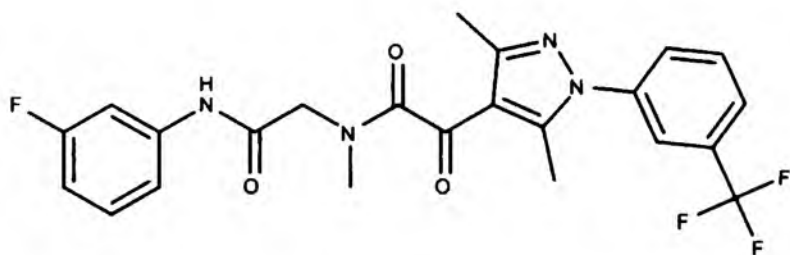
Metoda IBk wygenerowała 10665 hitów, natomiast metoda SMO 3 hity. Łączna liczba związków wyodrębnionych na drodze skriningu wyniosła więc 10668. Jednak jeden ze związków (indeks w bazie Enamine: T5553419) został wskazany zarówno przez metodę SMO, jak i IBk oraz związki o indeksie w bazie Enamine: T6815564 i T6861283 okazały się duplikatami. Dwie nadmiarowe struktury usunięto. Ostatecznie metodą VS bazy Enamine wskazano 10666 związków o potencjalnej aktywności względem enzymu IN wirusa HIV-1.

W celu zawężenia liczby związków, które zostaną zakupione i następnie skierowane do eksperymentalnej weryfikacji wyników poprzez wykonanie testów biologicznych *in vitro* w kierunku potencjalnej aktywności względem enzymu IN wirusa HIV-1, wykorzystano informacje na temat fragmentów uprzywilejowanych. W tym celu wykorzystano skonstruowany moduł PS. Opis modułu znajduje się w Rozdziale 4.4.7.1.2. Dla 10666 związków pochodzących z bazy Enamine i wskazanych

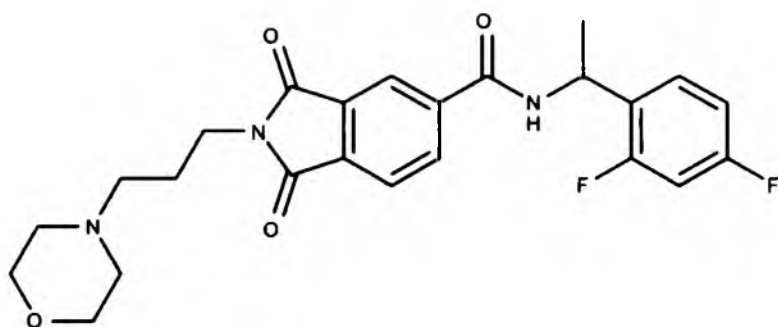
w pierwszym etapie procedury VS przez moduł ML, obliczono powtórnie deskryptor daktyloskopowy KlekFP i następnie przeprowadzono analizę zliczeń fragmentów uprzywilejowanych występujących jako podstruktury w poszczególnych cząsteczkach rozpatrywanego zbioru (zgodnie z procedurą modułu PS opisaną w Rozdziale 4.4.7.1.2). Ze zbioru odrzucono związki o niskiej wartości punktowej w skali PS. Średnia wartość punktowa wynosiła 23,53. Zbiór wyjściowy pomniejszono o związki, które uzyskały wartość punktową mniejszą niż 23. Po tej selekcji zostało 5290 struktur.

W kolejnym etapie filtrowania dla wszystkich związków wygenerowano deskryptor daktyloskopowy Molprint2D, a następnie wykonano klasteryzację hierarchiczną w oparciu o podobieństwo chemiczne z wykorzystaniem metryki Tanimoto (Tc) oraz metody łączenia klastrów Average (aplikacja Canvas pakietu Schrödinger). Otrzymano 385 klastrów. Z każdego klastra wybrano do dalszej analizy jedynie ten związek, który otrzymał największą liczbę punktów. W przypadku większej niż 1 liczby związków o maksymalnej wartości rozpatrywanego parametru – uwzględniano wszystkie struktury. Po tym etapie zostało 978 związków. Związki te podzielono na 10 grup, tak aby w każdej wyodrębnionej grupie znalazły się związki pochodzące z sąsiednich klastrów (39 klastrów na 9 kolejnych grup oraz 34 klastry w ostatniej dziesiątej grupie). Wykonano następnie inspekcję wizualną struktur i kierując się różnorodnością chemiczną wybrano z każdej grupy po jednej strukturze. W ten sposób wybrano dziesięć związków spośród całej puli wskazanej przez klasyfikator IBk. Związki te (wraz z trzema strukturami wskazanymi przez klasyfikator SMO) zakupiono w celu testowania ich potencjalnej aktywności względem wirusa HIV. Rysunek 4.48 przedstawia struktury trzynastu wyłonionych związków.

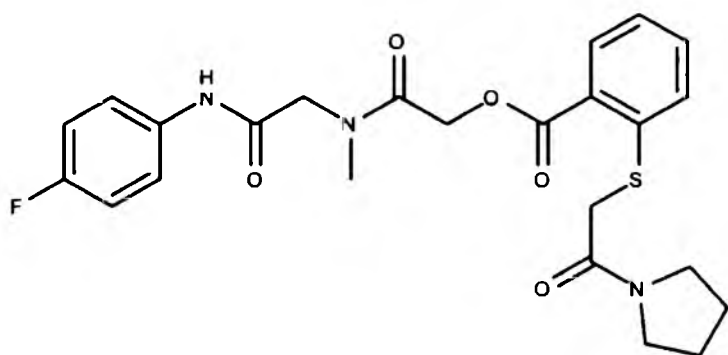
**T553419; SMO | IBk****T6412586; SMO****T6412424; SMO****T0510-0122; IBk****T6802914; IBk****T6420392; IBk****T6492823; IBk**



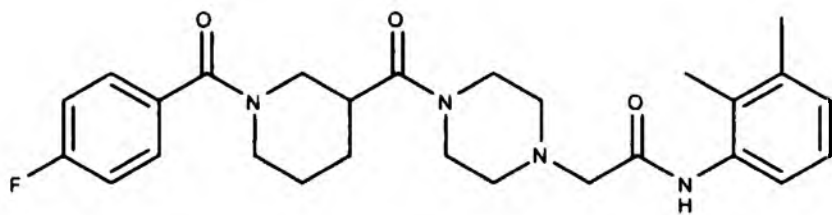
T6070186; IBk



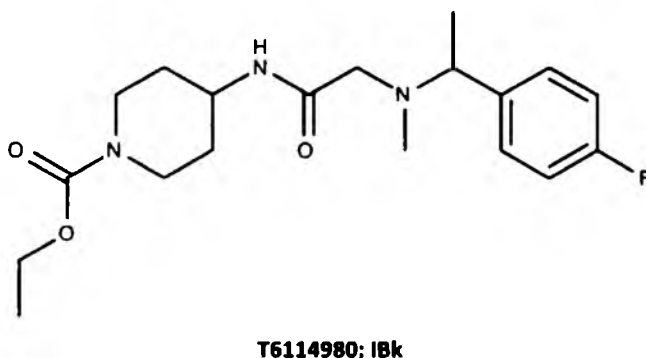
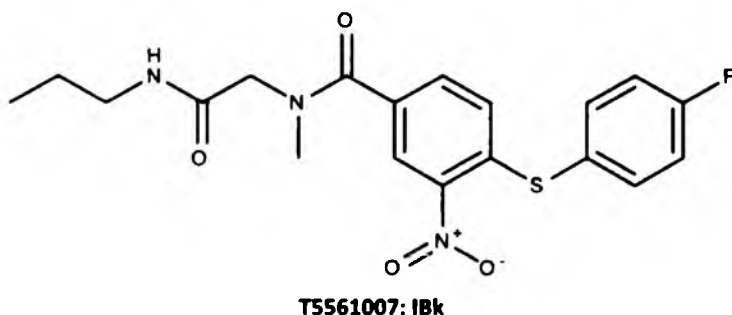
T6002810; IBk



T5537982; IBk



T5894322; IBk



Rysunek 4.48 Struktury związków wyłonionych na drodze opracowanego protokołu VS.

Wyniki uzyskane metodą wirtualnego skriningu należy weryfikować pod względem różnorodności strukturalnej, ponieważ odkrywanie nowych leków wymaga eksploracji różnych obszarów przestrzeni chemicznej. Konieczna więc staje się konstrukcja molekuł wypełniających możliwie równomiernie takie obszary. Z wyników uzyskanych na drodze VS odrzucane są związki znane i opatentowane lub stanowiące proste analogi strukturalne znanych już substancji aktywnych.

Wykonano obliczenia związane z podobieństwem strukturalnym cząsteczek wyodrębnionych na drodze VS, aby wykluczyć ryzyko zamówienia związków, których struktury mogłyby stanowić analogi znanych już inhibitorów IN wirusa HIV-1. Podobieństwo badano w obrębie trzech grup:

1. wzajemnie pomiędzy trzynastoma związkami wskazanymi na drodze VS;
2. pomiędzy trzynastoma związkami wskazanymi na drodze VS, a związkami aktywnymi tworzącymi zbiór act_A (Rozdział 4.4.2, Rysunek 4.35);
3. oraz pomiędzy trzynastoma związkami wskazanymi na drodze VS, a wszystkimi substancjami, które posiadają adnotacje o wykonaniu testów biologicznych w kierunku aktywności względem wirusa HIV (Rysunek 4.35 w Rozdziale 4.4.2).

Do obliczeń użyto deskryptor daktyloskopowy KlekFP oraz Daylight. Jako miarę podobieństwa stosowano metrykę Tanimoto (T_c). Parametr T_c przyjmuje wartości od 0 do 1. Zakłada się, iż dwie struktury są do siebie podobne, jeżeli wartość parametru T_c jest wyższa niż 0,85. Wyniki wskazały, iż związki między sobą są mocno zróżnicowane strukturalnie (średnia wartość parametru T_c wynosi 0,544 dla deskryptora Daylight i 0,450 dla KlekFP). Związki są odmienne strukturalnie od znanych substancji aktywnych (średnia wartość parametru T_c wynosi 0,486 dla deskryptora Daylight i 0,481 dla KlekFP). Niewiele bardziej przypominają związki przebadane w kierunku aktywności względem wirusa HIV (średnia wartość parametru T_c wynosi 0,525 dla deskryptora Daylight i 0,485 dla KlekFP).

4.4.9 TESTY BIOLOGICZNE *IN VITRO* ZWIĄZKÓW WYŁONIONYCH NA DRODZE WIRTUALNEGO SKRININGU

Związki chemiczne wyłonię na drodze opracowanego protokołu VS, zakupiono i skierowano do wykonania testów biologicznych *in vitro* w celu weryfikacji uzyskanych drogą obliczeń wyników. Badania biologiczne przeprowadzono w Narodowym Instytucie Leków w Warszawie⁹. Wykonano oznaczenie cytotoksyczności i aktywności względem wirusa HIV preparatów w hodowlach komórek ustalonej linii CEM-T4 inokulowanej laboratoryjnym izolatem wirusa HIV-1. W części eksperymentalnej niniejszej rozprawy – w Rozdziale 6.3 – zamieszczono opis procedury wykonanego oznaczenia.

Dwa preparaty (T6802914, T5561007) wykazały istotną aktywność antywirusową, jednak towarzysząca im cytotoksyczność powoduje konieczność optymalizacji ich struktur chemicznych w celu poprawienia profilu farmakokinetycznego. Tabela 4.11 przedstawia wyniki oznaczeń aktywności biologicznej trzynastu przebadanych preparatów. Dwa (wymienione powyżej) związki chemiczne mogą stanowić struktury wiodące w badaniach nad poszukiwaniem nowych leków aktywnych względem wirusa

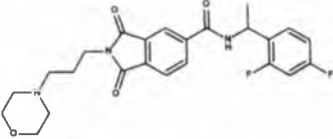
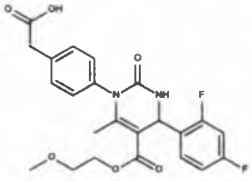
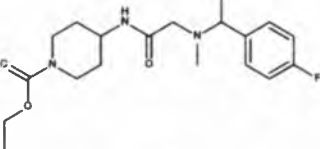
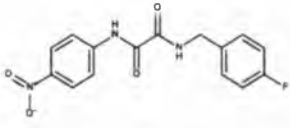
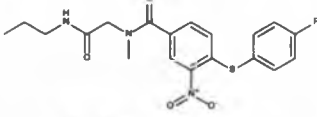
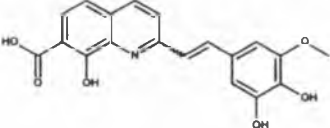
⁹ Badania biologiczne przeprowadzono w Samodzielnej Pracowni Wirusologii Narodowego Instytutu Leków w Warszawie pod kierownictwem Pana dr Andrzeja Piasek.

HIV. Wyniki oznaczeń biologicznych potwierdzają efektywność skonstruowanego protokołu VS.

Tabela 4.11 Wyniki oznaczeń aktywności biologicznej oraz cytotoksyczności.

Indeks związku	Struktura cząsteczki	IC ₅₀ [μM]	CC ₅₀ [μM]	TI ^a (CC ₅₀ /IC ₅₀)
T5894322		65	> 65	> 1
T5537982		65	> 68,3	> 1,05
T5553419		105	> 119	> 1,13
T0510-0122		ND ^b	< 1,4	ND
T6070186		22	15	0,68
T6802914		10	15	1,50
T6412424		102	> 110	> 1,08
T6412586		112	> 115,3	> 1,03

4. Badania własne

T6002810		> 72,5	> 72,7	~ 1,00
T6492823		> 72,3	> 72,3	~ 1,00
T6114980		85	> 91,3	> 1,07
T6420392		32	> 35	> 1,09
T5561007		25	70	2,80
FZ-41 (związek referencyjny)		2	6	3,00

^a Indeks terapeutyczny
^b Preparat T 0510-0122 wykazał cytotoksyczność już w najniższym badanym stężeniu (1,4uM), powodując całkowite zabicie komórek w hodowli. W tej sytuacji nie oznaczano jego aktywności antywirusowej.

4.5 PODSUMOWANIE

Zakres przeprowadzonych i opisanych w niniejszej rozprawie badań obejmuje poszukiwanie związków biologicznie aktywnych z wykorzystaniem metod obliczeniowych ze szczególnym uwzględnieniem eksploracji baz danych molekularno-strukturalnych. W poniższych punktach omówiono uzyskane wyniki w odniesieniu do celu badań sformułowanym w Rozdziale 2 – „Idea i cel projektu badawczego”.

1. *Opracowanie nowych metod eksploracji chemicznych baz danych w celu poszukiwania wiedzy koniecznej dla znajdowania fragmentów molekularnych przydatnych do projektowania leków.*

Przeprowadzono studium lekotypii fragmentu mono- i diazanaftalenowego (Rozdział 4.2), na podstawie którego zaproponowano i zdefiniowano parametry pozwalające na oszacowanie lekopodobieństwa, bioaktywności oraz dostępności syntetycznej związków badanej podprzestrzeni chemicznej. Parametry wyliczane są w oparciu o wyniki uzyskiwane metodami eksploracji baz danych molekularno-strukturalnych. Wartości wyliczanych w ten sposób parametrów mogą stanowić praktyczne wskazówki w procesie projektowania związków biologicznie aktywnych. Badanie lekotypii fragmentów azanaftalenowych opisano w załączonej publikacji:

- **Załącznik 2a**

Polanski J., Kurczyk A., Bak A., Musiol R.; Privileged structures – dream or reality: preferential organization of azanaphthalene scaffold., *Curr. Med. Chem.*, 2012, 19(13), 1921-1945.

2. *Wykorzystanie metod eksploracji chemicznych baz danych w celu przeprowadzenia polifarmakologicznej analizy leków aktywnych względem wirusa HIV w odniesieniu do koncepcji lekotypii oraz uprzywilejowania wybranych motywów strukturalnych.*

Zbadano topologię lekotypii fragmentów molekularnych wydzielonych w strukturach cząsteczek leków stosowanych w terapii przeciw HIV (Rozdział 4.3). Zaproponowano algorytm procedury fragmentacji dekrementacyjnej, którego kluczowym aspektem jest wykorzystanie metod eksploracji chemicznych baz danych w celu zbadania stopnia rozpowszechnienia danego fragmentu w odpowiednio zdefiniowanej przestrzeni chemicznej. Tego typu badanie pozwala na przeprowadzenie analizy topologicznej badanej grupy związków w szerokim aspekcie problemów polifarmakologii leków. Analiza ścieżek fragmentacji dekrementacyjnej pozwala na wskazanie w strukturze badanej cząsteczki takich obszarów, dla których wyodrębniane elementy podstruktury cechuje większa lub mniejsza swoistość. Zidentyfikowane w ten sposób fragmenty mogą stanowić punkt wyjściowy w projektowaniu nowych związków biologicznie aktywnych.

Procedury dekrementacyjnego podziału struktury chemicznej innego typu (ang. fragmental topology-activity landscapes, FRAGTAL) oraz ich wykorzystanie we fragonomice leków przeciw HIV opisano w załączonej publikacji:

- **Załącznik 2b**

Bak A., Magdziarz T., Kurczyk A., Serafin K., Polanski J.; Probing a Chemical Space for Fragmental Topology-Activity Landscapes (FRAGTAL): Application for Diketo Acid and Catechol HIV Integrase Inhibitor Offspring Fragments., *Comb. Chem. High Throughput Screen.*, 2013, 16(4), 274-287.

3. Poszukiwanie nowych związków aktywnych względem wirusa HIV.

Skonstruowano protokół wirtualnego skriningu (VS), który wykorzystano do przeszukiwania baz komercyjnie dostępnych związków w celu poszukiwania nowych inhibitorów enzymu integrazy (IN) wirusa HIV. Potencjalne inhibitory IN wyłonione na drodze VS, zakupiono i skierowano do wykonania testów biologicznych *in vitro* w celu weryfikacji prognozowanej bioaktywności tych struktur. Dwa preparaty wykazały istotną aktywność antywirusową, jednak towarzysząca im cytotoksyczność powoduje konieczność optymalizacji ich struktur chemicznych w celu poprawienia profilu

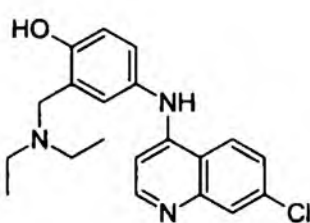
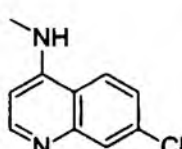
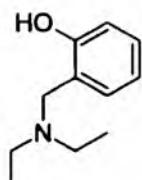
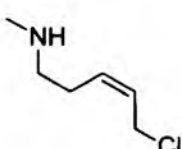
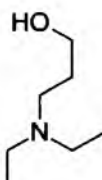
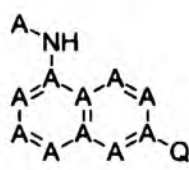
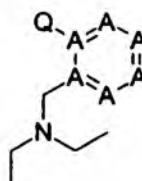
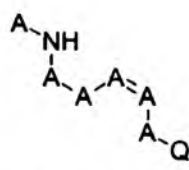
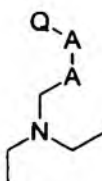
farmakokinetycznego. Związki te mogą więc stanowić struktury wiodące w dalszych badaniach nad poszukiwaniem nowych leków aktywnych względem wirusa HIV.

5 METODY

W poniższych rozdziałach opisano wykorzystane w pracy metody i stosowane algorytmy (Rozdziały 5.1 - 5.3).

5.1 PRZESZUKIWANIE BAZ DANYCH W OPARCIU O PODSTRUKTURY

Przestrzeń chemiczną badanych związków eksplorowano z wykorzystaniem metod przeszukiwania baz danych w oparciu o podstruktury. Zapytanie do bazy stanowił odpowiednio zdefiniowany fragment molekularny. Rysunek 5.1 przedstawia reguły definiowania zapytań strukturalnych w metodach eksploracji chemicznych baz danych w oparciu o podstruktury.

molecule			
exact substructures	 	exact linear/branched substructures	 
generic substructures	  A – any atom Q – any heteroatom	generic linear/branched substructures	  A – any atom Q – any heteroatom

Rysunek 5.1 Definiowanie podstruktury. Schemat zaczerpnięty z publikacji [90].

5.2 PROCEDURA GENEROWANIA ZBIORU ZWIĄZKÓW NIEAKTYWNYCH DUD

Procedura DUD [157] pozwala na konstrukcję zbioru związków nieaktywnych, dla którego wartości deskryptorów molekularnych mieszczą się w tych samych granicach jak dla badanej populacji związków aktywnych, ale z zachowaniem zróżnicowania strukturalnego pomiędzy związkami z obydwu zbiorów (Rysunek 5.2).

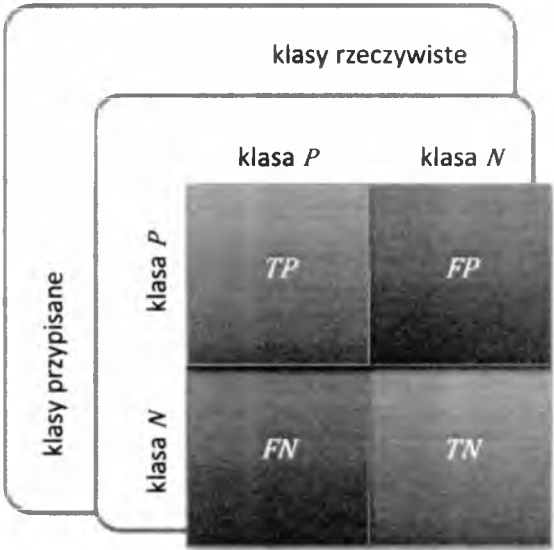


Rysunek 5.2 Schemat poszukiwania zbioru związków nieaktywnych w oparciu o algorytm DUD [157].

Zbiór związków nieaktywnych generowany jest z podzbioru „druglike” bazy danych ZINC [42,43]. W pierwszym kroku wybierane są tylko te związki, których wartość współczynnika podobieństwa Tanimoto (ang. Tanimoto coefficient, T_c) mierzona względem badanego zbioru ligandów nie przekracza 0,9. W kolejnym etapie obliczane są wartości deskryptorów molekularnych (masa molowa, logP, liczba donorów i akceptorów wiązania wodorowego oraz liczba wiązań rotacyjnych) dla wszystkich związków obydwu zbiorów. Następnie związki należące do zbioru DUD są sortowane według podobieństwa właściwości fizykochemicznych do któregośkolwiek z badanego zbioru związków aktywnych. Zgodnie z opisaną procedurą [157], algorytm przewiduje wybór trzydziestu sześciu związków DUD na każdy związek aktywny, a więc liczebność zbioru DUD względem liczebności zbioru związków aktywnych powinna pozostawać w stosunku 36:1.

5.3 METRYKI OCENY EFEKTYWNOŚCI KLASYFIKATORÓW BINARNYCH

Efektywność klasyfikacji oceniana jest za pomocą różnych metryk [174]. Na drodze klasyfikacji binarnej dla każdej instancji zbioru testowego możliwy jest jeden z czterech wyników: dla przykładów pozytywnych uzyskuje się dobre (ang. true positives, TP) lub złe (ang. false positives, FP) przypisanie do klasy i podobnie dla instancji klasy negatywnej uzyskuje się poprawne (ang. true negatives, TN) lub błędne (ang. false negatives, FN) przypisanie. W analizie wyników klasyfikacji pomocne jest utworzenie macierzy porównującej stan faktyczny ze wskazaniem modelu (ang. confusion matrix, contingency table, error matrix) (Rysunek 5.3). Jest to macierz kwadratowa drugiego stopnia, której elementy wyrażają sumaryczne zestawienie wyników uzyskanych w trakcie procesu testowania modelu klasyfikatora. Macierz klasyfikacji stanowi punkt wyjścia dla wyliczenia różnorodnych metryk służących ewaluacji skuteczności modelu. Równania 5.1 - 5.5 definiują pięć najczęściej stosowanych miar szacujących wartość predykcyjną reguły decyzyjnej.



Rysunek 5.3 Macierz klasyfikacji.

$$\text{czułość} = \frac{TP}{TP + FN} = \frac{TP}{P}$$

Równanie 5.1

$$\text{precyzja} = \frac{TP}{TP + FP} \quad \text{Równanie 5.2}$$

$$\text{specyficzność} = \frac{TN}{TN + FP} = \frac{TN}{N} \quad \text{Równanie 5.3}$$

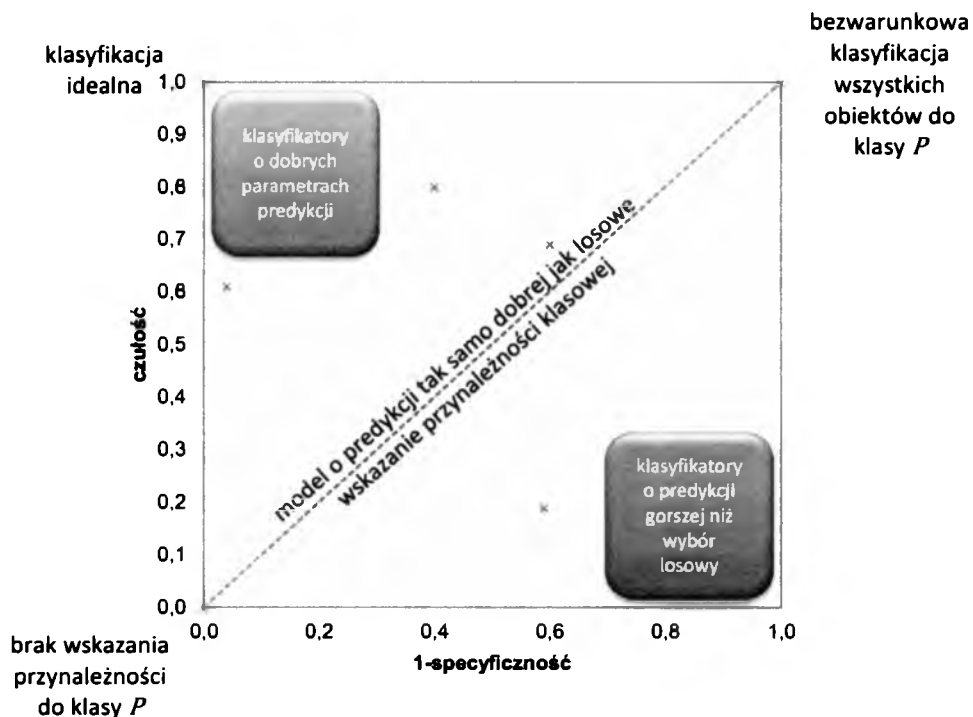
$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad \text{Równanie 5.4}$$

$$\text{metryka F} = \frac{2}{1/\text{precyzja} + 1/\text{czułość}} \quad \text{Równanie 5.5}$$

Czułość (ang. recall, sensitivity) podaje stosunek poprawnie wskazanych obiektów klasy P względem całkowitej liczebności tej klasy w zbiorze testowym. Precyzja (ang. precision) odnosi liczbę poprawnie wskazanych instancji klasy P do wszystkich obiektów sklasyfikowanych przez model do kategorii P . Specyficzność (ang. specificity) wylicza stosunek poprawnie wskazanych obiektów klasy N względem całkowitej liczebności tej klasy. Parametr korelacyjny Matthews (MCC) to uniwersalna metryka stosowana w ocenie klasyfikatorów binarnych, która rozpatruje zależności pomiędzy TP , FP , TN i FN , a jej niewątpliwą zaletą jest stosowalność nawet w przypadku zbiorów o dużej dysproporcji klas. MCC przyjmuje wartości z przedziału od -1 do 1, gdzie 1 oznacza bezbłędną predykcję, 0 predykcję nie lepszą niż losowa, a -1 oznacza klasyfikację błędną. Czułość, precyzja i specyficzność przyjmują wartości od 0 do 1.

Krzywe ROC (ang. receiver operating characteristics) są pomocnym narzędziem do oceny i wizualizacji efektywności klasyfikatora [174]. Wykres ROC ilustruje związek między czułością a specyficznością dla danego modelu. Rysunek 5.4 przedstawia schemat konstrukcji i sposób interpretacji wykresu ROC. Punkt o współrzędnych (0,1) to punkt reprezentujący klasyfikator, który poprawnie wskazał przynależność wszystkich obiektów klasy wyróżnionej (klasy P) i równocześnie nie popełnił błędnego przypisania obiektów klasy N do klasy P . W przypadku klasyfikatora binarnego uzyskuje się pojedynczy punkt na wykresie ROC. Dla każdego wyniku modelu można utworzyć regułę decyzyjną, wybierając odpowiedni punkt odcięcia. Dla każdego z możliwych punktów odcięcia obliczana jest czułość i specyficzność. Otrzymane wyniki naniesione na wykres ROC tworzą krzywą. Pole powierzchni pod krzywą ROC (ang. area under ROC

curve, AUC) pozwala porównać różne klasyfikatory. Idealny klasyfikator posiada wartość wskaźnika AUC = 1.



Rysunek 5.4 Wykres ROC. Punkty odzwierciedlają przykładowe wyniki uzyskane dla siedmiu różnych klasyfikatorów binarnych. Opisy wskazują sposób interpretacji poszczególnych obszarów wykresu.

6 CZĘŚĆ EKSPERYMENTALNA

W części eksperymentalnej przedstawiono szczegółowe wyniki i wszelkie dane liczbowe uzyskane w przeprowadzonych eksperymentach. (Rozdział 6.1 oraz Rozdział 6.2). W Rozdziale 6.3 umieszczono opis procedury oznaczenia cytotoksyczności i aktywności względem wirusa HIV dostarczony przez wykonawcę testów biologicznych.¹⁰

6.1 STUDIUM LEKOTYPII FRAGMENTU MONO- I DIAZANAFTALENOWEGO – WYNIKI EKSPLOKACJI CHEMICZNYCH BAZ DANYCH

Przeprowadzono studium lekotypii fragmentów molekularnych dla ugrupowania mono- i diazanaftalenowego (Rozdział 4.2). Przestrzeń chemiczną azanaftalenów zdefiniowano jako pulę związków organicznych zawierających fragment mono- lub diazanaftalenowy jako podstrukturę. W ten sposób w obrębie badanej przestrzeni wskazano dwanaście zbiorów związków chemicznych posiadających jeden z fragmentów molekularnych **f1 - f12** jako element podstruktury (struktury fragmentów przedstawia Rysunek 4.1 w Rozdziale 4.2.1). Analizy porównawcze wyodrębnionych zbiorów cząsteczek przeprowadzono na podstawie informacji uzyskanych metodą eksploracji chemicznych baz danych. Tabele poniżej przedstawiają dane liczbowe, stanowiące wyniki przeszukiwań chemicznych baz danych w oparciu o podstrukturę zdefiniowaną jednym z dwunastu fragmentów o indeksie od **f1** do **f12**. Dane liczbowe posłużyły do wykonania analiz opisanych w Rozdziale 4.2.1 (Tabela 6.2), Rozdziale 4.2.2 i Rozdziale 4.2.3 (Tabela 6.3) oraz w Rozdziale 4.2.4 (Tabela 6.4). Tabela 6.1 przedstawia notację SMILES badanych fragmentów. Notację liniowego zapisu struktur chemicznych SMILES wykorzystano do formułowania zapytań strukturalnych do baz.

¹⁰ Badania biologiczne przeprowadzono w Samodzielnej Pracowni Wirusologii Narodowego Instytutu Leków w Warszawie pod kierownictwem Pana dr Andrzeja Piasek.

Tabela 6.1 Notacja SMILES fragmentów azanaftalenowych.

Fragment molekularny	SMILES	Fragment molekularny	SMILES
f1	C1=CC=CC2=C1C=CC=N2	f7	C1=NC2=C(C=C1)N=CC=C2
f2	C1=CC=CC2=C1C=CN=C2	f8	C1=CC2=C(C=N1)C=CC=N2
f3	C1=CC=CC2=C1C=CN=N2	f9	C1=CN=CC2=C1C=CC=N2
f4	C1=CC=CC2=C1C=NC=N2	f10	C1=CC=NC2=C1C=CC=N2
f5	C1=CC=CC2=C1N=CC=N2	f11	C1=NC=CC2=C1C=CN=C2
f6	C1=CC=CC2=C1C=NN=C2	f12	C1=CN=CC2=C1C=CN=C2

Tabela 6.2 Liczba rekordów uzyskanych w wyniku eksploracji bazy danych DrugBank i Beilstein w oparciu o podstrukturę zdefiniowaną jednym z dwunastu fragmentów molekularnych o indeksie od f1 do f12. W ostatniej kolumnie podano wyliczone wartości parametru L.

Fragment molekularny	DrugBank ^{a, b}	Beilstein ^b	param_L ^c
f1	32	185470	0,000173
f2	1	32137	0,000031
f3	1	3909	0,000256
f4	9	18331	0,000491
f5	4	32399	0,000123
f6	2	4393	0,000455
f7	0	1176	0,000000
f8	0	1640	0,000000
f9	0	695	0,000000
f10	3	4694	0,000639
f11	0	198	0,000000
f12	0	827	0,000000

^a Liczba rekordów bazy DrugBank została ograniczona do leków posiadających status zatwierdzenia przynajmniej w jednym kraju.

^b Dane na dzień 4 kwietnia 2010.

^c Parametr L zdefiniowano w Rozdziale 4.2.1 poprzez Równanie 4.1.

Tabela 6.3 Liczba rekordów uzyskanych w wyniku eksploracji bazy danych PubChem w oparciu o podstrukturę zdefiniowaną jednym z dwunastu fragmentów molekularnych o indeksie od f1 do f12. W piątej i szóstej kolumnie podano wyliczone wartości parametrów I oraz B.

Fragment molekularny	Liczba wszystkich rekordów, zbiór <i>W</i>	Liczba związków testowanych, zbiór <i>T</i>	Liczba związków aktywnych, zbiór <i>A</i>	param_I ^a	param_B ^b	Data
f1	469868	32650	9510	0,07	0,29	27/05/2010
f2	42431	3192	719	0,08	0,23	
f3	2858	158	41	0,06	0,26	
f4	85948	6872	2621	0,08	0,38	
f5	71772	4582	1738	0,06	0,38	
f6	11758	1036	275	0,09	0,27	
f7	2640	175	132	0,07	0,75	
f8	3488	222	61	0,06	0,27	
f9	1253	82	4	0,07	0,05	
f10	4584	206	46	0,04	0,22	
f11	346	3	0	0,01	0,00	
f12	1421	131	65	0,09	0,50	
f1	507148	43661	10582	0,09	0,24	06/11/2010
f2	45899	4446	850	0,10	0,19	
f3	3207	214	41	0,07	0,19	
f4	102946	10180	3051	0,10	0,30	
f5	82123	6792	2046	0,08	0,30	
f6	14640	1599	391	0,11	0,24	
f7	2833	245	134	0,09	0,55	
f8	3857	292	69	0,08	0,24	
f9	1330	123	5	0,09	0,04	
f10	5085	343	54	0,07	0,16	
f11	403	3	0	0,01	0,00	
f12	1471	243	70	0,17	0,29	
f1	521221	44998	11200	0,09	0,25	06/05/2011
f2	47793	4594	921	0,10	0,20	
f3	3239	221	42	0,07	0,19	
f4	105682	10750	3249	0,10	0,30	
f5	85975	7095	2154	0,08	0,30	
f6	14948	1700	449	0,11	0,26	
f7	2864	245	134	0,09	0,55	
f8	3923	338	71	0,09	0,21	
f9	1374	134	6	0,10	0,04	
f10	5296	428	56	0,08	0,13	
f11	452	33	2	0,07	0,06	
f12	1479	244	70	0,16	0,29	

^a Parametr I zdefiniowano w Rozdziale 4.2.2 poprzez Równanie 4.3.

^b Parametr B zdefiniowano w Rozdziale 4.2.2 poprzez Równanie 4.2.

Tabela 6.4 Liczba rekordów uzyskanych w wyniku eksploracji bazy danych PubChem w oparciu o podstrukturę zdefiniowaną jednym z dwunastu fragmentów molekularnych o indeksie od f1 do f12, podstawionym jednym bądź dwoma heteroatomami. W szóstej i siódmej kolumnie podano wyliczone wartości parametrów I oraz B.

Fragment molekularny	SMILES	Liczba wszystkich rekordów ^a , zbiór W	Liczba związków testowanych ^a , zbiór T	Liczba związków aktywnych ^a , zbiór A	param_I ^b	param_B ^c
fragment molekularny podstawiony jednym heteroatomem						
f12-Q-c	[Q]c1cncc2cnccc12	44	13	0	0,30	0,00
f3-Q-e	[Q]c1cccc2ccnnc12	214	31	0	0,14	0,00
f8-Q-a	[Q]c1cnc2ccncc2c1	174	7	0	0,04	0,00
f9-Q-c	[Q]c1ccc2ccncc2n1	111	3	0	0,03	0,00
f9-Q-d	[Q]c1cncc2ncccc12	50	2	0	0,04	0,00
f9-Q-f	[Q]c1ncc2ncccc2c1	282	61	1	0,22	0,02
f9-Q-e	[Q]c1ccnc2cnccc12	287	59	1	0,21	0,02
f2-Q-c	[Q]c1cncc2ccccc12	2868	199	10	0,07	0,05
f3-Q-b	[Q]c1ccc2ccnnc2c1	827	37	2	0,04	0,05
f8-Q-b	[Q]c1nccc2ncccc12	1858	109	6	0,06	0,06
f10-Q-c	[Q]c1ccnc2ncccc12	1309	175	11	0,13	0,06
f8-Q-c	[Q]c1cncc2ccnnc12	684	47	3	0,07	0,06
f8-Q-d	[Q]c1ccnc2ccncc12	185	13	1	0,07	0,08
f4-Q-f	[Q]c1cccc2ncncc12	2187	400	46	0,18	0,12
f3-Q-c	[Q]c1cnnc2ccccc12	1548	95	12	0,06	0,13
f6-Q-a	[Q]c1ccc2cnnc2c1	1066	66	9	0,06	0,14
f2-Q-e	[Q]c1cccc2ccncc12	4828	425	59	0,09	0,14
f10-Q-b	[Q]c1ccc2ccncc2n1	2363	177	26	0,07	0,15
f2-Q-b	[Q]c1ncc2ccccc2c1	2371	159	24	0,07	0,15
f3-Q-d	[Q]c1ccc2nnccc2c1	960	63	10	0,07	0,16
f1-Q-f	[Q]c1ccc2ccncc2c1	60731	7197	1181	0,12	0,16
f2-Q-a	[Q]c1cccc2cnccc12	10130	630	112	0,06	0,18
f10-Q-a	[Q]c1cnc2ncccc2c1	598	28	5	0,05	0,18
f1-Q-a	[Q]c1cnc2ccccc2c1	26239	1921	345	0,07	0,18
f2-Q-f	[Q]c1ccc2ccncc2c1	9059	1237	230	0,14	0,19
f6-Q-b	[Q]c1cccc2cnccc12	318	21	4	0,07	0,19
f4-Q-b	[Q]c1ccc2cncnc2c1	25534	2593	532	0,10	0,21
f1-Q-g	[Q]c1ccnc2ccccc12	91562	13666	2821	0,15	0,21
f5-Q-c	[Q]c1cccc2nccnc12	2301	304	63	0,13	0,21
f4-Q-a	[Q]c1cccc2cncnc12	2481	157	33	0,06	0,21
f5-Q-b	[Q]c1ccc2nccnc2c1	11143	1277	271	0,11	0,21
f4-Q-e	[Q]c1ccc2ncncc2c1	30856	3963	849	0,13	0,21
f2-Q-d	[Q]c1nccc2ccccc12	11288	976	216	0,09	0,22
f9-Q-b	[Q]c1nccc2ccncc12	535	13	3	0,02	0,23
f1-Q-e	[Q]c1ccc2ncccc2c1	91596	9634	2281	0,11	0,24
f2-Q-g	[Q]c1ccc2cnccc2c1	8300	1203	289	0,14	0,24
f1-Q-b	[Q]c1cccc2ncccc12	36701	2814	685	0,08	0,24
f4-Q-d	[Q]c1ncc2ccccc2n1	32246	3732	915	0,12	0,25
f3-Q-a	[Q]c1nnc2ccccc2c1	178	4	1	0,02	0,25
f6-Q-c	[Q]c1nncc2ccccc12	13557	1544	389	0,11	0,25

f1-Q-c	[Q]c1cccc2ccccc2n1	99389	8270	2133	0,08	0,26
f1-Q-d	[Q]c1cccc2ccnc12	78152	7097	2038	0,09	0,29
f4-Q-c	[Q]c1ncnc2ccccc12	86472	8896	2722	0,10	0,31
f12-Q-a	[Q]c1nccc2ccncc12	1084	149	49	0,14	0,33
f5-Q-a	[Q]c1cnc2ccccc2n1	46233	4190	1404	0,09	0,34
f3-Q-f	[Q]c1cccc2nnccc12	132	28	13	0,21	0,46
f7-Q-c	[Q]c1ccnc2ccnc12	1457	134	69	0,09	0,51
f12-Q-b	[Q]c1ncc2cnccc2c1	560	85	44	0,15	0,52
f8-Q-e	[Q]c1ccc2cnccc2n1	520	84	51	0,16	0,61
f7-Q-b	[Q]c1ccc2ncccc2n1	2110	182	117	0,09	0,64
f8-Q-f	[Q]c1ncc2ccnc2c1	270	70	53	0,26	0,76
f7-Q-a	[Q]c1cnc2ccnc2c1	849	52	45	0,06	0,87
f11-Q-a	[Q]c1cncc2ccncc12	4	0	0	0,00	-
f11-Q-b	[Q]c1ncc2ccncc2c1	6	0	0	0,00	-
f11-Q-c	[Q]c1nccc2cnccc12	292	0	0	0,00	-
f9-Q-a	[Q]c1cnc2cnccc2c1	48	0	0	0,00	-
fragment molekularny podstawiony dwoma heteroatomami						
f10-QQ-a	[Q]c1cc([Q])c2ccnc2n1	284	69	4	0,24	0,06
f10-QQ-b	[Q]c1ccc2c([Q])ccnc2n1	121	15	1	0,12	0,07
f10-QQ-c	[Q]c1cnc2nccc([Q])c2c1	203	14	0	0,07	0,00
f10-QQ-d	[Q]c1ccnc2nccc([Q])c12	20	1	1	0,05	1,00
f10-QQ-e	[Q]c1nc2ncccc2cc1[Q]	142	13	1	0,09	0,08
f10-QQ-f	[Q]c1cnc2nc([Q])ccc2c1	84	5	2	0,06	0,40
f10-QQ-g	[Q]c1cnc2ncc([Q])cc2c1	26	1	0	0,04	0,00
f10-QQ-h	[Q]c1ccc2ccc([Q])nc2n1	865	21	3	0,02	0,14
f10-QQ-i	[Q]c1cnc2ncccc2c1[Q]	95	0	0	0,00	-
f11-QQ-a	[Q]c1ncc2cncc([Q])c2c1	0	0	0	-	-
f11-QQ-b	[Q]c1nccc2c([Q])nccc12	6	0	0	0,00	-
f11-QQ-c	[Q]c1ncc([Q])c2cnccc12	0	0	0	-	-
f11-QQ-d	[Q]c1cc2cnccc2c([Q])n1	2	0	0	0,00	-
f11-QQ-e	[Q]c1ncc2ccncc2c1[Q]	0	0	0	-	-
f11-QQ-f	[Q]c1cncc2c([Q])cncc12	0	0	0	-	-
f11-QQ-g	[Q]c1ncc2c(cncc2[Q])c1	0	0	0	-	-
f11-QQ-h	[Q]c1nccc2cncc([Q])c12	0	0	0	-	-
f11-QQ-i	[Q]c1ncc2cc([Q])ncc2c1	0	0	0	-	-
f12-QQ-a	[Q]c1ncc2cncc([Q])c2c1	4	0	0	0,00	-
f12-QQ-b	[Q]c1nccc2c([Q])cncc12	1	0	0	0,00	-
f12-QQ-c	[Q]c1ncc([Q])c2ccncc12	3	0	0	0,00	-
f12-QQ-d	[Q]c1ncc2cnccc2c1[Q]	9	1	0	0,11	0,00
f12-QQ-e	[Q]c1cc2ccncc2c([Q])n1	535	84	44	0,16	0,52
f12-QQ-f	[Q]c1cncc2cncc([Q])c12	5	0	0	0,00	-
f12-QQ-g	[Q]c1ncc2c(ccnc2[Q])c1	529	76	44	0,14	0,58
f12-QQ-h	[Q]c1nccc2ccnc([Q])c12	523	76	44	0,15	0,58
f12-QQ-i	[Q]c1ncc2cncc([Q])cc2c1	513	76	44	0,15	0,58
f1-QQ-a	[Q]c1ccc2nc([Q])ccc2c1	17924	2331	568	0,13	0,24
f1-QQ-b	[Q]c1ccc2ncc([Q])cc2c1	7824	445	71	0,06	0,16
f1-QQ-c	[Q]c1ccc2nccc([Q])c2c1	23016	3579	829	0,16	0,23
f1-QQ-d	[Q]c1ccc2ncccc2c1[Q]	4913	443	57	0,09	0,13
f1-QQ-e	[Q]c1ccc2ccc([Q])nc2c1	15203	1260	290	0,08	0,23
f1-QQ-f	[Q]c1cnc2cc([Q])ccc2c1	2009	104	9	0,05	0,09

6. Część eksperymentalna

f1-QQ-g	[Q]c1ccc2c([Q])ccnc2c1	26006	4601	692	0,18	0,15
f1-QQ-h	[Q]c1cc([Q])c2cccnc2c1	5135	382	97	0,07	0,25
f1-QQ-i	[Q]c1cc2cccc2nc1[Q]	4955	544	26	0,11	0,05
f1-QQ-j	[Q]c1cc2ncccc2cc1[Q]	16099	1994	305	0,12	0,15
f1-QQ-k	[Q]c1ccc2cccc([Q])c2n1	7211	330	47	0,05	0,14
f1-QQ-l	[Q]c1cnc2c(cccc2[Q])c1	1195	78	32	0,07	0,41
f1-QQ-m	[Q]c1cccc2c([Q])ccnc12	8064	1087	169	0,13	0,16
f1-QQ-n	[Q]c1ccc([Q])c2cccnc12	16575	1456	409	0,09	0,28
f1-QQ-o	[Q]c1cc([Q])c2cccc2n1	8723	995	170	0,11	0,17
f1-QQ-p	[Q]c1cc2cccnc2c([Q])c1	6988	1132	190	0,16	0,17
f1-QQ-r	[Q]c1cnc2cccc2c1[Q]	9265	476	80	0,05	0,17
f1-QQ-s	[Q]c1ccc2cccnc2c1[Q]	5672	362	111	0,06	0,31
f1-QQ-t	[Q]c1ccc2c([Q])cccc2n1	2554	98	34	0,04	0,35
f1-QQ-u	[Q]c1cnc2cccc([Q])c2c1	482	20	3	0,04	0,15
f1-QQ-v	[Q]c1cccc2nccc([Q])c12	3354	613	129	0,18	0,21
f2-QQ-a	[Q]c1ccc2cnc([Q])cc2c1	507	54	4	0,11	0,07
f2-QQ-b	[Q]c1ccc2c(ccnc2[Q])c1	1731	91	18	0,05	0,20
f2-QQ-c	[Q]c1cc2ccncc2c([Q])c1	509	86	11	0,17	0,13
f2-QQ-d	[Q]c1cc2cnccc2cc1[Q]	4334	660	147	0,15	0,22
f2-QQ-e	[Q]c1ncc2cccc([Q])c2c1	132	3	0	0,02	0,00
f2-QQ-f	[Q]c1cccc2c([Q])nccc12	704	43	10	0,06	0,23
f2-QQ-g	[Q]c1ccc([Q])c2ccncc12	2189	34	13	0,02	0,38
f2-QQ-h	[Q]c1cc2cnccc2c([Q])c1	390	45	5	0,12	0,11
f2-QQ-i	[Q]c1cc2cccc2c([Q])n1	758	34	5	0,04	0,15
f2-QQ-j	[Q]c1ccc2cnccc2c1[Q]	467	49	4	0,10	0,08
f2-QQ-k	[Q]c1ncc2cccc2c1[Q]	206	5	0	0,02	0,00
f2-QQ-l	[Q]c1ncc([Q])c2cccc12	1157	87	0	0,08	0,00
f2-QQ-m	[Q]c1cncc2c([Q])cccc12	111	12	0	0,11	0,00
f2-QQ-n	[Q]c1ccc2c(cncc2[Q])c1	922	63	1	0,07	0,02
f2-QQ-o	[Q]c1ncc2c(cccc2[Q])c1	79	4	3	0,05	0,75
f2-QQ-p	[Q]c1ccc2cncc([Q])c2c1	285	13	1	0,05	0,08
f2-QQ-r	[Q]c1cccc2cnc([Q])c12	192	6	3	0,03	0,50
f2-QQ-s	[Q]c1cccc2ncc([Q])c12	219	5	0	0,02	0,00
f2-QQ-t	[Q]c1ccc2cc([Q])ncc2c1	480	55	4	0,11	0,07
f2-QQ-u	[Q]c1ccc2ccnc([Q])c2c1	1642	124	16	0,08	0,13
f2-QQ-v	[Q]c1ccc2ccncc2c1[Q]	1606	284	31	0,18	0,11
f3-QQ-a	[Q]c1nnc2cccc([Q])c2c1	7	0	0	0,00	-
f3-QQ-b	[Q]c1ccc([Q])c2ccnnc12	3	0	0	0,00	-
f3-QQ-c	[Q]c1cc([Q])c2ccnnc2c1	18	1	0	0,06	0,00
f3-QQ-d	[Q]c1ccc2nnccc2c1[Q]	21	8	5	0,38	0,63
f3-QQ-e	[Q]c1nnc2cccc2c1[Q]	41	1	0	0,02	0,00
f3-QQ-f	[Q]c1cccc2c([Q])cnnc12	103	17	0	0,17	0,00
f3-QQ-g	[Q]c1ccc2c([Q])cnnc2c1	519	32	2	0,06	0,06
f3-QQ-h	[Q]c1ccc2nncc([Q])c2c1	554	48	5	0,09	0,10
f3-QQ-i	[Q]c1nnc2c(cccc2[Q])c1	4	0	0	0,00	-
f3-QQ-j	[Q]c1cccc2nncc([Q])c12	15	1	0	0,07	0,00
f3-QQ-k	[Q]c1ccc2cc([Q])nnc2c1	29	0	0	0,00	-
f3-QQ-l	[Q]c1ccc2ccnnc2c1[Q]	11	0	0	0,00	-
f3-QQ-m	[Q]c1ccc2nncc([Q])cc2c1	24	0	0	0,00	-
f3-QQ-n	[Q]c1cc2ccnnc2c([Q])c1	28	1	0	0,04	0,00

f3-QQ-o	[Q]c1cc2nnccc2cc1[Q]	619	31	2	0,05	0,06
f4-QQ-a	[Q]c1ccc2cnc([Q])nc2c1	6939	881	217	0,13	0,25
f4-QQ-b	[Q]c1ccc2c(ncnc2[Q])c1	22830	2437	517	0,11	0,21
f4-QQ-c	[Q]c1cc([Q])c2cncnc2c1	582	63	5	0,11	0,08
f4-QQ-d	[Q]c1cc2ncncc2cc1[Q]	17694	2201	418	0,12	0,19
f4-QQ-e	[Q]c1ncc2cccc([Q])c2n1	979	46	7	0,05	0,15
f4-QQ-f	[Q]c1ncnc2c([Q])cccc12	1596	136	31	0,09	0,23
f4-QQ-g	[Q]c1ccc([Q])c2cncnc12	179	2	0	0,01	0,00
f4-QQ-h	[Q]c1cc2cncnc2c([Q])c1	1100	86	18	0,08	0,21
f4-QQ-i	[Q]c1nc2cccc2c([Q])n1	20439	2730	626	0,13	0,23
f4-QQ-j	[Q]c1ccc2cncnc2c1[Q]	621	69	8	0,11	0,12
f4-QQ-k	[Q]c1ncc2c([Q])cccc2n1	722	226	24	0,31	0,11
f4-QQ-l	[Q]c1cccc2ncnc([Q])c12	1777	391	40	0,22	0,10
f4-QQ-m	[Q]c1ccc2nc([Q])ncc2c1	9798	1640	424	0,17	0,26
f4-QQ-n	[Q]c1ccc2ncnc([Q])c2c1	24993	3455	736	0,14	0,21
f4-QQ-o	[Q]c1ccc2ncncc2c1[Q]	319	57	19	0,18	0,33
f5-QQ-a	[Q]c1cnc2c([Q])cccc2n1	467	77	12	0,16	0,16
f5-QQ-b	[Q]c1ccc([Q])c2nccnc12	395	14	2	0,04	0,14
f5-QQ-c	[Q]c1cc([Q])c2nccnc2c1	515	66	6	0,13	0,09
f5-QQ-d	[Q]c1ccc2nccnc2c1[Q]	701	55	15	0,08	0,27
f5-QQ-e	[Q]c1nc2cccc2nc1[Q]	14733	1623	504	0,11	0,31
f5-QQ-f	[Q]c1ccc2nc([Q])cnc2c1	4049	447	41	0,11	0,09
f5-QQ-g	[Q]c1ccc2ncc([Q])nc2c1	4075	564	73	0,14	0,13
f5-QQ-h	[Q]c1cnc2cccc([Q])c2n1	429	73	6	0,17	0,08
f5-QQ-i	[Q]c1cc2nccnc2cc1[Q]	3534	241	26	0,07	0,11
f6-QQ-a	[Q]c1ccc2cnnc([Q])c2c1	702	35	5	0,05	0,14
f6-QQ-b	[Q]c1ccc2c(cnnc2[Q])c1	580	23	3	0,04	0,13
f6-QQ-c	[Q]c1cccc2c([Q])nncc12	139	2	0	0,01	0,00
f6-QQ-d	[Q]c1nncc([Q])c2cccc12	3062	389	41	0,13	0,11
f6-QQ-e	[Q]c1ccc2cncc2c1[Q]	88	14	4	0,16	0,29
f6-QQ-f	[Q]c1cc2cncc2c([Q])c1	37	0	0	0,00	-
f6-QQ-g	[Q]c1ccc([Q])c2cncc12	18	1	0	0,06	0,00
f6-QQ-h	[Q]c1cc2cncc2cc1[Q]	271	12	1	0,04	0,08
f6-QQ-i	[Q]c1cccc2cnnc([Q])c12	125	10	4	0,08	0,40
f7-QQ-a	[Q]c1cc([Q])c2ncccc2n1	187	12	0	0,06	0,00
f7-QQ-b	[Q]c1ccnc2c([Q])ccnc12	5	1	1	0,20	1,00
f7-QQ-c	[Q]c1cnc2c([Q])ccnc2c1	101	8	2	0,08	0,25
f7-QQ-d	[Q]c1ccc2nccc([Q])c2n1	907	99	56	0,11	0,57
f7-QQ-e	[Q]c1nc2cccnc2cc1[Q]	196	9	9	0,05	1,00
f7-QQ-f	[Q]c1cnc2cc([Q])cnc2c1	20	4	4	0,20	1,00
f7-QQ-g	[Q]c1cnc2ccc([Q])nc2c1	451	39	39	0,09	1,00
f7-QQ-h	[Q]c1cnc2cccnc2c1[Q]	267	3	3	0,01	1,00
f7-QQ-i	[Q]c1ccc2nc([Q])ccc2n1	22	4	0	0,18	0,00
f8-QQ-a	[Q]c1ncc2ccc([Q])nc2c1	133	56	50	0,42	0,89
f8-QQ-b	[Q]c1ccc2c(ccnc2[Q])n1	20	9	1	0,45	0,11
f8-QQ-c	[Q]c1cc([Q])c2ncccc2n1	10	0	0	0,00	-
f8-QQ-d	[Q]c1nc2ccncc2cc1[Q]	63	6	0	0,10	0,00
f8-QQ-e	[Q]c1ncc2cccnc2c1[Q]	13	1	0	0,08	0,00
f8-QQ-f	[Q]c1ncc([Q])c2ncccc12	331	24	3	0,07	0,13
f8-QQ-g	[Q]c1cncc2c([Q])ccnc12	8	0	0	0,00	-

6. Część eksperymentalna

f8-QQ-h	[Q]c1cnc2c(cnc2[Q])c1	8	1	0	0,13	0,00
f8-QQ-i	[Q]c1cc2ncccc2c([Q])n1	65	2	0	0,03	0,00
f8-QQ-j	[Q]c1ccc2cncc([Q])c2n1	6	2	0	0,33	0,00
f8-QQ-k	[Q]c1ncc2c([Q])ccnc2c1	26	7	0	0,27	0,00
f8-QQ-l	[Q]c1ccnc2ccnc([Q])c12	11	0	0	0,00	-
f8-QQ-m	[Q]c1ncc2cc([Q])cnc2c1	1	0	0	0,00	-
f8-QQ-n	[Q]c1cnc2ccnc([Q])c2c1	8	1	0	0,13	0,00
f8-QQ-o	[Q]c1cnc2ccncc2c1[Q]	10	0	0	0,00	-
f9-QQ-a	[Q]c1ncc2nccc([Q])c2c1	92	54	1	0,59	0,02
f9-QQ-b	[Q]c1ccnc2c([Q])nccc12	29	0	0	0,00	-
f9-QQ-c	[Q]c1cc([Q])c2ccncc2n1	20	1	0	0,05	0,00
f9-QQ-d	[Q]c1cnc2cnccc2c1[Q]	19	0	0	0,00	-
f9-QQ-e	[Q]c1ncc2ncccc2c1[Q]	2	0	0	0,00	-
f9-QQ-f	[Q]c1cnc([Q])c2ncccc12	8	0	0	0,00	-
f9-QQ-g	[Q]c1ccc2c([Q])cncc2n1	1	0	0	0,00	-
f9-QQ-h	[Q]c1cnc2cncc([Q])c2c1	1	0	0	0,00	-
f9-QQ-i	[Q]c1cc2cccnc2c([Q])n1	92	0	0	0,00	-
f9-QQ-j	[Q]c1cncc2nccc([Q])c12	8	0	0	0,00	-
f9-QQ-k	[Q]c1ncc2nc([Q])ccc2c1	14	0	0	0,00	-
f9-QQ-l	[Q]c1ccc2ccnc([Q])c2n1	10	0	0	0,00	-
f9-QQ-m	[Q]c1ncc2ncc([Q])cc2c1	1	0	0	0,00	-
f9-QQ-n	[Q]c1cnc2c(ccnc2[Q])c1	6	0	0	0,00	-
f9-QQ-o	[Q]c1nc2cnccc2cc1[Q]	18	0	0	0,00	-

^a Dane na dzień 5 listopada 2010.

^b Parametr I zdefiniowano w Rozdziale 4.2.2 poprzez Równanie 4.3.

^c Parametr B zdefiniowano w Rozdziale 4.2.2 poprzez Równanie 4.2.

6.2 FRAGMENTACJA DEKREMENTACYJNA LEKÓW AKTYWNYCH WZGLĘDEM ENZYMU INTEGRAZY WIRUSA HIV – DANE LICZBOWE

Fragmentacja dekrementacyjna cząsteczek leków stosowanych w terapii przeciw HIV polegała na wydzielaniu fragmentów molekularnych poprzez stopniowe zmniejszanie wyjściowej struktury chemicznej. Po każdym odcięciu jednostki strukturalnej, przeprowadzano przeszukiwanie bazy ChemDB HIV, Opportunistic Infection and Tuberculosis Therapeutics Database (ChemDB HIV/OI/TB) w oparciu o podstrukturę zminimalizowanego fragmentu molekularnego. Opis repozytorium sporządzono w Rozdziale 3.4.1.3. Zastosowano protokół fragmentacji dekrementacyjnej dla cząsteczki raltegraviru i elwitegraviru. Algorytm procedury opisano w Rozdziale 4.3.1. Wyboru leków dokonano w oparciu o mechanizm stymulacji farmakologicznej, wybierając środki terapeutyczne należące do klasy inhibitorów IN. Produkty lecznicze zarejestrowane w terapii HIV/AIDS opisano w Rozdziale 3.7.1. Tabele poniżej (Tabela 6.5, Tabela 6.6 oraz Tabela 6.7) przedstawiają szczegółowe wyniki uzyskane dla poszczególnych ścieżek rozłączeń. Indeks fragmentu koduje sekwencyjny zapis kierunku przeprowadzanej instrukcji rozłączeń (Rysunek 4.22 w Rozdziale 4.3.2 oraz Rysunek 4.29 w Rozdziale 4.3.3). W trzeciej kolumnie tabel podano liczbę rekordów odnotowaną w bazie, czyli pulę związków zawierających dany fragment molekularny jako podstrukturę.

Tabela 6.5 Wyniki uzyskane dla cząsteczki raltegrawiru (ralt) na drodze przeprowadzonej procedury fragmentacji dekrementacyjnej.

Fragment molekularny ^a	SMILES	Liczba hitów ^b
raltegrawir	<chem>CN1C(=O)C(O)=C(N=C1C(C)(C)NC(=O)C1=NN=C(C)O1)C(=O)NCC1=CC=C(F)C=C1</chem>	1 (2)
Ścieżka fragmentacji I		
ralt_A	<chem>CN1C(=O)C(O)=C(N=C1C(C)(C)NC(=O)C1=NN=CO1)C(=O)NCC1=CC=C(F)C=C1</chem>	1 (2)
ralt_AB	<chem>CN1C(=O)C(O)=C(N=C1C(C)(C)NC(=O)C(=O)NC1=CC=C(F)C=C1</chem>	7 (9)
ralt_ABC	<chem>CNC(C)(C)C1=NC(C(=O)NCC2=CC=C(F)C=C2)=C(O)C(=O)N1C</chem>	66 (76)
ralt_ABCD	<chem>CN1C(=O)C(O)=C(N=C1C(C)(C)N)C(=O)NCC1=CC=C(F)C=C1</chem>	68 (78)
ralt_ABCDE	<chem>CC(C)C1=NC(C(=O)NCC2=CC=C(F)C=C2)=C(O)C(=O)N1C</chem>	179 (292)
ralt_ABCDEF	<chem>CCC1=NC(C(=O)NCC2=CC=C(F)C=C2)=C(O)C(=O)N1C</chem>	294 (458)
ralt_ABCDEFG	<chem>CN1C(C)=NC(C(=O)NCC2=CC=C(F)C=C2)=C(O)C1=O</chem>	308 (492)
ralt_ABCDEFGH	<chem>CN1C=NC(C(=O)NCC2=CC=C(F)C=C2)=C(O)C1=O</chem>	308 (492)
ralt_ABCDEFGHI	<chem>OC1=C(N=CNC1=O)C(=O)NCC1=CC=C(F)C=C1</chem>	308 (492)
ralt_ABCDEFGHIJ	<chem>OC1=C(N=CNC1)C(=O)NCC1=CC=C(F)C=C1</chem>	308 (493)
ralt_ABCDEFGHIJK	<chem>FC1=CC=C(CNC(=O)C2=CCNC=N2)C=C1</chem>	308 (493)
ralt_ABCDEFGHIJKL	<chem>FC1=CC=C(CNC=O)C=C1</chem>	712 (1420)
ralt_ABCDEFGHIJKLM	<chem>CNCC1=CC=C(F)C=C1</chem>	942 (2413)
ralt_ABCDEFGHIJKLMN	<chem>NCC1=CC=C(F)C=C1</chem>	1051 (2681)
ralt_ABCDEFGHIJKLMNO	<chem>CC1=CC=C(F)C=C1</chem>	1411 (6207)
ralt_ABCDEFGHIJKLMNOP	<chem>FC1=CC=CC=C1</chem>	2896 (14903)
Ścieżka fragmentacji II		
ralt_R	<chem>CN1C(=O)C(O)=C(N=C1C(C)(C)NC(=O)C1=NN=C(C)O1)C(=O)NCC1=CC=CC=C1</chem>	1 (2)
ralt_RP	<chem>CNC(=O)C1=C(O)C(=O)N(C)C(=N1)C(C)(C)NC(=O)C1=NN=C(C)O1</chem>	1 (2)
ralt_RPO	<chem>CN1C(=O)C(O)=C(N=C1C(C)(C)NC(=O)C1=NN=C(C)O1)C(N)=O</chem>	1 (2)
ralt_RPON	<chem>CN1C(=O)C(O)=C(C(=O)N=C1C(C)(C)NC(=O)C1=NN=C(C)O1</chem>	1 (2)
ralt_RPONM	<chem>CN1C(=O)C(O)=C(C)N=C1C(C)(C)NC(=O)C1=NN=C(C)O1</chem>	1 (2)
ralt_RPONML	<chem>CN1C(=O)C(O)=CN=C1C(C)(C)NC(=O)C1=NN=C(C)O1</chem>	1 (2)
ralt_RPONMLK	<chem>CN1C(=O)C=CN=C1C(C)(C)NC(=O)C1=NN=C(C)O1</chem>	1 (2)
ralt_RPONMLKJ	<chem>CN1CC=CN=C1C(C)(C)NC(=O)C1=NN=C(C)O1</chem>	1 (2)
ralt_RPONMLKJI	<chem>CC1=NN=C(O1)C(=O)NC(C)(C)C1=NC=CCN1</chem>	1 (2)
ralt_RPONMLKJIH	<chem>CC(C)NC(=O)C1=NN=C(C)O1</chem>	4 (6)
ralt_RPONMLKJIHG	<chem>CCNC(=O)C1=NN=C(C)O1</chem>	4 (11)
ralt_RPONMLKJIHGF	<chem>CNC(=O)C1=NN=C(C)O1</chem>	4 (11)

ralt_RPONMLKJIHGFE	CC1=NN=C(O)C(N)=O	4 (11)
ralt_RPONMLKJIHGFED	CC1=NN=C(O)C=O	5 (13)
ralt_RPONMLKJIHGFEDC	CC1=NN=C(C)O1	24 (297)
ralt_RPONMLKJIHGFEDCB	CC1=NN=CO1	31 (605)
Ścieżka fragmentacji III		
ralt_(AR)	CN1C(=O)C(O)=C(N=C1C(C)(C)NC(=O)C1=NN=CO1)C(=O)NCC1=CC=CC=C1	1 (2)
ralt_(AR)B	CN1C(=O)C(O)=C(N=C1C(C)(C)NC(=O)C(=O)NC1=CC=CC=C1	7 (9)
ralt_(AR)BP	CNC(=O)C1=C(O)C(=O)N(C)C(=N1)C(C)(C)NC=O	7 (9)
ralt_(AR)BPC	CNC(=O)C1=C(O)C(=O)N(C)C(=N1)C(C)(C)NC	67 (78)
ralt_(AR)BPCO	CNC(C)(C)C1=NC(C(N)=O)=C(O)C(=O)N1C	67 (78)
ralt_(AR)BPCOD	CN1C(=O)C(O)=C(N=C1C(C)(C)N)C(N)=O	69 (80)
ralt_(AR)BPCODN	CN1C(=O)C(O)=C(C=O)N=C1C(C)(C)N	69 (80)
ralt_(AR)BPCODNE	CC(C)C1=NC(C=O)=C(O)C(=O)N1C	322 (534)
ralt_(AR)BPCODNEF	CCC1=NC(C=O)=C(O)C(=O)N1C	449 (712)
ralt_(AR)BPCODNEFM	CCC1=NC(C)=C(O)C(=O)N1C	449 (712)
ralt_(AR)BPCODNEFMG	CN1C(C)=NC(C)=C(O)C1=O	463 (751)
ralt_(AR)BPCODNEFMGL	CN1C(C)=NC=C(O)C1=O	463 (751)
ralt_(AR)BPCODNEFMGLH	CN1C=NC=C(O)C1=O	463 (752)
ralt_(AR)BPCODNEFMGLHI	OC1=CN=CNC1=O	463 (753)
ralt_(AR)BPCODNEFMGLHIK	O=C1NC=NC=C1	1061 (5045)
ralt_(AR)BPCODNEFMGLHIKJ	C1NC=NC=C1	1067 (5398)
Ścieżka fragmentacji IV		
ralt_(RPONML)	CN1C(=O)C(O)=CN=C1C(C)(C)NC(=O)C1=NN=C(C)O1	1 (2)
ralt_(RPONML)A	CN1C(=O)C(O)=CN=C1C(C)(C)NC(=O)C1=NN=CO1	1 (2)
ralt_(RPONML)AB	CN1C(=O)C(O)=CN=C1C(C)(C)NC=O	7 (9)
ralt_(RPONML)ABC	CNC(C)(C)C1=NC=C(O)C(=O)N1C	67 (78)
ralt_(RPONML)ABCD	CN1C(=O)C(O)=CN=C1C(C)(C)N	69 (80)
ralt_(RPONML)ABCDE	CC(C)C1=NC=C(O)C(=O)N1C	322 (534)
ralt_(RPONML)ABCDEF	CCC1=NC=C(O)C(=O)N1C	449 (712)
ralt_(RPONML)ABCDEFG	CN1C(C)=NC=C(O)C1=O	463 (751)
ralt_(RPONML)ABCDEFGH	CN1C=NC=C(O)C1=O	463 (752)
ralt_(RPONML)ABCDEFGHI	OC1=CN=CNC1=O	463 (753)
ralt_(RPONML)ABCDEFGHIJ	OC1=CN=CNC1	463 (755)
ralt_(RPONML)ABCDEFGHIJK	C1NC=NC=C1	1067 (5398)
Ścieżka fragmentacji V		
ralt_(ABCDEFGH)	CN1C=NC(C(=O)NCC2=CC=C(F)C=C2)=C(O)C1=O	308 (492)
ralt_(ABCDEFGH)R	CN1C=NC(C(=O)NCC2=CC=CC=C2)=C(O)C1=O	456 (645)
ralt_(ABCDEFGH)RP	CNC(=O)C1=C(O)C(=O)N(C)C=N1	463 (746)
ralt_(ABCDEFGH)RPO	CN1C=NC(C(N)=O)=C(O)C1=O	463 (746)
ralt_(ABCDEFGH)RPON	CN1C=NC(C=O)=C(O)C1=O	463 (751)
ralt_(ABCDEFGH)RPONM	CN1C=NC(C)=C(O)C1=O	463 (751)
ralt_(ABCDEFGH)RPONML	CN1C=NC=C(O)C1=O	463 (752)
ralt_(ABCDEFGH)RPONMLK	CN1C=NC=CC1=O	471 (1216)
ralt_(ABCDEFGH)RPONMLKJ	CN1CC=CN=C1	476 (1360)

- ralt_(ABCDEFGH)RPONMLKJI

C1NC=NC=C1

1067 (5398)
- a

Indeks fragmentu koduje sekwencyjny zapis kierunku przeprowadzanej instrukcji rozłączeń. W nawiasach okrągłych podano elementy podstruktury odłączone w jednym, wstępnym etapie, poprzedzającym procedurę właściwej fragmentacji.
- b

Wyniki uzyskano stosując składnię zapytania strukturalnego, połączonego z filtrami zapytań tekstowych, ograniczających liczbę uzyskanych rekordów tylko do struktur posiadających adnotację w bazie w kierunku zastosowań w terapii przeciw HIV. Dodatkowo w nawiasach podano całkowitą liczbę rekordów (bez zastosowania filtra przynależności klasowej przeciw HIV). Wyniki uzyskano metodą eksploracji bazy ChemDB HIV/OI/TB. Dane na dzień 19 grudnia 2010.

Tabela 6.6 Wyniki uzyskane dla uogólnionej reprezentacji struktury cząsteczki raltegrawiru (ralt-Q) na drodze przeprowadzonej procedury fragmentacji dekrementacyjnej.

Fragment molekularny ^a	SMILES	Liczba hitów ^a
raltegrawir-Q	CQ1C(=Q)C(Q)=C(Q=C1C(C)(C)QC(=Q)C1=QQ =C(C)Q1)C(=Q)QCC1=CC=C(Q)C=C1	1
Ścieżka fragmentacji I-Q		
ralt-Q_A	CQ1C(=Q)C(Q)=C(Q=C1C(C)(C)QC(=Q)C1=QQ =CQ1)C(=Q)QCC1=CC=C(Q)C=C1	1
ralt-Q_AB	CQ1C(=Q)C(Q)=C(Q=C1C(C)(C)QC(=Q)C(=Q)Q CC1=CC=C(Q)C=C1	7
ralt-Q_ABC	CQC(C)(C)C1=QC(C(=Q)QCC2=CC=C(Q)C=C2) =C(Q)C(=Q)Q1C	197
ralt-Q_ABCD	CQ1C(=Q)C(Q)=C(Q=C1C(C)(C)Q)C(=Q)QCC1= CC=C(Q)C=C1	199
ralt-Q_ABCDE	CC(C)C1=QC(C(=Q)QCC2=CC=C(Q)C=C2)=C(Q)C(=Q)Q1C	199
ralt-Q_ABCDEF	CCC1=QC(C(=Q)QCC2=CC=C(Q)C=C2)=C(Q)C(=Q)Q1C	316
ralt-Q_ABCDEFG	CQ1C(C)=QC(C(=Q)QCC2=CC=C(Q)C=C2)=C(Q)C1=Q	330
ralt-Q_ABCDEFGH	CQ1C=QC(C(=Q)QCC2=CC=C(Q)C=C2)=C(Q)C 1=Q	330
ralt-Q_ABCDEFGHI	QC1=C(Q=CQC1=Q)C(=Q)QCC1=CC=C(Q)C=C 1	330
ralt-Q_ABCDEFGHIJ	QC1=C(Q=CQC1)C(=Q)QCC1=CC=C(Q)C=C1	330
ralt-Q_ABCDEFGHIJK	QC1=CC=C(CQC(=Q)C2=CCQC=Q2)C=C1	330
ralt-Q_ABCDEFGHIJKL	QC1=CC=C(CQC=Q)C=C1	1727
ralt-Q_ABCDEFGHIJKLM	CQCC1=CC=C(Q)C=C1	2942
ralt-Q_ABCDEFGHIJKLMN	QCC1=CC=C(Q)C=C1	3617
ralt-Q_ABCDEFGHIJKLMNO	CC1=CC=C(Q)C=C1	8660
ralt-Q_ABCDEFGHIJKLMNOP	QC1=CC=CC=C1	18388
Ścieżka fragmentacji II-Q		
ralt-Q_R	CQ1C(=Q)C(Q)=C(Q=C1C(C)(C)QC(=Q)C1=QQ =C(C)Q1)C(=Q)QCC1=CC=CC=C1	1
ralt-Q_RP	CQC(=Q)C1=C(Q)C(=Q)Q(C)C(=Q1)C(C)(C)QC(	1

ralt-Q_RPO	=Q)C1=QQ=C(C)Q1 CQ1C(=Q)C(Q)=C(Q=C1C(C)(C)QC(=Q)C1=QQ =C(C)Q1)C(Q)=Q	1
ralt-Q_RPON	CQ1C(=Q)C(Q)=C(C=Q)Q=C1C(C)(C)QC(=Q)C1 =QQ=C(C)Q1	1
ralt-Q_RPONM	CQ1C(=Q)C(Q)=C(C)Q=C1C(C)(C)QC(=Q)C1=Q Q=C(C)Q1	1
ralt-Q_RPONML	CQ1C(=Q)C(Q)=CQ=C1C(C)(C)QC(=Q)C1=QQ= C(C)Q1	1
ralt-Q_RPONMLK	CQ1C(=Q)C=CQ=C1C(C)(C)QC(=Q)C1=QQ=C(C)Q1	1
ralt-Q_RPONMLKJ	CQ1CC=CQ=C1C(C)(C)QC(=Q)C1=QQ=C(C)Q1	1
ralt-Q_RPONMLKJI	CC1=QQ=C(Q1)C(=Q)QC(C)(C)C1=QC=CCQ1	1
ralt-Q_RPONMLKJIH	CC(C)QC(=Q)C1=QQ=C(C)Q1	4
ralt-Q_RPONMLKJIHG	CCQC(=Q)C1=QQ=C(C)Q1	5
ralt-Q_RPONMLKJIHGF	CQC(=Q)C1=QQ=C(C)Q1	8
ralt-Q_RPONMLKJIHGFE	CC1=QQ=C(Q1)C(Q)=Q	11
ralt-Q_RPONMLKJIHGFED	CC1=QQ=C(Q1)C=Q	62
ralt-Q_RPONMLKJIHGFEDC	CC1=QQ=C(C)Q1	133
ralt-Q_RPONMLKJIHGFEDCB	CC1=QQ=CQ1	388

^a Indeks fragmentu koduje sekwencyjny zapis kierunku przeprowadzanej instrukcji rozłączeń. W nawiasach okrągłych podano elementy podstruktury odłączone w jednym, wstępnym etapie, poprzedzającą procedurę właściwej fragmentacji.

^b Wyniki uzyskano stosując składnię zapytania strukturalnego, połączonego z filtrami zapytań tekstowych, ograniczających liczbę uzyskanych rekordów tylko do struktur posiadających adnotację w bazie w kierunku zastosowań w terapii przeciw HIV. Wyniki uzyskano metodą eksploracji bazy ChemDB HIV/OI/TB. Dane na dzień 19 grudnia 2010.

Tabela 6.7 Wyniki uzyskane dla cząsteczki elwitegrawiru (elwit) na drodze przeprowadzonej procedury fragmentacji dekrementacyjnej.

Fragment molekularny	SMILES	Liczba hitów ^a
elwitegrawir	COC1=C(CC2=C(F)C(CI)=CC=C2)C=C2C(=O)C(=CN([C@H])(CO)C(C)C)C2=C1)C(O)=O	12
Ścieżka fragmentacji I ^c		
elwit_M	COC1=C(CC2=C(F)C(CI)=CC=C2)C=C2C(=O)C(C=O)=CN([C@H])(CO)C(C)C)C2=C1	12
elwit_MN	COC1=C(CC2=C(F)C(CI)=CC=C2)C=C2C(=O)C(C)=CN([C@H])(CO)C(C)C)C2=C1	12
elwit_MNO	COC1=C(CC2=C(F)C(CI)=CC=C2)C=C2C(=O)C=CN([C@H])(CO)C(C)C)C2=C1	12
elwit_MNOP	COC1=C(CC2=C(F)C(CI)=CC=C2)C=C2CC=CN([C@H])(CO)C(C)C)C2=C1	12
elwit_MNOPJ	COC1=C(CC2=C(F)C(CI)=CC=C2)C=C2CC=CN([C@H])(C)C(C)C)C2=C1	12
elwit_MNOPJK	COC1=C(CC2=C(F)C(CI)=CC=C2)C=C2CC=CN(C(C)C)C2=C1	12
elwit_MNOPJKH	CCCN1C=CCC2=CC(CC3=C(F)C(CI)=CC=C3)=C(	18

6. Część eksperymentalna

	OC)C=C12	
elwit_MNOPJKHG	CCN1C=CCC2=CC(CC3=C(F)C(Cl)=CC=C3)=C(O)C=C12	28
elwit_MNOPJKHGI	COC1=C(CC2=C(F)C(Cl)=CC=C2)C=C2CC=CN(C)C2=C1	28
elwit_MNOPJKHGIL	COC1=C(CC2=C(F)C(Cl)=CC=C2)C=C2CC=CNC2=C1	28
elwit_MNOPJKHGILE	OC1=C(CC2=C(F)C(Cl)=CC=C2)C=C2CC=CNC2=C1	28
elwit_MNOPJKHGILEF	FC1=C(CC2=CC=C3NC=CCC3=C2)C=CC=C1Cl	82
elwit_MNOPJKHGILEFD	CC1=C(F)C(Cl)=CC=C1	179
elwit_MNOPJKHGILEFDC	FC1=CC=CC=C1Cl	195
elwit_MNOPJKHGILEFDCB	C1C1=CC=CC=C1	4624
elwit_MNOPJKHGILEFDCBA	C1=CC=CC=C1	24202
Ścieżka fragmentacji II ^c		
elwit_A	COC1=C(CC2=C(F)C=CC=C2)C=C2C(=O)C(=CN([C@H](CO)C(C)C)C2=C1)C(O)=O	12
elwit_AB	COC1=C(CC2=CC=CC=C2)C=C2C(=O)C(=CN([C@H](CO)C(C)C)C2=C1)C(O)=O	12
elwit_ABC	COC1=C(C)C=C2C(=O)C(=CN([C@H](CO)C(C)C)C2=C1)C(O)=O	12
elwit_ABCD	COC1=CC=C2C(=O)C(=CN([C@H](CO)C(C)C)C2=C1)C(O)=O	12
elwit_ABCDE	CC(C)[C@@H](CO)N1C=C(C(O)=O)C(=O)C2=CC=C(O)C=C12	12
elwit_ABCDEF	CC(C)[C@@H](CO)N1C=C(C(O)=O)C(=O)C2=CC=CC=C12	23
elwit_ABCDEFG	CC[C@@H](CO)N1C=C(C(O)=O)C(=O)C2=CC=CC=C12	39
elwit_ABCDEFGH	C[C@@H](CO)N1C=C(C(O)=O)C(=O)C2=CC=CC=C12	48
elwit_ABCDEFGHI	OCCN1C=C(C(O)=O)C(=O)C2=CC=CC=C12	133
elwit_ABCDEFGHIJ	CCN1C=C(C(O)=O)C(=O)C2=CC=CC=C12	178
elwit_ABCDEFGHIJK	CN1C=C(C(O)=O)C(=O)C2=CC=CC=C12	203
elwit_ABCDEFGHIJKL	OC(=O)C1=CNC2=CC=CC=C2C1=O	213
elwit_ABCDEFGHIJKLM	O=CC1=CNC2=CC=CC=C2C1=O	217
elwit_ABCDEFGHIJKLMN	CC1=CNC2=CC=CC=C2C1=O	217
elwit_ABCDEFGHIJKLMNO	O=C1C=CNC2=CC=CC=C12	220
elwit_ABCDEFGHIJKLMNOP	C1C=CNC2=CC=CC=C12	220
Ścieżka fragmentacji III ^c		
elwit_G	CC[C@@H](CO)N1C=C(C(O)=O)C(=O)C2=CC(CC3=C(F)C(Cl)=CC=C3)=C(OC)C=C12	18
elwit_GH	COC1=C(CC2=C(F)C(Cl)=CC=C2)C=C2C(=O)C(=CN([C@@H](C)CO)C2=C1)C(O)=O	20
elwit_GHI	COC1=C(CC2=C(F)C(Cl)=CC=C2)C=C2C(=O)C(=CN(CCO)C2=C1)C(O)=O	28
elwit_GHIJ	CCN1C=C(C(O)=O)C(=O)C2=CC(CC3=C(F)C(Cl)=CC=C3)=C(OC)C=C12	28
elwit_GHIJK	COC1=C(CC2=C(F)C(Cl)=CC=C2)C=C2C(=O)C(=CN(C)C2=C1)C(O)=O	28

elwit_GHIJKL	<chem>COC1=C(CC2=C(F)C(Cl)=CC=C2)C=C2C(=O)C(=CNC2=C1)C(O)=O</chem>	28
elwit_GHIJKLM	<chem>COC1=C(CC2=C(F)C(Cl)=CC=C2)C=C2C(=O)C(C=O)=CNC2=C1</chem>	28
elwit_GHIJKLMN	<chem>COC1=C(CC2=C(F)C(Cl)=CC=C2)C=C2C(=O)C(C)=CNC2=C1</chem>	28
elwit_GHIJKLMNO	<chem>COC1=C(CC2=C(F)C(Cl)=CC=C2)C=C2C(=O)C=CNC2=C1</chem>	28
elwit_GHIJKLMNOP	<chem>COC1=C(CC2=C(F)C(Cl)=CC=C2)C=C2CC=CNC2=C1</chem>	28
elwit_GHIJKLMNOPB	<chem>COC1=C(CC2=CC(Cl)=CC=C2)C=C2CC=CNC2=C1</chem>	29
elwit_GHIJKLMNOPBA	<chem>COC1=C(CC2=CC=CC=C2)C=C2CC=CNC2=C1</chem>	29
elwit_GHIJKLMNOPBAE	<chem>OC1=C(CC2=CC=CC=C2)C=C2CC=CNC2=C1</chem>	30
elwit_GHIJKLMNOPBAEF	<chem>C(C1=CC=CC=C1)C1=CC=C2NC=CCC2=C1</chem>	191
Ścieżka fragmentacji IV^d		
elwit_(ABC)	<chem>COC1=C(C)C=C2C(=O)C(=CN([C@H](CO)C(C)C)C2=C1)C(O)=O</chem>	12
elwit_(ABC)M	<chem>COC1=C(C)C=C2C(=O)C(C=O)=CN([C@H](CO)C(C)C)C2=C1</chem>	12
elwit_(ABC)MN	<chem>COC1=C(C)C=C2C(=O)C(C)=CN([C@H](CO)C(C)C)C2=C1</chem>	12
elwit_(ABC)MNO	<chem>COC1=C(C)C=C2C(=O)C=CN([C@H](CO)C(C)C)C2=C1</chem>	12
elwit_(ABC)MNOP	<chem>COC1=C(C)C=C2CC=CN([C@H](CO)C(C)C)C2=C1</chem>	12
elwit_(ABC)MNOPJ	<chem>COC1=C(C)C=C2CC=CN([C@H](C)C(C)C)C2=C1</chem>	12
elwit_(ABC)MNOPJK	<chem>COC1=C(C)C=C2CC=CN(CC(C)C)C2=C1</chem>	12
elwit_(ABC)MNOPJKH	<chem>CCCN1C=CCC2=CC(C)=C(OC)C=C12</chem>	18
elwit_(ABC)MNOPJKHG	<chem>CCN1C=CCC2=CC(C)=C(OC)C=C12</chem>	29
elwit_(ABC)MNOPJKHGI	<chem>COC1=C(C)C=C2CC=CN(C)C2=C1</chem>	29
elwit_(ABC)MNOPJKHGIL	<chem>COC1=C(C)C=C2CC=CNC2=C1</chem>	29
elwit_(ABC)MNOPJKHGILE	<chem>CC1=C(O)C=C2NC=CCC2=C1</chem>	30
elwit_(ABC)MNOPJKHGILEF	<chem>CC1=CC=C2NC=CCC2=C1</chem>	195
elwit_(ABC)MNOPJKHGILEFD	<chem>C1C=CNC2=CC=CC=C12</chem>	220
Ścieżka fragmentacji V^e		
elwit_(ABC)	<chem>COC1=C(C)C=C2C(=O)C(=CN([C@H](CO)C(C)C)C2=C1)C(O)=O</chem>	12
elwit_(ABC)D	<chem>COC1=CC=C2C(=O)C(=CN([C@H](CO)C(C)C)C2=C1)C(O)=O</chem>	12
elwit_(ABC)DE	<chem>CC(C)[C@@H](CO)N1C=C(C(O)=O)C(=O)C2=CC=C(O)C=C12</chem>	12
elwit_(ABC)DEF	<chem>CC(C)[C@@H](CO)N1C=C(C(O)=O)C(=O)C2=CC=CC=C12</chem>	23
elwit_(ABC)DEFG	<chem>CC[C@@H](CO)N1C=C(C(O)=O)C(=O)C2=CC=CC=C12</chem>	39
elwit_(ABC)DEFGH	<chem>C[C@@H](CO)N1C=C(C(O)=O)C(=O)C2=CC=CC=C12</chem>	48
elwit_(ABC)DEFGHI	<chem>OCCN1C=C(C(O)=O)C(=O)C2=CC=CC=C12</chem>	133

6. Część eksperymentalna

elwit_(ABC)DEFGHIJ	CCN1C=C(C(O)=O)C(=O)C2=CC=CC=C12	178
elwit_(ABC)DEFGHIJK	CN1C=C(C(O)=O)C(=O)C2=CC=CC=C12	203
elwit_(ABC)DEFGHIJKL	OC(=O)C1=CNC2=CC=CC=C2C1=O	213
elwit_(ABC)DEFGHIJKLM	O=CC1=CNC2=CC=CC=C2C1=O	217
elwit_(ABC)DEFGHIJKLMN	CC1=CNC2=CC=CC=C2C1=O	217
elwit_(ABC)DEFGHIJKLMNO	O=C1C=CNC2=CC=CC=C12	220
elwit_(ABC)DEFGHIJKLMNOP	C1C=CNC2=CC=CC=C12	220
Ścieżka fragmentacji VI ^e		
elwit_(ABC)	COC1=C(C)C=C2C(=O)C(=CN([C@H](CO)C(C)C)C2=C1)C(O)=O	12
elwit_(ABC)G	CC[C@H](CO)N1C=C(C(O)=O)C(=O)C2=CC(C)=C(OC)C=C12	18
elwit_(ABC)GH	COC1=C(C)C=C2C(=O)C(=CN([C@@H](C)CO)C2=C1)C(O)=O	20
elwit_(ABC)GHI	COC1=C(C)C=C2C(=O)C(=CN(CCO)C2=C1)C(O)=O	29
elwit_(ABC)GHII	CCN1C=C(C(O)=O)C(=O)C2=CC(C)=C(OC)C=C12	29
elwit_(ABC)GHIJK	COC1=C(C)C=C2C(=O)C(=CN(C)C2=C1)C(O)=O	29
elwit_(ABC)GHIJKL	COC1=C(C)C=C2C(=O)C(=CNC2=C1)C(O)=O	29
elwit_(ABC)GHIJKLM	COC1=C(C)C=C2C(=O)C(C=O)=CNC2=C1	29
elwit_(ABC)GHIJKLMN	COC1=C(C)C=C2C(=O)C(C)=CNC2=C1	29
elwit_(ABC)GHIJKLMNO	COC1=C(C)C=C2C(=O)C=CNC2=C1	29
elwit_(ABC)GHIJKLMNOP	COC1=C(C)C=C2CC=CNC2=C1	29
elwit_(ABC)GHIJKLMNOPD	COC1=CC=C2CC=CNC2=C1	29
elwit_(ABC)GHIJKLMNOPDE	OC1=CC=C2CC=CNC2=C1	30
elwit_(ABC)GHIJKLMNOPDEF	C1C=CNC2=CC=CC=C12	220
^a	Indeks fragmentu koduje sekwencyjny zapis kierunku przeprowadzanej instrukcji rozłączeń. W nawiasach okrągłych podano elementy podstruktury odłączone w jednym, wstępnym etapie, poprzedzającym procedurę właściwej fragmentacji.	
^b	Wyniki uzyskano stosując składnię zapytania strukturalnego, połączonego z filtrami zapytań tekstowych, ograniczających liczbę uzyskanych rekordów tylko do struktur posiadających adnotację w bazie w kierunku zastosowań w terapii przeciw HIV. Wyniki uzyskano metodą eksploracji ChemDB HIV/OI/TB.	
^c	Dane na dzień 19 grudnia 2010.	
^d	Dane na dzień 7 lutego 2011.	
^e	Dane na dzień 8 kwietnia 2011.	

6.3 OZNACZENIE CYTOTOKSYCZNOŚCI I AKTYWNOŚCI WZGLĘDEM WIRUSA HIV¹¹

Badane preparaty (w ilości 5 mg każdy) rozpuszczono w 500 µl DMSO. Związek referencyjny FZ-41 (w ilości 3,94 mg) rozpuszczono w 394 µl DMSO. Następnie 50 µl otrzymanego roztworu rozpuszczono w 5000 µl RPMI1640/FCS 10%. Końcowe stężenie DMSO poniżej 1%. Tak otrzymane roztwory związków użyto do oznaczeń w hodowlach komórkowych. Hodowle wirusa HIV prowadzono w komórkach ustalonej linii komórkowej CEM-T4. W hodowlach do inokulacji wykorzystano laboratoryjny izolat HIV wyprowadzony w naszym laboratorium (wild type, fenotyp SI indukujący syncycja).

Proces zahamowania replikacji wirusa w hodowlach badano poprzez oznaczenie ilości białka p24 (najbardziej reprezentatywnego białka wirusowego) w nadsączach znad hodowli komórek CEM-T4 inokulowanych wirusem HIV prowadzonych w medium wzbogaconym w znane stężenia testowanych związków. Hodowle prowadzono na 96-cio dołkowych płytkach hodowlanych w standardowych warunkach (20 tys./studzienka, temp. 37°C, 5% CO₂).

Inokulację poprzedzała dwudziestoczerogodzinna preinkubacja komórek w znanych stężeniach testowanych substancji. Dla każdego testowanego stężenia badanych preparatów hodowle były wykonane w trzech powtórzeniach. Kontrolę referującą 100% replikację stanowiły hodowle prowadzone w standardowym medium. Po siedmiu dniach hodowli z każdej studzienki zbierano 80 µl nadsączy i oznaczano zawartość białka p24 przy użyciu testu Genscreen HIV-1 Ag Assay (Bio-Rad, France). Po pobraniu nadsączy w hodowlach oznaczano żywotność komórek metodą MTT (Sigma, USA). W trakcie prowadzenia hodowli wykonywano codzienne obserwacje komórek w mikroskopie odwróconym w celu obserwacji tworzenia się lub braku syncycji będących wynikiem zakażenia.

Cytotoksyczność preparatów oceniano po siedmiu dniach hodowli komórek linii CEM-T4 prowadzonych w standardowych warunkach (medium RPMI wzbogacone

¹¹ Opis procedury oznaczenia cytotoksyczności i aktywności względem wirusa HIV dostarczony przez wykonawcę testów biologicznych – zespół badawczy Samodzielnej Pracowni Wirusologii Narodowego Instytutu Leków w Warszawie pod kierownictwem Pana dr Andrzeja Piasek.

w 10% surowicy płodowej (FCS), temp. 37°C w atmosferze zawierającej 5% CO₂). Hodowle komórkowe prowadzono na 96-studzienkowych płytkach hodowlanych w szeregu rozcieńczeń testowanych preparatów. Hodowle kontrolne prowadzono w standardowym medium hodowlanym (RPMI, 10% FCS, liczba komórek w momencie rozpoczęcia hodowli 20 tys/studzienka). Po siedmiu dniach hodowli żywotność komórek oznaczano metodą MTT. Hodowle prowadzono w trzech powtórzeniach dla każdego stężenia poszczególnych preparatów.

7 SPIS ILUSTRACJI

RYSUNEK 3.1	LEK OD POMYSŁU DO WDROŻENIA. SKRÓT SAR ODNOSI SIĘ DO BADAŃ ZALEŻNOŚCI POMIĘDZY STRUKTURĄ CZĄSTECZKI, A AKTYWNOŚCIĄ BIOLOGICZNĄ.	12
RYSUNEK 3.2	LICZBA LEKÓW ZATWIERDZONYCH PRZEZ FDA I WPROWADZONYCH NA RYNEK W OSTATNICH DZIESIĘCIU LATACH. DANE PRZEDSTAWIONE NA WYKRESIE OBEJMUJĄ TYLKO LEKI ZAWIERAJĄCE NOWE STRUKTURY CHEMICZNE JAKO SUBSTANCJE AKTYWNE (LEKI TYPU NME). POMINIĘTO INFORMACJE DOTYCZĄCE LICZBY NOWYCH LEKÓW GENERYCZNYCH.	13
RYSUNEK 3.3	EWOLUCJA MODELU OBOWIĄZUJĄCEGO W BADANIACH NAD NOWYM LEKIEM.	15
RYSUNEK 3.4	LICZBA PUBLIKACJI TRAKTUJĄCYCH O BADANIACH NAD POSZUKIWANIEM NOWYCH LEKÓW Z UWZGLĘDNIENIEM PODZIAŁU W ZALEŻNOŚCI OD PRZYJĘTEJ STRATEGII BADAWCZEJ. DANE UZYSKANO METODĄ EKSPLOKACJI LITERATUROWEJ BAZY WEB OF SCIENCE. WYRÓŻNIONO PIERWSZE PUBLIKACJE DLA KAŻDEJ STRATEGII. (BADANIA WŁASNE – MATERIAŁ PRZYGOTOWYWANY DO DRUKU).	16
RYSUNEK 3.5	PODZIAŁ METOD W ZADANIACH WIRTUALNEGO SKRININGU.	17
RYSUNEK 3.6	SCHEMAT PROCEDURY WIRTUALNEGO SKRININGU.	18
RYSUNEK 3.7	SCHEMAT KLASYFIKATORA BINARNEGO W ROZWIĄZYWANIU ZADAŃ CHEMOINFORMATYCZNYCH.	21
RYSUNEK 3.8	REPREZENTACJA NUMERYCZNA ZBIORU K MOLEKUŁ W POSTACI N -WYMIAROWEGO WEKTORA.	22
RYSUNEK 3.9	PREFERENCYJNOŚĆ WYBRANYCH MOLEKUŁ DO TWORZENIA WIĘKSZEJ LICZBY NOWYCH POŁĄCZEŃ W FUNKCJI CZASU. CZĄSTECZKA M_1 JEST ŹRÓDŁEM DUŻEJ LICZBY NOWYCH SZLAKÓW SYNTETYCZNYCH (A), NATOMIAST CZĄSTECZKA M_2 W ZNACZNIE MNIEJSZYM STOPNIU ROZBUDOWUJE SIĘ CHEMICZNYCH POŁĄCZEŃ (B). PREFERENCYJNOŚĆ JEST ZACHOWANA W FUNKCJI CZASU.	30
RYSUNEK 3.10	PRZYKŁADY ZWIĄZKÓW ZAWIERAJĄCYCH PODSTRUKTURĘ CHINOLINY I NALEŻĄCYCH DO GRUPY LEKÓW ZATWIERDZONYCH PRZEZ FDA, U. S. FOOD AND DRUG ADMINISTRATION. W NAWIASACH PODANO LICZBĘ RECEPTORÓW Z KTÓRYMI LEKI TE ODDZIAŁUJĄ. DANE UZYSKANO METODĄ EKSPLOKACJI BAZY DRUGBANK. DANE NA DZIEŃ 14 LUTY 2013.	36
RYSUNEK 3.11	POLIFARMAKOLOGICZNA DWUCZĘŚCIOWA SIĘ ODDZIAŁYWAŃ LEK-RECEPTOR (B) ORAZ SKONSTRUOWANE SIECI JEDNOCZĘŚCIOWE LEK-LEK (C) I RECEPTOR-RECEPTOR (A). DWA WĘZŁY SIECI JEDNOCZĘŚCIOWEJ POŁĄCZONE SĄ JEŻELI W SIECI DWUCZĘŚCIOWEJ POSIADAJĄ ELEMENTY WSPÓLNE.	37
RYSUNEK 3.12	PROTOTYP UPRZYWILEJOWANYCH MOTYWÓW STRUKTURALNYCH [64].	38
RYSUNEK 3.13	LICZBA PUBLIKOWANYCH COROCZNIE ARTYKUŁÓW TRAKTUJĄCYCH O STRUKTURACH UPRZYWILEJOWANYCH ZINDEKSOWANYCH W WYBRANYCH LITERATUROWYCH BAZACH DANYCH. WYNIKI UZYSKANO FORMUŁUJĄC ZAPYTANIE TEKSTOWE JAKO KOMBINACJE OKREŚLENIA „PRIVILEGED” W POŁĄCZENIU Z NASTĘPUJĄCYMI WYRAZAMI „STRUCTURE/ -S”, „SUBSTRUCTURE/ -S”, „SCAFFOLD/ -S”, „FRAGMENT/ -S”. BAZA SCIENCE DIRECT ORAZ ACS POSIADAJĄ OPCJĘ PRZESZUKIWANIA PEŁNOTEKSTOWEGO (ZA WYJĄTKIEM REFERENCJI). W PRZYPADKU POZOSTAŁYCH BAZ MOŻLIWE BYŁO PRZESZUKIWANIE JEDYNI W OBRĘBIE TYTUŁU/TEMATU (WEB OF SCIENCE) LUB TYTUŁU/SŁÓW KLUCZOWYCH/ABSTRAKTU (REAXYS).	39
RYSUNEK 3.14	UPRZYWILEJOWANIE STRUKTUR CHEMICZNYCH W UKŁADACH BIOLOGICZNYCH.	44

RYSUNEK 3.15	KLASYFIKACJA ŚRODKÓW TERAPEUTYCZNYCH STOSOWANYCH W LECZENIU INFЕКCJI HIV/AIDS ZE WZGLĘDU NA MECHANIZM STYMULACJI FARMAKOLOGICZNEJ.....	46
RYSUNEK 3.16	LEKI ZATWIERDZONE PRZEZ FDA W TERAPII PRZECIW HIV. KOLORY KWADRATÓW KODUJĄ PRZYNALEŻNOŚĆ PRODUKTU HANDLOWEGO DO DANEJ KLASY FARMACEUTYKÓW ZGODNIE Z MECHANIZMEM STYMULACJI FARMAKOLOGICZNEJ SUBSTANCJI CZYNNEJ LEKU. NRTIs – NUKLEOZYDOWE INHIBITORY ODWROTNEJ TRANSKRYPTAZY (ANG. NUCLEOSIDE REVERSE TRANSCRIPTASE INHIBITORS); PIs – INHIBITORY PROTEAZY (ANG. PROTEASE INHIBITORS); NNRTIs – NIENUKLEOZYDOWE INHIBITORY ODWROTNEJ TRANSKRYPTAZY (ANG. NONNUCLEOSIDE REVERSE TRANSCRIPTASE INHIBITORS); MCCP – LEKI WIELOSKŁADNIKOWE (ANG. MULTI-CLASS COMBINATION PRODUCTS); FI – INHIBITORY FUZJI, (ANG. FUSION INHIBITORS); EI – INHIBITORY WNIKANIA WIRUSA DO KOMÓRKI GOSPODARZA, BLOKERY KORECEPTORÓW CCR5 (ANG. ENTRY INHIBITORS – CCR5 CO-RECEPTOR ANTAGONIST); HIV ISTI – INHIBITOR INTEGRAZY (ANG. HIV INTEGRASE STRAND TRANSFER INHIBITORS).	49
RYSUNEK 3.17	INHIBITORY PROTEAZY, PI.....	52
RYSUNEK 3.18	NUKLEOZYDOWE INHIBITORY ODWROTNEJ TRANSKRYPTAZY, NRTI.	52
RYSUNEK 3.19	NIENUKLEOZYDOWE INHIBITORY ODWROTNEJ TRANSKRYPTAZY, NNRTI.	53
RYSUNEK 3.20	INHIBITOR WNIKANIA WIRUSA DO KOMÓRKI GOSPODARZA – BLOKER KORECEPTORÓW CCR5, EI.	54
RYSUNEK 3.21	INHIBITORY INTEGRAZY, HIV ISTI.	54
RYSUNEK 4.1	FRAGMENT MONO- (F1 I F2) I DIAZANAFTALENOWY (F3 - F12).....	61
RYSUNEK 4.2	STATYSTYKA ROZPOWSZECHNIENIA FRAGMENTÓW AZANAFTALENOWYCH JAKO PODSTRUKTUR ZNANYCH ZWIĄZKÓW ORGANICZNYCH ORAZ LEKÓW. DANE UZYSKANO METODĄ EKSPLOKACJI BAZY BEILSTEIN ORAZ DRUGBANK. LICZBA REKORDÓW BAZY DRUGBANK ZOSTAŁA OGRANICZONA DO LEKÓW POSIADAJĄCYCH STATUS ZATWIERDZENIA PRZYNAJMNIEJ W JEDNYM KRAJU. DANE NA DZIEŃ 4 KWIEŃNIA 2010.	62
RYSUNEK 4.3	WARTOŚCI PARAMETRU L DLA FRAGMENTÓW MOLEKULARNYCH F1 - F12.	63
RYSUNEK 4.4	WARTOŚCI PARAMETRU L WYLICZONE DLA FRAGMENTÓW MOLEKULARNYCH F1 - F12 W ODNIESIENIU DO LICZEBNOŚCI ZBIORU.....	64
RYSUNEK 4.5	PRZESTRZEŃ BIOAKTYWNOŚCI FRAGMENTU MONO- I DIAZANAFTALENOWEGO. PARAMETR B ORAZ I ZDEFINIOWANO POPRZECZ RÓWNANIE 4.2 ORAZ RÓWNANIE 4.3. ŚREDNICA KÓŁ ODZWIERCIEDLA LICZEBNOŚĆ POSZCZEGÓLNYCH DWUNASTU ZBIORÓW ZWIĄZKÓW CHEMICZNYCH POSIADAJĄCYCH JEDEN Z FRAGMENTÓW MOLEKULARNYCH F1 - F12 JAKO ELEMENT PODSTRUKTURY. DIAGRAM SPORZĄDZONO NA PODSTAWIE DANYCH UZYSKANYCH METODĄ EKSPLOKACJI BAZY PUBCHEM COMPOUND. DANE NA DZIEŃ 27 MAJA 2010.	66
RYSUNEK 4.6	PRZYROST POPULACJI BADANYCH ZBIORÓW. DIAGRAM SPORZĄDZONO NA PODSTAWIE DANYCH UZYSKANYCH METODĄ EKSPLOKACJI BAZY PUBCHEM COMPOUND. DANE PRZEDSTAWIONO W DWÓCH SERIACH: PO SZEŚCIOMIESIĘCZNYM OKRESIE ANALIZY (ETAP 1) I NASTĘPNIE PO KOLEJNYCH SZEŚCIU MIESIĄCACH (ETAP 2). W OKREŚLA POPULACJĘ ZWIĄZKÓW POSIADAJĄCYCH JEDEN Z FRAGMENTÓW MOLEKULARNYCH F1 - F12 JAKO ELEMENT PODSTRUKTURY, NATOMIAST T oraz A WSKAZUJĄ ZBIORY ZWIĄZKÓW TESTOWANYCH I AKTYWNYCH WYODRĘBNIONE W KAŻDYM Z DWUNASTU BADANYCH ZBIORÓW W.	68
RYSUNEK 4.7	ZMIANA WARTOŚCI PARAMETRU B W CZASIE. DIAGRAM SPORZĄDZONO NA PODSTAWIE DANYCH UZYSKANYCH METODĄ EKSPLOKACJI BAZY PUBCHEM COMPOUND. DANE PRZEDSTAWIONO W TRZECH SERIACH KOLEJNO NA DZIEŃ 27 MAJA 2010, 6 LISTOPADA 2010 ORAZ 6 MAJA 2011.	70

RYSUNEK 4.8	TRAJEKTORIA ZMIAN WARTOŚCI PARAMETRÓW B ORAZ I W CZASIE. DIAGRAM SPORZĄDZONO NA PODSTAWIE DANYCH UZYSKANYCH METODĄ EKSPLOKACJI BAZY PUBCHEM COMPOUND. DANE PRZEDSTAWIONO W TRZECH SERIACH KOLEJNO NA DZIEŃ 27 MAJA 2010, 6 LISTOPADA 2010 ORAZ 6 MAJA 2011.	71
RYSUNEK 4.9	ZESTAWIENIE WYNIKÓW UZYSKANYCH W OKRESIE ROCZNEGO MONITORINGU DANYCH, DOTYCZĄCYCH FRAGMENTU AZANAFTALENOWEGO, EWIDENCJONOWANYCH W BAZIE PUBCHEM COMPOUND.	72
RYSUNEK 4.10	PRZESTRZEŃ BIOAKTYWNOŚCI FRAGMENTU MONO- I DIAZANAFTALENOWEGO PODSTAWIONEGO JEDNYM (A) I DWOMA (B) HETEROATOMAMI. PARAMETR B ORAZ I ZDEFINIOWANO POPRZEC RÓWNIANIE 4.2 ORAZ RÓWNIANIE 4.3. ŚREDNICA KÓŁ ODZWIERCIEDLA LICZEBNOŚĆ POSZCZEGÓLNYCH ZBIORÓW STRUKTUR CHEMICZNYCH. DIAGRAM SPORZĄDZONO NA PODSTAWIE DANYCH UZYSKANYCH METODĄ EKSPLOKACJI BAZY PUBCHEM COMPOUND. DANE NA DZIEŃ 5 LISTOPADA 2010.	74
RYSUNEK 4.11	UPRZYWILEJOWANA ARCHITEKTURA PODSTAWIEŃ FRAGMENTU AZANAFTALENOWEGO. GÓRNY RZĄD PRZEDSTAWIA TRZY TOPOLOGIE MONOPODSTAWIEŃ, NATOMIAST DOLNY RZĄD ILUSTRUJE TRZY KOMPOZYCJE DIPODSTAWIEŃ, DLA KTÓRYCH UZYSKANO NAJWYŻSZE WARTOŚCI PARAMETRU B. Q OZNACZA DOWOLNY HETEROATOM, A PODANE POD STRUKTURAMI LICZBY OZNACZAJĄ INDEKS FRAGMENTU ORAZ WYLICZONĄ WARTOŚĆ PARAMETRU B. .	75
RYSUNEK 4.12	PRZESTRZEŃ BIOAKTYWNOŚCI FRAGMENTU CHINOLINOWEGO F1 W ZALEŻNOŚCI OD ARCHITEKTURY PODSTAWIENIA TRZONU MOLEKULARNEGO. Q OZNACZA DOWOLNY HETEROATOM, A PODANE POD STRUKTURAMI LICZBY OZNACZAJĄ INDEKS FRAGMENTU ORAZ WYLICZONĄ WARTOŚĆ PARAMETRU B.	75
RYSUNEK 4.13	PROFIL UPRZYWILEJOWANIA ARCHITEKTURY PODSTAWIEŃ FRAGMENTU AZANAFTALENOWEGO JEDNYM (A) I DWOMA (B) HETEROATOMAMI. DIAGRAM SPORZĄDZONO NA PODSTAWIE DANYCH UZYSKANYCH METODĄ EKSPLOKACJI BAZY PUBCHEM COMPOUND. W OKREŚLA POPULACJĘ ZWIĄZKÓW POSIADAJĄCYCH JEDEN Z FRAGMENTÓW MOLEKULARNYCH F1-12 JAKO ELEMENT PODSTRUKTURY O OKREŚLONEJ ORNAMENTACJI PODSTAWNIKÓW, NATOMIAST T ORAZ A WSKAZUJĄ ZBIORY ZWIĄZKÓW TESTOWANYCH I AKTYWNYCH WYODRĘBNIONE W KAŻDYM Z BADANYCH ZBIORÓW W. DANE NA DZIEŃ 5 LISTOPADA 2010.	78
RYSUNEK 4.14	ZESTAWIENIE WARTOŚCI WSKAŹNIKÓW DOSTĘPNOŚCI SYNTETYCZNEJ ZWIĄZKÓW AZANAFTALENOWYCH. PARAMETR SA ORAZ P ZDEFINIOWANO ODPOWIEDNIO POPRZEC RÓWNIANIE 4.4 ORAZ RÓWNIANIE 4.5. DANE UZYSKANO METODĄ EKSPLOKACJI BAZY BEILSTEIN STOSUJĄC ZAPYTANIE O PODSTRUKTURĘ. DANE NA DZIEŃ 4 KWIETNIA 2010.	82
RYSUNEK 4.15	RANKING UWZGLĘDNIAJĄCY LICZEBNOŚĆ ZBIORÓW S, R ORAZ P DLA POSZCZEGÓLNYCH FRAGMENTÓW AZANAFTALENOWYCH.	83
RYSUNEK 4.16	ZESTAWIENIE WARTOŚCI WSKAŹNIKÓW DOSTĘPNOŚCI SYNTETYCZNEJ ZWIĄZKÓW AZANAFTALENOWYCH. PARAMETR SA, ORAZ P, ZDEFINIOWANO ODPOWIEDNIO POPRZEC RÓWNIANIE 4.6 ORAZ RÓWNIANIE 4.7. DANE UZYSKANO METODĄ EKSPLOKACJI BAZY BEILSTEIN STOSUJĄC ZAPYTANIE O PODSTRUKTURĘ. DANE NA DZIEŃ 4 KWIETNIA 2010.	84
RYSUNEK 4.17	KONSTRUKCJA ZAPYTANIA STRUKTURALNEGO DO BAZY. R OZNACZA DOWOLNY PODSTAWNIK. ZAPYTANIE STRUKTURALNE O REAKCJĘ CHEMICZNĄ STOSOWANO W TRYBIE DOPASOWANIA DOKŁADNEGO. WYNIKI NIE OBEJMUJĄ REAKCJI, W KTÓRYCH UDZIAŁ BIORĄ ZWIĄZKI CHEMICZNE, ZAWIERAJĄCE WSKAZANE MOTYWY STRUKTURALNE JAKO PODSTRUKTURY.	85

RYSUNEK 4.18	KIERUNKOWE PREFERENCJE TWORZENIA DOWOLNEGO WIĄZANIA (A) ORAZ WIĄZANIA C-C (B) DLA MOTYWU STRUKTURALNEGO CHINOLINY I IZOCHINOLINY. DANE UZYSKANO METODĄ EKSPLOKACJI BAZY REAXYS. DANE NA DZIEŃ 30 PAŹDZIERNIKA 2012.	87
RYSUNEK 4.19	LICZBA OPUBLIKOWANYCH ARTYKUŁÓW TRAKTUJĄCYCH O ZWIĄZKACH AZANAFTALENOWYCH Z UWZGLĘDNIENIEM PODZIAŁU NA DWANAŚCIE ZBIORÓW STRUKTUR CHEMICZNYCH POSIADAJĄCYCH JEDEN Z FRAGMENTÓW MOLEKULARNYCH F1 - F12 JAKO ELEMENT PODSTRUKTURY. DIAGRAM SPORZĄDZONO NA PODSTAWIE DANYCH UZYSKANYCH METODĄ EKSPLOKACJI BAZY BEILSTEIN. DANE NA DZIEŃ 14 MAJA 2010.	88
RYSUNEK 4.20	TENDENCJE I TRENDY SYNTEZY ORGANICZNEJ NA PRZESTRZENI OSTATNICH OSIEMDZIESIĘCIU DZIEWIĘCIU LAT (1820-2009). FLUKTUACJE W LICZBIE POZYCJI LITERATUROWYCH WIDOCZNE NA DIAGRAMACH W POSTACI SWOISTEGO FALUJĄCEGO PROFILU PRZEDSTAWIONO DLA TRZECH WYBRANYCH POPULACJI STRUKTUR CHEMICZNYCH ZAWIERAJĄCYCH ODPOWIEDNIO FRAGMENT MOLEKULARNY O TOPOLOGII OKREŚLONEJ INDEKSEM PORZĄDKOWYM F1 (A), F5 (B) ORAZ F2 (C). DANE UZYSKANO METODĄ EKSPLOKACJI BAZY BEILSTEIN. ODNOŚNIKI LITERATUROWE REFERUJĄ DO OPISANYCH REAKCJI SYNTEZ ZWIĄZKÓW ZAWIERAJĄCYCH JAKO PODSTRUKTURĘ ROZWAŻANE FRAGMENTY MOLEKULARNE. DANE NA DZIEŃ 14 MAJA 2010.	90
RYSUNEK 4.21	ALGORYTM TWORZENIA ŚCIEŻEK ROZŁĄCZEŃ W PROCEDURZE FRAGMENTACJI DEKREMENTACYJNEJ ORAZ SPOSÓB INDEKSOWANIA FRAGMENTÓW MOLEKULARNYCH NA PRZYKŁADZIE CZĄSTECZKI RALTEGRAWIRU.	92
RYSUNEK 4.22	SPOSÓB INDEKSOWANIA WIĄZAŃ W PROCEDURZE FRAGMENTACJI DEKREMENTACYJNEJ DLA CZĄSTECZKI RALTEGRAWIRU. STRUKTURA RALT-Q JEST UOGÓLNIONĄ REPREZENTACJĄ STRUKTURY CZĄSTECZKI RALTEGRAWIRU (RALT), W KTÓREJ W MIEJSCU ATOMU AZOTU, TLENU I FLUORU DOPUSZCZONO WYSTĄPIENIE DOWOLNEGO HETEROATOMU (SYMBOL Q).	94
RYSUNEK 4.23	LICZEBNOŚĆ FRAGMENTÓW MOLEKULARNYCH WYODRĘBNIONYCH NA DRODZE FRAGMENTACJI DEKREMENTACYJNEJ CZĄSTECZKI RALTEGRAWIRU (RALT) W ZALEŻNOŚCI OD ROZMIARU WYDZIELONEGO ELEMENTU PODSTRUKTURY. WIELKOŚĆ FRAGMENTU MOLEKULARNEGO ODPOWIADA LICZBIE ATOMÓW WCHODZĄCYCH W JEGO SKŁAD (BEZ UWZGLĘDNIENIA ATOMÓW WODORU).	94
RYSUNEK 4.24	POPULACJA FRAGMENTÓW MOLEKULARNYCH WYODRĘBNIONYCH NA DRODZE FRAGMENTACJI DEKREMENTACYJNEJ CZĄSTECZKI RALTEGRAWIRU (RALT) W ZALEŻNOŚCI OD ROZMIARU WYDZIELONEGO ELEMENTU PODSTRUKTURY. WIELKOŚĆ FRAGMENTU MOLEKULARNEGO ODPOWIADA LICZBIE ATOMÓW WCHODZĄCYCH W JEGO SKŁAD (BEZ UWZGLĘDNIENIA ATOMÓW WODORU). LICZBA SŁUPKÓW DLA POSZCZEGÓLNEJ KATEGORII ZWIĄZANEJ Z WIELKOŚCIĄ FRAGMENTU ODPOWIADA LICZBIE WYODRĘBNIONYCH ELEMENTÓW PODSTRUKTURY O DANEJ WIELKOŚCI, NP. WYODRĘBNIONO PIĘĆ RÓŻNYCH FRAGMENTÓW MOLEKULARNYCH POSIADAJĄCYCH SIEDEM ATOMÓW W STRUKTURZE, NATOMIAST WYODRĘBNIONO TYLKO JEDEN FRAGMENT MOLEKULARNY POSIADAJĄCY SZESNAŚCIE ATOMÓW W STRUKTURZE. WYNIKI UZYSKANO METODĄ EKSPLOKACJI BAZY CHEMDB HIV/OI/TB.	95
RYSUNEK 4.25	STRUKTURY NAJMNIEJSZYCH (OD SZEŚCIU DO DZIESIĘCIU ATOMÓW W STRUKTURZE) WYODRĘBNIONYCH FRAGMENTÓW MOLEKULARNYCH. PODANO STOPIEŃ ROZPOWSZECHNIENIA DANEGO MOTYWU JAKO ELEMENTU PODSTRUKTURY ZWIĄZKÓW ZDEPONOWANYCH W BAZIE CHEMDB HIV/OI/TB I POSIADAJĄCYCH ADNOTACJE NA TEMAT ZASTOSOWANIA W TERAPII PRZECIW HIV.	96
RYSUNEK 4.26	PORÓWNANIE DWÓCH ŚCIEŻEK FRAGMENTACJI DEKREMENTACYJNEJ CZĄSTECZKI RALTEGRAWIRU. PRZEDSTAWIONO LICZBĘ REKORDÓW UZYSKANYCH W BAZIE PO KAŻDYM KROKU PRZEPROWADZENIA INSTRUKCJI	

REDUKUJĄCEJ. WYNIKI UZYSKANO METODĄ EKSPORACJI BAZY CHEMDB HIV/OI/TB DLA STRUKTURY RALT ORAZ STRUKTURY RALT-Q. DANE NA DZIEŃ 19 GRUDNIA 2010.	97
RYSUNEK 4.27 SWOISTE ELEMENTY STRUKTURALNE CZĄSTECZKI RALTEGRAWIRU. LINIĄ PRZERYWANĄ ZAZNACZONO NAJMNIEJSZY WYODRĘBNIONY FRAGMENT, KTÓRY OBECNY JEST TYLKO I WYŁĄCZNIE W STRUKTURZE ROZWAŻANEGO LEKU I NIE WYSTĘPUJE JAKO ELEMENT PODSTRUKTURY INNYCH ZWIĄZKÓW ZDEPONOWANYCH W BAZIE CHEMDB HIV/OI/TB. FRAGMENTY O INDEKSACH RALT_RPONMLKJIH ORAZ RALT-Q_RPONMLKJIH Z NISKĄ CZĘSTOTLIWOŚCIĄ WYSTĘPUJĄ JAKO ELEMENTY PODSTRUKTUR ZWIĄZKÓW SKATALOGOWANYCH W BAZIE.	98
RYSUNEK 4.28 OBSZARY SPECYFICZNOŚCI STRUKTURALNEJ CZĄSTECZKI RALTEGRAWIRU.	99
RYSUNEK 4.29 SPOSÓB INDEKSOWANIA WIĄZAŃ W PROCEDURZE FRAGMENTACJI DEKREMENTACYJNEJ DLA CZĄSTECZKI ELWITEGRAWIRU.	100
RYSUNEK 4.30 LICZEBNOŚĆ FRAGMENTÓW MOLEKULARNYCH WYODRĘBNIONYCH NA DRODZE FRAGMENTACJI DEKREMENTACYJNEJ CZĄSTECZKI ELWITEGRAWIRU (ELWIT) W ZALEŻNOŚCI OD ROZMIARU WYDZIELONEGO ELEMENTU PODSTRUKTURY. WIELKOŚĆ FRAGMENTU MOLEKULARNEGO ODPOWIADA LICZBIE ATOMÓW WCHODZĄCYCH W JEGO SKŁAD (BEZ UWZGLĘDNIENIA ATOMÓW WODORU).	100
RYSUNEK 4.31 POPULACJA FRAGMENTÓW MOLEKULARNYCH WYODRĘBNIONYCH NA DRODZE FRAGMENTACJI DEKREMENTACYJNEJ CZĄSTECZKI ELWITEGRAWIRU (ELWIT) W ZALEŻNOŚCI OD ROZMIARU WYDZIELONEGO ELEMENTU PODSTRUKTURY. WIELKOŚĆ FRAGMENTU MOLEKULARNEGO ODPOWIADA LICZBIE ATOMÓW WCHODZĄCYCH W JEGO SKŁAD (BEZ UWZGLĘDNIENIA ATOMÓW WODORU). LICZBA SŁUPKÓW DLA POSZCZEGÓLNEJ KATEGORII ZWIĄZANEJ Z WIELKOŚCIĄ FRAGMENTU ODPOWIADA LICZBIE WYODRĘBNIONYCH ELEMENTÓW PODSTRUKTURY O DANEJ WIELKOŚCI, NP. WYODRĘBNIONO PIĘĆ RÓŻNYCH FRAGMENTÓW MOLEKULARNYCH POSIADAJĄCYCH DZIEWIĘTNĄSCIE ATOMÓW W STRUKTURZE, NATOMIAST WYODRĘBNIONO TYLKO JEDEN FRAGMENT MOLEKULARNY POSIADAJĄCY SZESĆ ATOMÓW W STRUKTURZE. WYNIKI UZYSKANO METODĄ EKSPORACJI BAZY CHEMDB HIV/OI/TB.	101
RYSUNEK 4.32 STRUKTURY NAJMNIEJSZYCH (OD SZESZCIU DO PIĘTNASTU ATOMÓW W STRUKTURZE) WYODRĘBNIONYCH FRAGMENTÓW MOLEKULARNYCH. PODANO STOPIEŃ ROZPOWSZECZENIA DANEGO MOTYWU JAKO ELEMENTU PODSTRUKTURY ZWIĄZKÓW ZDEPONOWANYCH W BAZIE CHEMDB HIV/OI/TB I POSIADAJĄCYCH ADNOTACJE NA TEMAT ZASTOSOWANIA W TERAPII PRZECIW HIV.	102
RYSUNEK 4.33 PORÓWNANIE DWÓCH ŚCIEŻEK FRAGMENTACJI DEKREMENTACYJNEJ CZĄSTECZKI ELWITEGRAWIRU. PRZEDSTAWIONO LICZBĘ REKORDÓW UZYSKANYCH W BAZIE PO KAŻDYM KROKU PRZEPROWADZENIA INSTRUKCJI REDUKUJĄCEJ. WYNIKI UZYSKANO METODĄ EKSPORACJI BAZY CHEMDB HIV/OI/TB DLA STRUKTURY ELWIT. DANE NA DZIEŃ 19 GRUDNIA 2010.	103
RYSUNEK 4.34 OBSZARY SPECYFICZNOŚCI STRUKTURALNEJ CZĄSTECZKI ELWITEGRAWIRU.	104
RYSUNEK 4.35 SCHEMAT PODZIAŁU BIBLIOTEKI INHIBITORÓW IN NA ZBIORY ZWIĄZKÓW AKTYWNYCH/NIEAKTYWNYCH W ZALEŻNOŚCI OD ZDEFINIOWANYCH ZAKRESÓW WARTOŚCI PARAMETRU IC ₅₀	108
RYSUNEK 4.36 DESKRYPTOR DAKTYLOSKOWPOWY KLEKOTY-ROTHA (KLEKFP) I PRZYKŁADY FRAGMENTÓW MOLEKULARNYCH, KODOWANYCH W WYBRANYCH BITACH.	111
RYSUNEK 4.37 SCHEMAT KLASFIKATORA.	114
RYSUNEK 4.38 PANEL WYNIKÓW EFEKTYWNOŚCI METOD ML, TESTOWANYCH NA ZBIORZE INHIBITORÓW IN WIRUSA HIV-1 Z ZASTOSOWANIEM REPREZENTACJI NUMERYCZNEJ CZĄSTECZEK CHEMICZNYCH W POSTACI DESKRYPTORA	

	DAKTYLOSKOPOWEGO KLEKFP. ZESTAWIONO WARTOŚCI TRZECH METRYK SŁUŻĄCYCH OCENIE KLASYFIKATORÓW: CZUŁOŚCI, PRECYZJI ORAZ PARAMETRU KORELACYJNEGO MATTHEWSA (MCC). UWZGLĘDNIONO ROZRÓŻNIENIE ZE WZGLĘDU NA ALGORYTM WYBORU ZBIORU UCZĄCEGO. PANEL OBEJMUJE WYNIKI UZYSKANE DLA KLASYFIKATORÓW UCZONYCH NA ZBIORZE TRENINGOWYM GENEROWANYM W OPARCIU O ALGORYTM ZRÓŻNICOWANY, LOSOWY ORAZ REPREZENTATYWNY.....	115
RYSUNEK 4.39	PANEL WYNIKÓW EFEKTYWNOŚCI METOD ML, TESTOWANYCH NA ZBIORZE INHIBITORÓW IN WIRUSA HIV-1 Z ZASTOSOWANIEM REPREZENTACJI NUMERYCZNEJ CZĄSTECZEK CHEMICZNYCH W POSTACI DESKRYPTORA DAKTYLOSKOPOWEGO KLEKFP. ZESTAWIONO WARTOŚCI TRZECH METRYK SŁUŻĄCYCH OCENIE KLASYFIKATORÓW: CZUŁOŚCI, PRECYZJI ORAZ PARAMETRU KORELACYJNEGO MATTHEWSA (MCC). PANEL OBEJMUJE WYNIKI UZYSKANE DLA KLASYFIKATORÓW UCZONYCH NA ZBIORZE TRENINGOWYM GENEROWANYM W OPARCIU O METODĘ LSCO.	116
RYSUNEK 4.40	SCHEMAT OBLICZENIA WARTOŚCI WSKAŹNIKA MI-DSE DLA POSZCZEGÓLNYCH BITÓW DESKRYPTORA DAKTYLOSKOPOWEGO KLEKFP W TRZYNASTU WYODRĘBNIONYCH ZBIORACH INHIBITORÓW IN WIRUSA HIV-1.....	119
RYSUNEK 4.41	STOPIEŃ POKRYCIA STRUKTURY RALTEGRAWIRU FRAGMENTAMI WSKAZANYMI METODĄ MI-DSE JAKO UPRZYWILEJOWANE DLA INHIBITORÓW IN WIRUSA HIV-1. KOLOREM NIEBIESKIM ZILUSTROWANO WSPÓLNE MOTYWY STRUKTURALNE CZĄSTECZKI LEKU I DWUDZIEŚCIU DZIEWIĘCIU FRAGMENTÓW MOLEKULARNYCH WSKAZANYCH METODĄ MI-DSE. LINIĄ PRZERYWANĄ ZAKREŚLONO FRAGMENT O NAJWYŻSZEJ WARTOŚCI WSKAŹNIKA UPRZYWILEJOWANIA PS – BIT KR590. 123	
RYSUNEK 4.42	WYCINEK PLIKU TEKSTOWEGO ZAWIERAJĄCEGO OPIS KONSTRUKCJI KLASYFIKATORA BINARNEGO W OPARCIU O ZBIÓR REGUŁ TWORZONYCH METODĄ PART. KOLOREM NIEBIESKIM ZAZNACZONO ATRYBUTY WARUNKOWE WSKAZUJĄCE OBECNOŚĆ DANEGO FRAGMENTU MOLEKULARNEGO (ATT > 0). ATRYBUTY TE STANOWIĄ BITY WSPÓLNE Z BITAMI DESKRYPTORA DAKTYLOSKOPOWEGO KLEKFP WSKAZANYMI PRZEZ METODĘ MI-DSE.	124
RYSUNEK 4.43	SCHEMAT OPRACOWANEGO PROTOKOŁU VS.....	125
RYSUNEK 4.44	SKRINING WIRTUALNY BAZY DANYCH MDL DRUG DATA REPORT (MDDR).....	127
RYSUNEK 4.45	WYKRES ROC DLA MODUŁU PS.	129
RYSUNEK 4.46	WYKRES ROC DLA MODUŁÓW ML I PS POŁĄCZONYCH W KONFIGURACJI KASKADOWEJ.	130
RYSUNEK 4.47	SCHEMAT OPRACOWANEGO PROTOKOŁU VS ZASTOSOWANEGO DO PRZESZUKIWANIA KOMERCYJNEJ BAZY ZWIĄZKÓW CHEMICZNYCH W CELU IDENTYFIKACJI NOWYCH INHIBITORÓW IN WIRUSA HIV-1.	131
RYSUNEK 4.48	STRUKTURY ZWIĄZKÓW WYŁONIONYCH NA DRODZE OPRACOWANEGO PROTOKOŁU VS.....	135
RYSUNEK 5.1	DEFINIOWANIE PODSTRUKTURY. SCHEMAT ZACZERPIĘTY Z PUBLIKACJI [90].....	142
RYSUNEK 5.2	SCHEMAT POSZUKIWANIA ZBIORU ZWIĄZKÓW NIEAKTYWNYCH W OPARCIU O ALGORYTM DUD [157]. ..	143
RYSUNEK 5.3	MACIERZ KLASYFIKACJI.	144
RYSUNEK 5.4	WYKRES ROC. PUNKTY ODZWIERCIAŁAJĄ PRZYKŁADOWE WYNIKI UZYSKANE DLA SIĘDMIU RÓŻNYCH KLASYFIKATORÓW BINARNYCH. OPISY WSKAZUJĄ SPOSÓB INTERPRETACJI POSZCZEGÓLNYCH OBSZARÓW WYKRESU.	146

8 BIBLIOGRAFIA

- [1] Schneider G, Baringhaus K-H. Molecular design. Concepts and applications. Alemanha: Wiley-VCH Verlag GmbH & Co. KGaA; 2008.
- [2] Silverman RB. The Organic Chemistry of Drug Design and Drug Action. 2nd Editio. Elsevier; 2004.
- [3] Polański J, Bąk A. Podstawy chemoinformatyki leków. Oficyna Wydawnicza WW; 2010.
- [4] Patrick GL. An Introduction to Medicinal Chemistry. 4th Editio. Oxford University Press; 2009.
- [5] Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov* 2010;9:203–14.
- [6] Cuatrecasas P. Science and society Drug discovery in jeopardy. *J Clin Investig* 2006;116:2837–42.
- [7] Munos B. Lessons from 60 years of pharmaceutical innovation. *Nat Rev Drug Discov* 2009;8:959–68.
- [8] Pammolli F, Magazzini L, Riccaboni M. The productivity crisis in pharmaceutical R&D. *Nat Rev Drug Discov* 2011;10:428–38.
- [9] Mullard A. 2011 FDA drug approvals. *Nat Rev Drug Discov* 2012;11:91–4.
- [10] Hann MM, Leach AR, Harper G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *J Chem Inf Comput Sci* 2001;41:856–64.
- [11] Oprea TI, Davis AM, Teague SJ, Leeson PD. Is there a difference between leads and drugs? A historical perspective. *J Chem Inf Comput Sci* 2001;41:1308–15.
- [12] Teague S, Davis A, Leeson P, Oprea T. The Design of Leadlike Combinatorial Libraries. *Angew Chem Int Ed Engl* 1999;38:3743–8.
- [13] Tsinopoulos C, McCarthy IP. An Evolutionary Classification of the Strategies for Drug Discovery. In: Frizelle G, Richards H, editors. *Proc. 2002 Conf. Manuf. Complex. Netw. (Tackling Ind. Complex. ideas that make a Differ., University of Cambridge, U.K; 2001, p. 373–86.*
- [14] Ledford H. Translational research: the full cycle. *Nature* 2008;453:843–5.

- [15] Fishburn CS. Translational research: the changing landscape of drug discovery. *Drug Discov Today* 2013;18:487–94.
- [16] Woolf SH. The meaning of translational research and why it matters. *JAMA* 2008;299:211–3.
- [17] Johnson MA, Maggiora GM. Concepts and applications of molecular similarity. John Wiley & Sons; 1990.
- [18] Sadowski J, Kubinyi H. A scoring scheme for discriminating between drugs and nondrugs. *J Med Chem* 1998;41:3325–9.
- [19] Ajay A, Walters WP, Murcko MA. Can we learn to distinguish between “drug-like” and “nondrug-like” molecules? *J Med Chem* 1998;41:3314–24.
- [20] Lipinski CA. Drug-like properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol Methods* 2001;44:235–49.
- [21] Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem* 2002;45:2615–23.
- [22] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 1997;23:3–25.
- [23] Fialkowski M, Bishop KJM, Chubukov VA, Campbell CJ, Grzybowski BA. Architecture and evolution of organic chemistry. *Angew Chem Int Ed Engl* 2005;44:7263–9.
- [24] Ghose AK, Herbertz T, Salvino JM, Mallamo JP. Knowledge-based chemoinformatic approaches to drug discovery. *Drug Discov Today* 2006;11:1107–14.
- [25] Melville JL, Burke EK, Hirst JD. Machine learning in virtual screening. *Comb Chem High Throughput Screen* 2009;12:332–43.
- [26] Geppert H, Vogt M, Bajorath J. Current trends in ligand-based virtual screening: molecular representations, data mining methods, new application areas, and performance evaluation. *J Chem Inf Model* 2010;50:205–16.
- [27] Schwaighofer A, Schroeter T, Mika S, Blanchard G. How wrong can we get? A review of machine learning approaches and error bars. *Comb Chem High Throughput Screen* 2009;12:453–68.
- [28] Martins IF, Teixeira AL, Pinheiro L, Falcao AO. A bayesian approach to in silico blood-brain barrier penetration modeling. *J Chem Inf Model* 2012;52:1686–97.

- [29] Klon AE, Glick M, Davies JW. Application of machine learning to improve the results of high-throughput docking against the HIV-1 protease. *J Chem Inf Comput Sci* 2004;44:2216–24.
- [30] Hammann F, Gutmann H, Baumann U, Helma C, Drewe J. Classification of cytochrome p(450) activities using machine learning methods. *Mol Pharm* 2009;6:1920–6.
- [31] Schroeter T, Schwaighofer A, Mika S, Laak A Ter, Suelzle D, Ganzer U, et al. Machine learning models for lipophilicity and their domain of applicability. *Mol Pharm* 2007;4:524–38.
- [32] Podolyan Y, Walters MA, Karypis G. Assessing synthetic accessibility of chemical compounds using machine learning methods. *J Chem Inf Model* 2010;50:979–91.
- [33] Todeschini R, Consonni V, editors. *Handbook of Molecular Descriptors*. Weinheim, Germany: Wiley-VCH Verlag GmbH; 2000.
- [34] Hert J, Willett P, Wilton DJ, Acklin P, Azzaoui K, Jacoby E, et al. Comparison of topological descriptors for similarity-based virtual screening using multiple bioactive reference structures. *Org Biomol Chem* 2004;2:3256–66.
- [35] Plewczynski D, Spieser SAH, Koch U. Assessing different classification methods for virtual screening. *J Chem Inf Model* 2006;46:1098–106.
- [36] Bruce CL, Melville JL, Pickett SD, Hirst JD. Contemporary QSAR classifiers compared. *J Chem Inf Model* 2007;47:219–27.
- [37] Liu XH, Ma XH, Tan CY, Jiang YY, Go ML, Low BC, et al. Virtual screening of Abl inhibitors from large compound libraries by support vector machines. *J Chem Inf Model* 2009;49:2101–10.
- [38] Smusz S, Kurczab R, Bojarski AJ. The influence of the inactives subset generation on the performance of machine learning methods. *J Cheminform* 2013;5:17.
- [39] Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* 2006;34:D668–72.
- [40] Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res* 2008;36:D901–6.
- [41] Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, et al. DrugBank 3.0: a comprehensive resource for “omics” research on drugs. *Nucleic Acids Res* 2011;39:D1035–41.

- [42] Irwin JJ, Shoichet BK. ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model* 2005;45:177–82.
- [43] Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG. ZINC: A Free Tool to Discover Chemistry for Biology. *J Chem Inf Model* 2012.
- [44] Chen JH, Linstead E, Swamidass SJ, Wang D, Baldi P. ChemDB update--full-text search and virtual chemical space. *Bioinformatics* 2007;23:2348–51.
- [45] Bolton EE, Wang Y, Thiessen PA, Bryant SH. PubChem: Integrated Platform of Small Molecules and Biological Activities. *Annu. Reports Comput. Chem.* Vol. 4, Washington, DC: American Chemical Society; 2008.
- [46] Seiler KP, George GA, Happ MP, Bodycombe NE, Carrinski HA, Norton S, et al. ChemBank: a small-molecule screening and cheminformatics resource database. *Nucleic Acids Res* 2008;36:D351–9.
- [47] Overington J. ChEMBL. An interview with John Overington, team leader, chemogenomics at the European Bioinformatics Institute Outstation of the European Molecular Biology Laboratory (EMBL-EBI). Interview by Wendy A. Warr. *J Comput Aided Mol Des* 2009;23:195–8.
- [48] Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, et al. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res* 2012;40:D1100–7.
- [49] Bak A, Polanski J, Kurczyk A. The use of MoStBioDat for rapid screening of molecular diversity. *Molecules* 2009;14:3436–45.
- [50] Bak A, Polanski J, Stockner T, Kurczyk A. MoStBioDat--molecular and structural bioinformatics database. *Comb Chem High Throughput Screen* 2010;13:366–74.
- [51] Bak A, Magdziarz T, Kurczyk A, Polanski J. Mapping drug architecture by MoStBioDat: rapid screening of intramolecular hydrogen bonded motifs in catechols. *Drug Dev Res* 2011;72:209–18.
- [52] Bak A, Magdziarz T, Kurczyk A, Polanski J. Mapping Fragmental Drug-Likeness in the MoStBioDat Environment: Intramolecular Hydrogen Bonding Motifs in β -Ketoenols. *Comb Chem High Throughput Screen* 2011;14:560–9.
- [53] Frawley WJ, Piatetsky-Shapiro G, Matheus CJ. Knowledge Discovery in Databases : An Overview. *AI Mag* 1992;13:57–70.
- [54] Fayyad U, Piatetsky-Shapiro G, Smyth P. From Data Mining to Knowledge Discovery in Databases. *AI Mag* 1996;17:37–54.
- [55] Grzybowski BA, Bishop KJM, Kowalczyk B, Wilmer CE. The “wired” universe of organic chemistry. *Nat Chem* 2009;1:31–6.

- [56] Chen B, Wild D, Guha R. PubChem as a source of polypharmacology. *J Chem Inf Model* 2009;49:2044–55.
- [57] Costall B, Naylor RJ. The hypotheses of different dopamine receptor mechanisms. *Life Sci* 1981;28:215–29.
- [58] Vogt I, Mestres J. Drug-Target Networks. *Mol Inform* 2010;29:10–4.
- [59] Keiser MJ, Setola V, Irwin JJ, Laggner C, Abbas AI, Hufeisen SJ, et al. Predicting new molecular targets for known drugs. *Nature* 2009;462:175–81.
- [60] Cavalli A, Bolognesi ML, Minarini A, Rosini M, Tumiatti V, Recanatini M, et al. Multi-target-directed ligands to combat neurodegenerative diseases. *J Med Chem* 2008;51:347–72.
- [61] Cavalli A, Bolognesi ML. Neglected tropical diseases: multi-target-directed ligands in the search for novel lead candidates against *Trypanosoma* and *Leishmania*. *J Med Chem* 2009;52:7339–59.
- [62] Morphy R, Kay C, Rankovic Z. From magic bullets to designed multiple ligands. *Drug Discov Today* 2004;9:641–51.
- [63] Melchiorre C, Bolognesi ML, Minarini A, Rosini M, Tumiatti V. Polyamines in drug discovery: from the universal template approach to the multitarget-directed ligand design strategy. *J Med Chem* 2010;53:5906–14.
- [64] Evans B, Rittle K, Bock M, DiPardo R, Freidinger R, Whitter W, et al. Methods for drug discovery: development of potent, selective, orally effective cholecystokinin antagonists. *J Med Chem* 1988;31:2235–46.
- [65] Nicolaou KC, Pfefferkorn JA, Roecker AJ, Cao G-Q, Barluenga S, Mitchell HJ. Natural Product-like Combinatorial Libraries Based on Privileged Structures. 1. General Principles and Solid-Phase Synthesis of Benzopyrans. *J Am Chem Soc* 2000;122:9939–53.
- [66] Nicolaou KC, Pfefferkorn JA, Mitchell HJ, Roecker AJ, Barluenga S, Cao G-Q, et al. Natural Product-like Combinatorial Libraries Based on Privileged Structures. 2. Construction of a 10 000-Membered Benzopyran Library by Directed Split-and-Pool Chemistry Using NanoKans and Optical Encoding. *J Am Chem Soc* 2000;122:9954–67.
- [67] Nicolaou KC, Pfefferkorn JA, Barluenga S, Mitchell HJ, Roecker AJ, Cao G-Q, et al. Natural Product-like Combinatorial Libraries Based on Privileged Structures. 3. The “Libraries from Libraries” Principle for Diversity Enhancement of Benzopyran Libraries. *J Am Chem Soc* 2000;122:9968–76.
- [68] Ding S, Gray NS, Ding Q, Schultz PG. Expanding the diversity of purine libraries. *Tetrahedron Lett* 2001;42:8751–5.

- [69] Chang YT, Gray NS, Rosania GR, Sutherlin DP, Kwon S, Norman TC, et al. Synthesis and application of functionally diverse 2,6,9-trisubstituted purine libraries as CDK inhibitors. *Chem Biol* 1999;6:361–75.
- [70] Brun V, Legraverend M, Grierson DS. Cyclin-dependent kinase (CDK) inhibitors: development of a general strategy for the construction of 2,6,9-trisubstituted purine libraries. Part 1. *Tetrahedron Lett* 2001;42:8161–4.
- [71] Norman TC, Gray NS, Koh JT, Schultz PG. A Structure-Based Library Approach to Kinase Inhibitors 1996;7863:7430–1.
- [72] Hirschmann R, Hynes J, Cichy-Knight MA, van Rijn RD, Sprengeler PA, Spoor PG, et al. Modulation of receptor and receptor subtype affinities using diastereomeric and enantiomeric monosaccharide scaffolds as a means to structural and biological diversity. A new route to ether synthesis. *J Med Chem* 1998;41:1382–91.
- [73] Liu J, Underwood DJ, Cascieri MA, Rohrer SP, Cantin LD, Chicchi G, et al. Synthesis of a substance P antagonist with a somatostatin scaffold: factors affecting agonism/antagonism at GPCRs and the role of pseudosymmetry. *J Med Chem* 2000;43:3827–31.
- [74] Hirschmann R, Ducry L, Smith AB. Development of an efficient, regio- and stereoselective route to libraries based on the beta-D-glucose scaffold. *J Org Chem* 2000;65:8307–16.
- [75] Mason JS, Morize I, Menard PR, Cheney DL, Hulme C, Labaudiniere RF. New 4-point pharmacophore method for molecular similarity and diversity applications: overview of the method and applications, including a novel approach to the design of combinatorial libraries containing privileged substructures. *J Med Chem* 1999;42:3251–64.
- [76] Nilsson JW, Thorstensson F, Kvarnström I, Oprea T, Samuelsson B, Nilsson I. Solid-phase synthesis of libraries generated from a 4-phenyl-2-carboxy-piperazine scaffold. *J Comb Chem* 2001;3:546–53.
- [77] Willoughby CA, Hutchins SM, Rosauer KG, Dhar MJ, Chapman KT, Chicchi GG, et al. Combinatorial synthesis of 3-(amidoalkyl) and 3-(aminoalkyl)-2-arylindole derivatives: discovery of potent ligands for a variety of G-protein coupled receptors. *Bioorg Med Chem Lett* 2002;12:93–6.
- [78] Lee J-Y, Kim Y-C. Combinatorial library synthesis and biological evaluation of pyrazolo[4,3-e][1,4]diazepine as a potential privileged structure. *ChemMedChem* 2009;4:733–7.
- [79] Horton DA, Bourne GT, Smythe ML. The combinatorial synthesis of bicyclic privileged structures or privileged substructures. *Chem Rev* 2003;103:893–930.

- [80] DeSimone RW, Currie KS, Mitchell SA, Darrow JW, Pippin D a. Privileged structures: applications in drug discovery. *Comb Chem High Throughput Screen* 2004;7:473–94.
- [81] Costantino L, Barlocco D. Privileged structures as leads in medicinal chemistry. *Curr Med Chem* 2006;13:65–85.
- [82] Duarte CD, Barreiro EJ, Fraga CAM. Privileged structures: a useful concept for the rational design of new lead drug candidates. *Mini Rev Med Chem* 2007;7:1108–19.
- [83] Che Y, Marshall GR. Privileged scaffolds targeting reverse-turn and helix recognition. *Expert Opin Ther Targets* 2008;12:101–14.
- [84] Welsch ME, Snyder SA, Stockwell BR. Privileged scaffolds for library design and drug discovery. *Curr Opin Chem Biol* 2010;14:347–61.
- [85] Nicolaou KC, Roecker AJ, Barluenga S, Pfefferkorn JA, Cao GQ. Discovery of novel antibacterial agents active against methicillin-resistant *Staphylococcus aureus* from combinatorial benzopyran libraries. *Chembiochem* 2001;2:460–5.
- [86] Patchett AA, Nargund RP. Privileged structures — An update. *Annu. Reports Med. Chem.* Vol 35, Elsevier Inc.; 2000, p. 289–98.
- [87] Hajduk PJ, Bures M, Praestgaard J, Fesik SW. Privileged molecules for protein binding identified from NMR-based screening. *J Med Chem* 2000;43:3443–7.
- [88] Jacobson KA, Kim Y ong-C, King BF. In search of selective P2 receptor ligands: interaction of dihydropyridine derivatives at recombinant rat P2X(2) receptors. *J Auton Nerv Syst* 2000;81:152–7.
- [89] Austin JF, MacMillan DWC. Enantioselective organocatalytic indole alkylations. Design of a new and highly effective chiral amine for iminium catalysis. *J Am Chem Soc* 2002;124:1172–3.
- [90] Polanski J, Kurczyk A, Bak A, Musiol R. Privileged structures - dream or reality: preferential organization of azanaphthalene scaffold. *Curr Med Chem* 2012;19:1921–45.
- [91] Gathwala G, Goel M, Singh J, Mittal K. Intravenous diazepam, midazolam and lorazepam in acute seizure control. *Indian J Pediatr* 2012;79:327–32.
- [92] Fukinaga M, Ishizawa K, Kamei C. Anticonvulsant properties of 1,4-benzodiazepine derivatives in amygdaloid-kindled seizures and their chemical structure-related anticonvulsant action. *Pharmacology* 1998;57:233–41.
- [93] Taira N. [Nifedipine: a novel vasodilator]. *Drugs* 2006;66 Spec No:1–3.

- [94] Kuwajima I. [Antihypertensive effects of nifedipine and changing for modified-release formula]. *Drugs* 2006;66 Spec No:7–9.
- [95] Peacock WF, Hilleman DE, Levy PD, Rhoney DH, Varon J. A systematic review of nicardipine vs labetalol for the management of hypertensive crises. *Am J Emerg Med* 2011;981–93.
- [96] Rosenthal J. Nilvadipine: profile of a new calcium antagonist. An overview. *J Cardiovasc Pharmacol* 1994;24 Suppl 2:S92–107.
- [97] Matsuda H, Araki N, Kuji I, Ohkubo T, Imabayashi E, Shimazu K. Effect of nilvadipine on regional cerebral blood flow in a patient with early Alzheimer disease. *Clin Nucl Med* 2008;33:34–5.
- [98] Gradman AH. The evolving role of calcium channel blockers in the treatment of angina pectoris: focus on felodipine. *Can J Cardiol* 1995;11 Suppl B:14B–21B.
- [99] Maksimov ML, Mochkin IA, Starodubtsev AK. [Application of AT1-angiotensin II receptor blocker valsartan in clinical practice]. *Kardiologiya* 2011;51:77–84.
- [100] Effects of telmisartan, irbesartan, valsartan, candesartan, and losartan on cancers in 15 trials enrolling 138,769 individuals. *J Hypertens* 2011;29:623–35.
- [101] Vachhani R, Olds G, Velanovich V. Esomeprazole: a proton pump inhibitor. *Expert Rev Gastroenterol Hepatol* 2009;3:15–27.
- [102] Dorman S, Perkins P. Droperidol for treatment of nausea and vomiting in palliative care patients. *Cochrane Database Syst Rev* 2010:CD006938.
- [103] Cañete R, Escobedo AA, Almirall P, González ME, Brito K, Cimerman S. Mebendazole in parasitic infections other than those caused by soil-transmitted helminths. *Trans R Soc Trop Med Hyg* 2009;103:437–42.
- [104] Murota H, Katayama I. Emedastine difumarate: a review of its potential ameliorating effect for tissue remodeling in allergic diseases. *Expert Opin Pharmacother* 2009;10:1859–67.
- [105] Dowson AJ, Charlesworth B. Review of zolmitriptan and its clinical applications in migraine. *Expert Opin Pharmacother* 2002;3:993–1005.
- [106] Dvornik DM. Tissue selective inhibition of prostaglandin biosynthesis by etodolac. *J Rheumatol Suppl* 1997;47:40–7.
- [107] Foote AP, Harmon DL, Strickland JR, Bush LP, Klotz JL. Effect of ergot alkaloids on contractility of bovine right ruminal artery and vein. *J Anim Sci* 2011;89:2944–9.
- [108] Bemis GW, Murcko MA. The properties of known drugs. 1. Molecular frameworks. *J Med Chem* 1996;39:2887–93.

- [109] Bemis GW, Murcko MA. Properties of Known Drugs. 2. Side Chains. *J Med Chem* 1999;42:5095–9.
- [110] Lewell XQ, Judd DB, Watson SP, Hann MM. 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;38:511–22.
- [111] Gianti E, Sartori L. Identification and selection of “privileged fragments” suitable for primary screening. *J Chem Inf Model* 2008;48:2129–39.
- [112] Schnur DM, Beno BR, Tebben AJ, Cavallaro C. Methods for combinatorial and parallel library design. *Methods Mol Biol* 2011;672:387–434.
- [113] Schnur DM, Hermsmeier MA, Tebben AJ. Are target-family-privileged substructures truly privileged? *J Med Chem* 2006;49:2000–9.
- [114] Bongarzone S, Bolognesi ML. The concept of privileged structures in rational drug design: focus on acridine and quinoline scaffolds in neurodegenerative and protozoan diseases. *Expert Opin Drug Discov* 2011;6:1–18.
- [115] De Sá Alves FR, Barreiro EJ, Fraga CAM. From nature to drug discovery: the indole scaffold as a “privileged structure”. *Mini Rev Med Chem* 2009;9:782–93.
- [116] Chang RS, Lotti VJ, Monaghan RL, Birnbaum J, Stapley EO, Goetz MA, et al. A potent nonpeptide cholecystokinin antagonist selective for peripheral tissues isolated from *Aspergillus alliaceus*. *Science* 1985;230:177–9.
- [117] Bock MG, DiPardo RM, Rittle KE, Evans BE, Freidinger RM, Veber DF, et al. Cholecystokinin antagonists. Synthesis of asperlicin analogues with improved potency and water solubility. *J Med Chem* 1986;29:1941–5.
- [118] Evans BE, Rittle KE, Bock MG, DiPardo RM, Freidinger RM, Whitter WL, et al. Design of nonpeptidal ligands for a peptide receptor: cholecystokinin antagonists. *J Med Chem* 1987;30:1229–39.
- [119] Van der Bent A, Ter Laak AM, IJzerman AP, Soudijn W. Molecular modelling of asperlicin derived cholecystokinin A receptor antagonists. *Eur J Pharmacol* 1992;226:327–34.
- [120] Lattmann E, Billington DC, Poyner DR, Howitt SB, Offel M. Synthesis and evaluation of asperlicin analogues as non-peptidal cholecystokinin-antagonists. *Drug Des Discov* 2001;17:219–30.
- [121] Mcgaughey GB, Gagne M, Rappe AK. -Stacking Interactions 1998;273:15458–63.

- [122] Poulain R, Horvath D, Bonnet B, Eckhoff C, Chapelain B, Bodinier MC, et al. From hit to lead. Analyzing structure-profile relationships. *J Med Chem* 2001;44:3391–401.
- [123] Abrous L, Hynes J, Friedrich SR, Smith AB, Hirschmann R. Design and Synthesis of Novel Scaffolds for Drug Discovery: Hybrids of β -D-Glucose with 1,2,3,4-Tetrahydrobenzo[e][1,4]diazepin-5-one, the Corresponding 1-Oxazepine, and 2- and 4-Pyridyldiazepines. *Org Lett* 2001;3:1089–92.
- [124] Sundberg EJ, Mariuzza RA. Luxury accommodations: the expanding role of structural plasticity in protein-protein interactions. *Structure* 2000;8:R137–42.
- [125] Ma B, Wolfson HJ, Nussinov R. Protein functional epitopes: hot spots, dynamics and combinatorial libraries. *Curr Opin Struct Biol* 2001;11:364–9.
- [126] MacCallum RM, Martin AC, Thornton JM. Antibody-antigen interactions: contact analysis and binding site topography. *J Mol Biol* 1996;262:732–45.
- [127] Al-Lazikani B, Leskela M, Chothia C. Standard conformations for the canonical structures of immunoglobulins. *J Mol Biol* 1997;273:927–48.
- [128] Hare S, Gupta SS, Valkov E, Engelman A, Cherepanov P. Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* 2010;464:232–6.
- [129] Li X, Krishnan L, Cherepanov P, Engelman A. Structural biology of retroviral DNA integration. *Virology* 2011;411:194–205.
- [130] Hare S, Vos AM, Clayton RF, Thuring JW, Cummings MD, Cherepanov P. Molecular mechanisms of retroviral integrase inhibition and the evolution of viral resistance. *Proc Natl Acad Sci U S A* 2010;107:20057–62.
- [131] Maertens GN, Hare S, Cherepanov P. The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* 2010;468:326–9.
- [132] Petersen U. Quinolone Antibiotics: The Development of Moxifloxacin. In: Fischer J, Ganellin CR, editors. *Analog. Drug Discov.*, Weinheim, FRG: Wiley-VCH Verlag GmbH & Co. KGaA; 2006, p. 315–70.
- [133] Tekwani BL, Walker LA. 8-Aminoquinolines: future role as antiprotozoal drugs. *Curr Opin Infect Dis* 2006;19:623–31.
- [134] Musiol R, Serda M, Hensel-Bielowka S, Polanski J. Quinoline-based antifungals. *Curr Med Chem* 2010;17:1960–73.
- [135] Font M, Monge A, Alvarez E, Cuartero A, Losa MJ, Fidalgo MJ, et al. Synthesis and evaluation of new Reissert analogs as HIV-1 reverse transcriptase inhibitors. 1. Quinoline and quinoxaline derivatives. *Drug Des Discov* 1997;14:305–32.

- [136] Knight SD, Adams ND, Burgess JL, Chaudhari AM, Darcy MG, Donatelli C a., et al. Discovery of GSK2126458, a Highly Potent Inhibitor of PI3K and the Mammalian Target of Rapamycin. *ACS Med Chem Lett* 2010;1:39–43.
- [137] Mekouar K, Mouscadet JF, Desmaële D, Subra F, Leh H, Savouré D, et al. Styrylquinoline derivatives: a new class of potent HIV-1 integrase inhibitors that block HIV-1 replication in CEM cells. *J Med Chem* 1998;41:2846–57.
- [138] Zouhiri F, Mouscadet JF, Mekouar K, Desmaële D, Savouré D, Leh H, et al. Structure-activity relationships and binding mode of styrylquinolines as potent inhibitors of HIV-1 integrase and replication of HIV-1 in cell culture. *J Med Chem* 2000;43:1533–40.
- [139] Bonnenfant S, Thomas CM, Vita C, Deprez E, Zouhiri F, Desmae D, et al. Styrylquinolines , Integrase Inhibitors Acting Prior to Integration : a New Mechanism of Action for Anti-Integrase Agents 2004;78:5728–36.
- [140] Firley D, Courcot B, Gillet J-M, Fraisse B, Zouhiri F, Desmaële D, et al. Experimental/theoretical electrostatic properties of a styrylquinoline-type HIV-1 integrase inhibitor and its progenitors. *J Phys Chem B* 2006;110:537–47.
- [141] Cotel P. Patented HIV-1 integrase inhibitors (1998-2005). *Recent Pat Antiinfect Drug Discov* 2006;1:1–15.
- [142] Polanski J, Niedbala H, Musiol R, Podeszwa B, Tabak D, Palka A, et al. 5-Hydroxy-6-Quinaldic Acid as a Novel Molecular Scaffold for HIV-1 Integrase Inhibitors. *Lett Drug Des Discov* 2006;3:175–8.
- [143] Polanski J, Niedbala H, Musiol R, Podeszwa B, Tabak D, Palka a., et al. Fragment Based Approach for the Investigation of HIV-1 Integrase Inhibition. *Lett Drug Des Discov* 2007;4:99–105.
- [144] Boros EE, Edwards CE, Foster S a, Fuji M, Fujiwara T, Garvey EP, et al. Synthesis and antiviral activity of 7-benzyl-4-hydroxy-1,5-naphthyridin-2(1H)-one HIV integrase inhibitors. *J Med Chem* 2009;52:2754–61.
- [145] Musiol R. Quinoline-based HIV integrase inhibitors. *Curr Pharm Des* 2013;19:1835–49.
- [146] Guare JP, Wai JS, Gomez RP, Anthony NJ, Jolly SM, Cortes AR, et al. A series of 5-aminosubstituted 4-fluorobenzyl-8-hydroxy-[1,6]naphthyridine-7-carboxamide HIV-1 integrase inhibitors. *Bioorg Med Chem Lett* 2006;16:2900–4.
- [147] Embrey MW, Wai JS, Funk TW, Hornick CF, Perlow DS, Young SD, et al. A series of 5-(5,6)-dihydrouracil substituted 8-hydroxy-[1,6]naphthyridine-7-carboxylic acid 4-fluorobenzylamide inhibitors of HIV-1 integrase and viral replication in cells. *Bioorg Med Chem Lett* 2005;15:4550–4.

- [148] Verschueren WG, Dierynck I, Amssoms KIE, Hu L, Boonants PMJG, Pille GME, et al. Design and optimization of tricyclic phthalimide analogues as novel inhibitors of HIV-1 integrase. *J Med Chem* 2005;48:1930–40.
- [149] Metobo SE, Jin H, Tsiang M, Kim CU. Design, synthesis, and biological evaluation of novel tricyclic HIV-1 integrase inhibitors by modification of its pyridine ring. *Bioorg Med Chem Lett* 2006;16:3985–8.
- [150] Baber JC, Feher M. Predicting synthetic accessibility: application in drug discovery and development. *Mini Rev Med Chem* 2004;4:681–92.
- [151] Ertl P, Schuffenhauer A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J Cheminform* 2009;1:8.
- [152] Boda K, Seidel T, Gasteiger J. Structure and reaction based evaluation of synthetic accessibility. *J Comput Aided Mol Des* 2007;21:311–25.
- [153] Lajiness MS, Maggiora GM, Shanmugasundaram V. Assessment of the consistency of medicinal chemists in reviewing sets of compounds. *J Med Chem* 2004;47:4891–6.
- [154] Takaoka Y, Endo Y, Yamanobe S, Kakinuma H, Okubo T, Shimazaki Y, et al. Development of a method for evaluating drug-likeness and ease of synthesis using a data set in which compounds are assigned scores based on chemists' intuition. *J Chem Inf Comput Sci* 2003;43:1269–75.
- [155] Gillet VJ, Myatt G, Zsoldos Z, Johnson AP. SPROUT, HIPPO and CAESA: Tools for de novo structure generation and estimation of synthetic accessibility. *Perspect Drug Discov Des* 1995;3:34–50.
- [156] Ihlenfeldt W-D, Gasteiger J. Computer-Assisted Planning of Organic Syntheses: The Second Generation of Programs. *Angew Chemie Int Ed English* 1996;34:2613–33.
- [157] Huang N, Shoichet BK, Irwin JJ. Benchmarking sets for molecular docking. *J Med Chem* 2006;49:6789–801.
- [158] Bender A, Mussa HY, Glen RC, Reiling S. Molecular similarity searching using atom environments, information-based feature selection, and a naïve Bayesian classifier. *J Chem Inf Comput Sci* 2004;44:170–8.
- [159] Bender A, Mussa HY, Glen RC, Reiling S. Similarity searching of chemical databases using atom environment descriptors (MOLPRINT 2D): evaluation of performance. *J Chem Inf Comput Sci* 2004;44:1708–18.
- [160] Klekota J, Roth FP. Chemical substructures that enrich for biological activity. *Bioinformatics* 2008;24:2518–25.

- [161] Yap CW. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J Comput Chem* 2011;32:1466–74.
- [162] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software. *ACM SIGKDD Explor Newsl* 2009;11:10.
- [163] John GH, Pat L. Estimating Continuous Distributions in Bayesian Classifiers. *Proc. Elev. Conf. Uncertain. Artif. Intell.*, San Mateo: Morgan Kaufmann Publishers; 1995.
- [164] Keerthi SS, Shevade SK, Bhattacharyya C, Murthy KRK. Improvements to Platt's SMO Algorithm for SVM Classifier Design. *Neural Comput* 2001;13:637–49.
- [165] Hastie BT, Tibshirani R. Classification by Pairwise Coupling. *Ann Stat* 1998;26:451–71.
- [166] Platt JC. Fast Training of Support Vector Machines using Sequential Minimal Optimization. In: Schölkopf B, Burges C, Smola A, editors. *Adv. Kernel Methods - Support Vector Learn.*, Cambridge, Mass.: MIT Press; 1999, p. 185–208.
- [167] Aha DW, Kibler D, Albert MK. Instance-based learning algorithms. *Mach Learn* 1991;6:37–66.
- [168] Melville P, Mooney RJ. Constructing Diverse Classifier Ensembles using Artificial Training Examples. *Proc. Eighteenth Int. Jt. Conf. Artif. Intell.*, Acapulco, Mexico: 2003, p. 505–10.
- [169] Melville P, Mooney RJ. Creating Diversity In Ensembles Using Artificial Data. *Inf Fusion* 2004;6:99–111.
- [170] Witten IH, Frank E. *Data Mining: Practical Machine Learning Tools and Techniques*. second. Morgan Kaufmann Publishers; 2005.
- [171] Frank E, Witten IH. Generating Accurate Rule Sets Without Global Optimization. *ICML '98 Proc. Fifteenth Int. Conf. Mach. Learn.*, Morgan Kaufmann Publishers; 1998, p. 144–51.
- [172] Breiman L. Random Forests. *Mach Learn* 2001;45:5–32.
- [173] Wassermann AM, Nisius B, Vogt M, Bajorath J. Identification of descriptors capturing compound class-specific features by mutual information analysis. *J Chem Inf Model* 2010;50:1935–40.
- [174] Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett* 2006;27:861–74.

9 ZAŁĄCZNIKI

Do treści niniejszej rozprawy dołączono dwa załączniki: (i) życiorys naukowy autora (**Załącznik 1.**) oraz (ii) najważniejsze publikacje w dotychczasowym dorobku naukowym autora (**Załącznik 2.**). Drugi z wymienionych załączników obejmuje kserokopie dwóch publikacji z dorobku naukowego autora:

Załącznik 2a. Kserokopia publikacji naukowej:

Polanski J., Kurczyk A., Bak A., Musiol R.; Privileged structures – dream or reality: preferential organization of azanaphthalene scaffold., *Curr. Med. Chem.*, 2012, 19(13), 1921-1945. (IF₂₀₁₁ = 4,859)

Załącznik 2b. Kserokopia publikacji naukowej:

Bak A., Magdziarz T., Kurczyk A., Serafin K., Polanski J.; Probing a Chemical Space for Fragmental Topology-Activity Landscapes (FRAGTAL): Application for Diketo Acid and Catechol HIV Integrase Inhibitor Offspring Fragments., *Comb. Chem. High Throughput Screen.*, 2013, 16(4), 274-287. (IF₂₀₁₂ = 2,000)

Załącznik 1. Życiorys naukowy autora

Curriculum Vitae

Dane osobowe

Imię i nazwisko	Agata Kurczyk
Data urodzenia	5 stycznia 1984 r.
Miejsce urodzenia	Bytom

Wykształcenie

2008-2013	Studia doktoranckie w zakresie chemii przy Instytucie Fizyki Uniwersytetu Śląskiego w Katowicach.
2003-2008	Studia stacjonarne magisterskie na kierunku chemia w zakresie chemii ogólnej na Wydziale Matematyki, Fizyki i Chemii Uniwersytetu Śląskiego w Katowicach; zdobycie tytułu zawodowego: magister.
1999-2003	Zdobycie wykształcenia średniego ogólnego. (Liceum Ogólnokształcące Nr IV w Bytomiu)
1991-1999	Szkoła Podstawowa Nr 37 w Bytomiu

Kwalifikacje oraz uprawnienia zawodowe

- Kwalifikacje pedagogiczne wymagane od nauczycieli zgodnie z § 1 pkt. 3 rozporządzenia Ministra Edukacji Narodowej i Sportu z 10 września 2002 r. Dz.U. z 2003 r. Nr 155 poz. 1288.

Działalność dydaktyczna

Prowadzone zajęcia dydaktyczne w Zakładzie Chemii Organicznej Instytutu Chemii Wydziału Matematyki, Fizyki i Chemii Uniwersytetu Śl. w Katowicach:

- laboratorium chemii organicznej;
- chemoinformatyka;
- informacja naukowa.

Udział w projektach

- Opracowanie multimedialnych materiałów dydaktycznych platformy e-learningowej „E-chem” Uniwersytetu Śląskiego w Katowicach (Projekt UPGOW - Uniwersytet Partnerem Gospodarki Opartej na Wiedzy współfinansowany przez Unię Europejską w ramach Europejskiego Funduszu Społecznego).

Zainteresowania naukowe i badawcze

Modelowanie molekularne, polifarmakologia, metody odkrywania wiedzy w bazach danych.

Udział w konferencjach

- 16-17. 11. 2012 XV Gliwickie Spotkania Naukowe, Gliwice, Polska
poster: Virtual screening of HIV integrase inhibitors and azanaphthalene compounds in a search for privileged substructural motifs; **Kurczyk A.**, Bak A., Magdziarz T., Polanski J.
- 10-14. 10. 2012 10th International Congress of Young Chemists "YoungChem 2012", Gdańsk, Polska
poster: Mining molecular and structural chemical databases for the exploration of the architecture of drugs; **Kurczyk A.**, Bak A., Magdziarz T., Polanski J.
- 13-15. 09. 2012 V Konwersatorium Chemii Medycznej, Lublin, Polska
poster: Development of multistep ligand-based virtual screening cascade methodology in a search for novel HIV-1 integrase inhibitors: 1. Machine learning; **Kurczyk A.**, Warszky D., Kafel R., Musiol R., Bojarski A., Polanski J.
- 18-19. 11. 2011 r. XIV Gliwickie Spotkania Naukowe, Gliwice, Polska
poster: Probing pharmacological space for the analysis of fragmental drug-likeness topology: Application to mono- and diazanaphthalene compounds; **Kurczyk A.**, Bak A., Magdziarz T., Polanski J.
- 4-5. 10. 2011 r. The Modern Techniques for Drug Design Purposes, Warszawa, Polska
poster: Drug-likeness topology determination using fragment-based approaches in molecular and structural chemical databases; **Kurczyk A.**, Bak A., Magdziarz T., Polanski J.
- 18-22. 09. 2011 r. 54 Zjazd Polskiego Towarzystwa Chemicznego i Stowarzyszenia Inżynierów i Techników Przemysłu Chemicznego, Lublin, Polska
poster: Drug-likeness topology determination using fragment-based approaches in molecular and structural chemical databases; **Kurczyk A.**, Bak A., Magdziarz T., Polanski J.
- 12-14. 09. 2011 r. 40th Conference. Drug Synthesis and Analysis, Brno, Czechy
poster: Screening chemical space in a search for fragmental druglikeness topology: Application for HIV integrase inhibitor offspring fragments and azanaphthalene polypharmacology; **Kurczyk A.**, Bak A., Magdziarz T., Polanski J.

- 8-10. 09. 2011 r. IV Konwersatorium Chemii Medycznej, Lublin, Polska
poster: Probing pharmacological space for the analysis of fragmental drug-likeness topology: Application to mono- and diazanaphthalene compounds; **Kurczyk A.**, Bak A., Magdziarz T., Polanski J.
- 26-27. 11. 2010 r. XIV Gliwickie Spotkania Naukowe, Gliwice, Polska
poster: Drug Architecture by Database Mining: Global Mapping of Pharmacological Space; **Kurczyk A.**, Magdziarz T., Bak A., Polanski J.
- 24-27. 04. 2010 r. VIII Seminarium Doktorantów „Na pograniczu biologii i chemii”, Duszniki Zdrój, Polska
komunikat: Identification of privileged molecular fragments using global mapping of pharmacological space: Application to mono- and diazanaphthalene scaffolds; **Kurczyk A.**
- 20-21. 11. 2009 r. XIII Gliwickie Spotkania Naukowe, Gliwice, Polska
poster: The use of MoStBioDat for rapid screening of molecular diversity; **Kurczyk A.**, Bąk A., Magdziarz T.

Staże naukowe

- 18.03.2012-14.04.2012 – miesięczny staż naukowy, Charles University in Prague, Faculty of Pharmacy in Hradec Kralove;
- 05.02.2012-03.03.2012 – miesięczny staż naukowy, Instytut Farmakologii Polskiej Akademii Nauk w Krakowie.

Prace naukowe

Współautor rozdziałów książek:

1. Musiol R., Magdziarz T., **Kurczyk A.**; Quinoline scaffold as a privileged substructure in antimicrobial drugs. W: A. Méndez-Vilas (red.) Science against microbial pathogens: communicating current research and technological advances, Vol. 1., *Formatex Research Center*, 2011, s. 72-83.
2. **Kurczyk A.**, Bąk A., Magdziarz T., Mazur P., Polański J.; Polifarmakologia i lekotypia związków azanaftalenowych. W: H. Koroniak, J. Barciszewski (red.) Na pograniczu chemii i biologii., t. XXV, *Wydawnictwo Uniwersytetu A. Mickiewicza w Poznaniu*, 2010, s. 161-170.
3. Bąk A., **Kurczyk A.**; Projektowanie molekularne – notacja liniowa cząsteczek, kodowanie SMILES, wybrane edytory molekularne. W: Bąk A. (red.) Laboratorium projektowania molekularnego. Materiały do ćwiczeń.; Katowice 2010.

4. Bąk A., **Kurczyk A.**; Projektowanie molekularne – standardy wymiany informacji strukturalnej, pakiety do modelowania molekularnego. W: Bąk A. (red.) Laboratorium projektowania molekularnego. Materiały do ćwiczeń.; Katowice 2010.

Publikacje:

1. Bak A., Magdziarz T., **Kurczyk A.**, Serafin K., Polanski J.; Probing a Chemical Space for Fragmental Topology-Activity Landscapes (FRAGTAL): Application for Diketo Acid and Catechol HIV Integrase Inhibitor Offspring Fragments., *Comb. Chem. High Throughput Screen.*, 2013, 16(4), 274-287. (IF₂₀₁₂ = 2,000)
2. Cieslik W., Serda M., **Kurczyk A.**, Musiol R.; Microwave Assisted Synthesis of Monoazanaphthalene Scaffolds., *Curr. Org. Chem.*, 2013, 2013, 17(5), 491–503. (IF₂₀₁₂ = 3,039)
3. Polanski J., **Kurczyk A.**, Bak A., Musiol R.; Privileged structures – dream or reality: preferential organization of azanaphthalene scaffold., *Curr. Med. Chem.*, 2012, 19(13), 1921-1945. (IF₂₀₁₁ = 4,859)
4. Bak A., Magdziarz T., **Kurczyk A.**, Polanski J.; Mapping Fragmental Drug-likeness in the MoStBioDat Environment: Intramolecular Hydrogen Bonding Motifs in β -Ketoenols., *Comb. Chem. High Throughput Screen.*, 2011, 14(7), 560-569. (IF₂₀₁₀ = 2,573)
5. Bak A., Magdziarz T., **Kurczyk A.**; Polanski J., Mapping Drug Architecture by MoStBioDat: Rapid Screening of Intramolecular Hydrogen Bonded Motifs in Catechols., *Drug Dev. Res.*, 2011, 72(2), 209-218. (IF₂₀₁₀ = 1,109)
6. Bąk A., Polański J., Stockner T., **Kurczyk A.**; MoStBioDat – Molecular and Structural Bioinformatics Database., *Comb. Chem. High Throughput Screen.*, 2010, 13(4), 366-374. (IF₂₀₀₉ = 2,464)
7. Bąk A., Polański J., **Kurczyk A.**; The use of MoStBioDat for rapid screening of molecular diversity., *Molecules*, 2009, 14(9), 3436-3445. (IF₂₀₀₈ = 1,252)

Załącznik 2a. Kserokopia publikacji naukowej:

Polanski J., Kurczyk A., Bak A., Musiol R.; Privileged structures – dream or reality: preferential organization of azanaphthalene scaffold., *Curr. Med. Chem.*, 2012, 19(13), 1921-1945. (IF₂₀₁₁ = 4,859)

Privileged Structures – Dream or Reality: Preferential Organization of Azanaphthalene Scaffold

J. Polanski*, A. Kurczyk, A. Bak and R. Musiol

Institute of Chemistry, University of Silesia, Szkolna 9, 40-006 Katowice, Poland

Abstract: The concept of privileged structures/substructures (PS) is the idea that certain structural features produce biological effects more often than others. The PS method can be seen as an offspring of fragonomics, which is based on recent experimental measurements of protein-ligand interactions. If PS prove to be true, then chemical motives that enrich biological activity can be used when designing new drugs. However, PS remain controversial because we cannot be sure whether the excess of active structures does not result from an abundance in chemical libraries. In this review, we will focus, in particular, on the preferential organization of azanaphthalene scaffolds (AN) in drugs and natural products (NP), which are preferred by Nature in evolution. We will show that knowledge discovery in molecular databases can reveal interesting time-trends profiles for important classes of potentially privileged scaffolds. The chemical library of AN is dominated by monoaza-compounds, among which quinoline appears to be the most frequently investigated scaffold; however, more sophisticated database mining seems to indicate different PS patterns within the AN scaffold family.

Keywords: Azanaphthalene scaffold, chemical spaces, database mining, fragment-based design, privileged motifs.

INTRODUCTION

The concept of privileged structures/substructures (PS) is the idea that certain structural features produce biological effects more often than others [1]. More formally, PS have been defined in the context of a minimal substructure capable of binding several types of receptors [2]. If PS prove to be true, then chemical motives that enrich biological activity can be used when designing new drugs [3]. However, PS remain controversial because we cannot be sure whether the excess of active structures does not result from an abundance in chemical libraries. For example, if PS in pioneering Ehrlich or Domagk chemotherapeutics (salvarsan, prontosil; Fig. (1)) are analyzed, an azo-dye related structure will be indicated as a privileged one; however, this could simply have resulted from the availability of dyes in XIX (or early XX) century chemistry, which determined that Ehrlich was fascinated by the fact that *aniline* or *other newly available dyes* could stain microbial tissues during his studies [4]. In fact, salvarsan, at least in its structure as presented in textbooks, resembles azo dyes very closely. However, it has just been reported that the actual salvarsan structure looks completely different [5]. According to another story, prontosil, which was discovered by Domagk, appeared to be a sulfonamide prodrug [6] of the azo dye family. Thus, both stories reveal the presence of serendipitous elements.

The above-mentioned examples indicate that privileged motifs that are identified are highly dependent on the compound space being screened. Thus, the concept of a chemical space (Fig. (2)) allows us to better control and organize chemical structures, e.g. the concept of organic chemistry architecture [7, 8], suggests how we can analyze and organize the information in molecular databases in order to find a *preferential chemistry landscape* Fig. (3).

The idea of *preferential chemistry* resembles this of *drug-likeness*. In turn, *drug-likeness* and *PS* concepts refer to similar problems. What are the differences between them? What we actually mean by *drug-likeness* in molecular design and do drug targets also have a likeness [9]? On the other hand, identifying a *preferential landscape* is related to molecular fragments, substructures, superstructures, scaffolds, linkers (compare Table 1); as well as their generation, identification, statistical analyses of frequencies etc. Intuitively, fragments, substructures or privilege (sub)structures refer to some building blocks of the molecule. Preferences may include real molecules but also virtual structures or molecular motifs. What are the relationships between them? The

aim of this publication was to review basic concepts essential for the PS identification and to compare this method to similar approaches in molecular design.

Another aim of this review was to compare the way in which *preferential chemistry* problems are addressed by computational chemists and those working in synthetic laboratories. Therefore, we additionally focused on the azanaphthalene scaffold, which can often be found in biologically active molecules. Can this observation be converted into a numeric form that supports the concepts of *likeness* and *privilegedness*? We screened molecular databases in an attempt to identify and illustrate the preferential organization of this scaffold. Since natural compounds are structures isolated from nature and thus clearly preferred and strongly influencing drug design [21], we concentrated on such compounds.

DIFFERENTIATING DRUGS FROM NONDRUGS

Various methods have been developed in molecular design in an attempt to find new drugs. Optimistic approaches, e.g. Quantitative Structure Activity Relationships (QSARs) were aimed at modeling/predicting a certain activity value for a new compound. In more pragmatic methods, we are satisfied when we can even postulate a lack or gain in activity, or even the probability of that. From among the optimistic methods, modeling QSAR appears to be especially efficient, provided that we proceed within a congeneric compound series. This works even better where membrane penetration is an effect limiting an activity value. If this is the case, then a simple Hansch approach provides both an efficient mathematical model and a clear mechanistic explanation. In multidimensional m-QSAR terminology, the Hansch approach can be classified as 0 or 1D QSAR. The increase of dimensionality m into (3-7)-QSAR protocols appears to be much less effective than expected [44-46].

Qualitatively differentiating drugs from non-drugs can be seen as a simplified, true or false (0 or 1) QSAR version, which is also not a minor problem. This can be approached in two ways. The first step is to choose the properties necessary to exhibit drug-like behavior. Then we can try to predict such properties for the structures under consideration in attempting to find drug candidates. This approach more or less mimics the QSAR protocols. Although some properties that determine the drug development ability of the molecular structure seem clear, e.g. toxicity, bioavailability or synthetic accessibility, choosing properties or their modeling is a rather complicated process. In approaches that are more pragmatic, we simply estimate the similarity of the designed structures to real drugs [25, 47]. This method still attracts attention, e.g. molecular

*Address correspondence to this author at the Institute of Chemistry, University of Silesia, Szkolna 9, 40-006 Katowice, Poland. Tel-Fax: 0048/32/2599978; E-mail: polanski@us.edu.pl

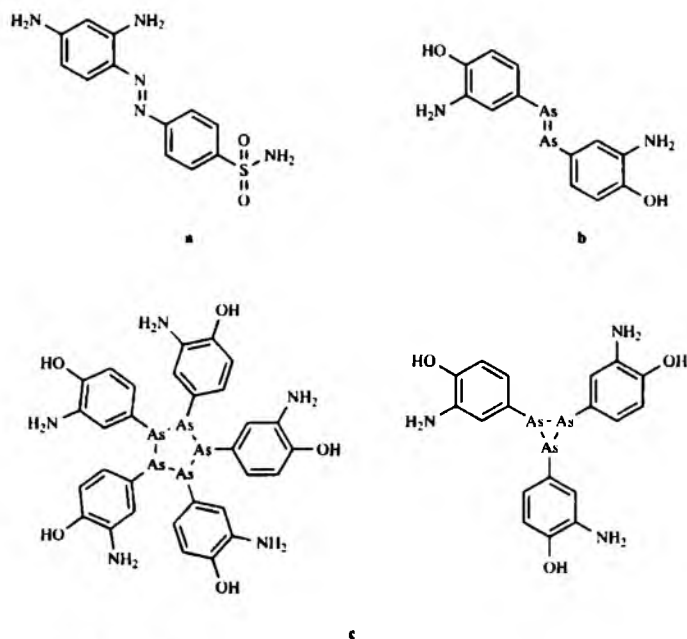


Fig. (1). Structures of prontosil and salvarsan. Prontosil – the first sulfonamide prodrug belonging to azo dyes (a). The original structure of salvarsan proposed by Ehrlich reflecting analogy with azo compounds (b). The actual salvarsan structure consisting of cyclic species being a mixture of cyclic trimers and pentamers [5] (c).

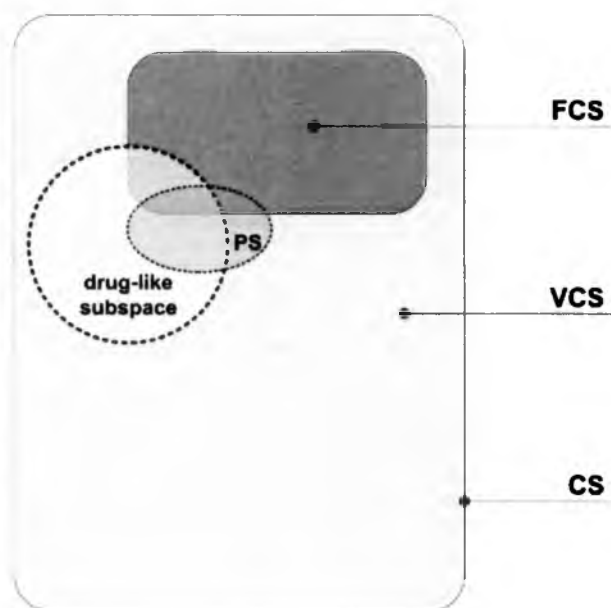


Fig. (2). Chemical space (CS) consisting of factual (FCS) and virtual (VCS) parts. FCS and VCS refer to all currently known and all possible molecules respectively. The CS representation may vary with the particular set of descriptors and properties used to define the subspace where the molecules will be placed, e. g., drug-like and PS subspaces.

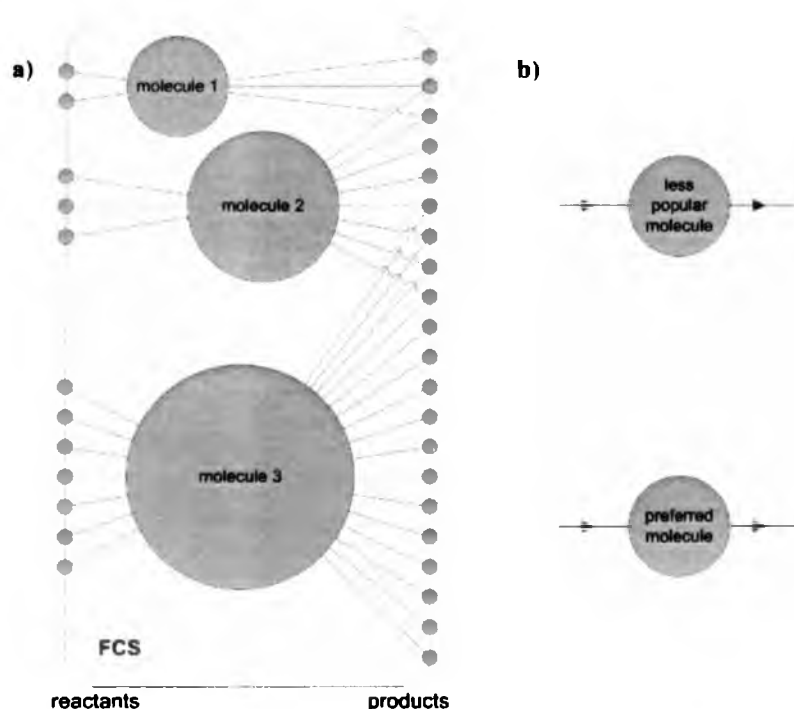


Fig. (3). Organic chemistry landscape. Representation of the preferential attachment responsible for time-evolving chemistry's network. Molecules correspond to nodes and reactions to edges connecting these nodes (a). Preferred molecules produce more connections thereby becoming even more popular in time; whereas less popular molecules form new relations less rapidly over the same period (b); thus, network evolution has been described as quite similar to that characterizing the World Wide Web [7, 8].

Table 1. Privileged Structure Related Concepts and Definitions

	Definitions and/or rules	References
Structure	- Topological graph which solely describe the linkage of the atoms at which every atoms is a vertex and every bond is an edge in the molecular graph. A molecule is segmented into four units: a. ring systems – cycles within the molecular graph or rings sharing a bond between atoms. b. linkers – atoms and/or bonds on the path connecting two different ring systems. c. side-chains – atoms not classified as ring system or linkers. d. framework – ring systems connected by linkers. e. 'complex frameworks' – framework which specifies atom types and bond orders	[10-12]
Scaffold	- Framework which exocyclic double bond (double bonded single atom side-chains mostly X=O) are not removed from rings. - The structural core upon which different types of substituents can be attached. - Common core structure characterizing the group of molecules that may share a common synthetic pathway	[10] [13] [14]
Privileged structure	- Scaffold which were capable of providing useful ligands for diverse receptors and judicious modifications of such structures could be a viable alternative for design a new receptor agonist and antagonists. - Scaffolds that very often produce biologically active analogues in a target family	[15] [16]
Frequent hitters	Compounds which show up as hits in many different biological assays covering a wide range of targets.	[17]
Chemical space	- Space can be viewed as being analogous to the cosmological universe in its vastness with chemical compounds populating space instead of stars - The total descriptor space that encompasses all the small carbon-based molecules that could be in principle created - The set of the all possible molecular structures (10^{16}-10^{18} with 10^6 of compounds congaing up to 30 C, N, O, S atoms) - Consists of factual and virtual subspaces where each molecule can be given a vector coding molecular structure which can be calculated from the molecular structure for the all molecules in the whole chemical space and compound property which basically should be measured.	[18] [19] [20] [21]

(Table I). Contd.....

	Definitions and/or rules	References
Factual Chemical Space	It is estimated that up to 10^{48} structures can be synthesized with potentially 10^{24} small molecules of less than 30 non-hydrogen atoms and up to 10^{16} pharmacologically relevant compounds.	[22]
Virtual Chemical Space	Theoretically possible collections of molecules	[21]
Drug Fragment Space	Small number of compounds based on scaffolds and functional groups commonly found in known drugs	[23]
Drug-Like Property Space	Composed of compounds which fulfill rule of five.	[24]
Drug-Like Chemical Space	Defined by fingerprint discrimination model	[25]
Drug-likeness	Lipinski rule of five The set of threshold values was imposed on the specific molecular descriptors including the octanol/water partition coefficient (ClogP ≤ 5), the molecular weight (MW ≤ 500), the number of hydrogen bond donors (HBD ≤ 5) and hydrogen bond acceptors (HBA ≤ 10)	[26]
	Veber rule Number of rotatable bonds (NRB) ≤ 10 , tPSA $\leq 140\text{\AA}^2$, HBA ≤ 12	[27]
	Drug-likeness for compound affinity - deviation of the docking scores with multiple protein.	[28]
	Drug-target-likeness - protein targets of drugs share characteristics in their sequence	[9]
	Model-Free drug-likeness - rapid elimination those of prospective drugs which are rich in nondrug-like fragments.	[29]
Lead-likeness	The computational criteria applicable in massive library screening for lead identification: MW ≤ 460 , -4 $\leq$ logP ≤ 4.2 or -4 $\leq$ logD ≤ 4 where logD $_7$ is logP value at pH=7.4, logS ≥ -5 , NRB ≤ 10 , HBD ≤ 5 , HBA ≤ 9 and number of rings RNC ≤ 4	[30]
Fragment-likeness	Congreve rule of three The rule restricts the boundary limits to MW ≤ 300 , logP ≤ 3 , NRB ≤ 3 , HBD ≤ 3 , HBA ≤ 3 and PSA $\leq 60\text{\AA}^2$	[31]
	Reduced complexity The paradigm in fragment library for lead generation have been proposed where MW ≤ 350 , NRB ≤ 6 , HBD ≤ 3 , HBA ≤ 8 , logP ≤ 2.2 and heavy atoms NHA ≤ 22 .	[30]
	Traffic Lights The promising set of rules where MW ≤ 400 , ClogP ≤ 3 , water solubility at pH 6.5 $\geq 50\text{mg L}^{-1}$, tPSA $\leq 120\text{\AA}^2$, NRB ≤ 7	[32]
Substructure	A subgraph of the graph that is associated with a chemical structure and is correspondingly labeled (or colored) in a manner that reflects the nature of the atoms and bonds comprised in the original parental molecule.	[22]
Superstructure	Given two chemical structures A and B, if structure A is fully contained in structure B, then A is a substructure of B, while B is a superstructure of A.	[33]
Substructure molecular descriptors (examples)	Molecular access systems (MACCS) - fragment dictionary based on the atom types, atom pairs and custom atom environments is encoded in bit vector.	[22]
	Structural (D)2D/3D descriptors - topological descriptors, physical property descriptors, atom and bond counts, surface area descriptors, charge descriptors.	[34]
	Vectorial descriptors collecting counts of occurrences of predefined structural features in molecule or binary variables specifying their presence/absence. Divided into topological representations (2D) or (3D) representations encoding the spatial relationship.	[35]
	Structural key descriptors e. g., DAIM fingerprints.	[36]
Substructure screening	The usage of the predefined list or dictionary of fragments to check for the presence or absence of each substructure in a set of active molecules	[22]
Maximum common substructure	A single contiguously connected common subgraph of a molecular structure present in a specified fraction of all molecules.	[37]
Fragment	Small molecules of less than approximately 250 Da and typically composed of just a few functionalities.	[38]
	Small molecule with a size typically less than 300Da which is expected to bind to its target protein with weaker activity than traditional drug-like molecule but with binding efficiency per atom at least as high as for target molecule	[39]
	Represents linear and/or branched substructure of 6-40 atoms. An exact fragment is a set of atoms and their connections. A generic fragment is a fuzzy substructure in which atoms and/or bonds are denoted by symbols representing lists of possible atoms, bonds or even structural properties such as 'H-bond acceptor' or the like.	[22]
	A set of atoms connected by unbreakable bonds e.g. terminal, double, triple or aromatic bonds in rings.	[36]
	Discriminative fragments - parts of molecule that are frequent in the set of active molecules and rare in the set of inactive.	[40]
Chemical cliches	The most often occurring fragments and frequently cooccurring pairs of fragments.	[41]
Structure fragmentation procedure	Fragment dictionary - predefined set of small molecules.	[22]
	Recursive partitioning - reiterative process whereby a set of structures is partitioned according to presence or absence of a discriminating fragment contained in an exhaustive list of atom pairs and topological torsions.	[22]
	Discrete fragmental analysis - fragment is scored according to its probability of chance occurrence in the dataset at hand.	[22]
	Decoration spheres - hits was decomposed resulting in the subset of substructures ordered according to a increasing level of complexity and belonging to distinct 'decoration spheres' around the main core present in the parental molecule.	[42]
	Graph splitting technique - breaks the molecule at topologically interesting points such as bond between a substituent and a ring.	[41]
	Decomposition and identification of molecules (DAIM).	[36]
	Cleavage rules - identifies fragments which can be synthesized easily from known chemistries	[43]

topology descriptors have been used recently to identify the differences between drugs, clinical candidate and bioactive molecules [48].

Drug-Likeness

Drug-likeness is a concept that originated from the idea that drug compounds differentiate from other molecules in their physical properties. Oprea performed a large database mining within the MACCS-II Drug Data Report (MDDR), Current Patents Fast-alert, Comprehensive Medicinal Chemistry, Physician Desk Reference, New Chemical Entities and the Available Chemical Directory (ACD) in an effort to identify drug-like structures and their property distributions [49]. Several approaches were developed in an attempt to identify the differences between drug- and non-drug-like molecules (chemicals). The *a priori* calculation of molecular descriptors, critical for the compound bioavailability and consequently fundamental for any potential drug candidate, is required to make predictions [18, 26]. The ADMET molecular properties absolutely necessary for drug bioavailability, are estimated on the basis of the molecular structures even before the synthesis of the molecules has been rationalized. This can be seen as the absolute prerequisite for the compound to survive through the expected clinical trials. A variety of modifications have resulted from the Lipinski drug-likeness strategy [28, 29].

Molecular descriptors calculated in order to quantify drug-like properties (DP) are easily available from molecular formulae, e.g. the number of hydrogen bond donors (HBD) or acceptors (HBA). In practice, even those that refer to molecular properties can be estimated (predicted) for virtual molecules by calculation today, e.g. log P. The reliability of the resulting DP values is one issue in this approach. The second is that DPs are usually extensive type parameters describing an entire molecule. Thus, they do not permit to discriminate easily and efficiently between diversified molecular structures. Drug-likeness appears to be an attractive idea in *privileged chemistry*; however, this concept is elusive and ambiguous [34, 50]. Thus, with a better understanding and the larger availability of databases equipped with chemoinformatic tools, we realized that a large portion of molecules satisfy the DP rules, e.g. 80% of the chemicals recorded in the ACD bases satisfy the Lipinski rule (Ro5) [50], while the distributions of the molecular weights (one of the Ro5 parameters) of drugs vs. other chemicals were found to be convergent [7].

Although drug-likeness has gained wide acceptance as an approach to reduce attrition in drug discovery and development, the analysis of recent trends has revealed that the actual properties of compounds being synthesized in the pharmaceutical industry differ significantly from those of recently discovered oral drugs and compounds in clinical development. One reason may be that the increase in lipophilicity (the prominent DP) enlarge the likelihood of a decrease in drug selectivity and further the probability of success in drug development [51]. The drug-likeness approach inspired a series of related concepts of *lead-likeness* [52-58], *protein druggability* [2] and *drug-target-likeness* [9], which have attracted a great deal of interest.

Does a Drug-Target have a Likeness?

Much effort is being expended to specify targets sufficiently robust enough to reach marketable therapeutics. Conventional target discovery procedures are generally termed as disease-dependent, which means the necessity of a great deal of experimental work and comprehensive domain knowledge; however, there are statistic-based attempts to predict new molecular targets based on polypharmacology that can be used to investigate the complex drug-receptor interaction network [7]. Like the Lipinski rule, a drug-target-likeness constraint has been proposed by Xu *et al.* for some protein families that show better tractability

when binding ligands with appropriate properties [9]. It was assumed that the structural and functional characteristics of commonly found drug targets taken from the Therapeutic Target Database might be demonstrated by analyzing their shared sequence coded as a fixed-length composition, transition and distribution pattern vector which is captured by statistical learning procedures, e.g. the support vector machine [59]. On the other hand, the provided model focuses more on the druggable potential of the protein hidden in its sequential features rather than on the accuracy of any drug/nondrug target-like distinction. Moreover, drug-target-like does not literally mean drug target. The proposed set of global profiling sequence descriptors coupled with statistical learning algorithms were employed in the negative labeling of proteins that are unlikely to bind prospective drug molecules; however, the accuracy of the approach is still questionable. In this context, the question of how to assess the drug-target-likeness is still open. Inversely, the target-related affinity profiling (TRAP) approach suggests that a protein-binding pattern called a reference panel characterizes the drug binding preferences encoded in the affinity fingerprint vector [60].

EVALUATING THE SIMILARITIES BETWEEN DRUGS

The similarity principle claims that structurally related compounds display a similar biological activity [61]. Although this seems reasonable, the similarity paradox drew our attention to the fact that even minor changes in chemical structure can result in substantial changes in biological activity [62-64]. Despite this paradox, the similarity principle is deeply rooted, not only in molecular design but also, more generally, in chemistry [65] where analogy plays a role, especially when we do not yet fully understand a certain process or effect. Therefore, molecular similarity still underlies the art of molecular design. In Fig. (4) we illustrated basic similarity aspects.

The Concept of Privileged Structures

Homochirality can be an example of a clear natural preference for certain chemical structures and an archetype for other privileged landscape searches. Although chemists are trying to find the laws that control Nature, we do not always succeed, e.g. although we do not know the reason for the olfactory properties of esters, this fact is described in the majority of organic chemistry handbooks. Similar comments refer to other chemical structures that favored a specific biological activity for unknown reasons. Although such observations appeared relatively early [66], the term *privileged fragments or scaffolds* was not coined by Evans until 1988 when he was investigating cholecystokinin (CCK-A) receptor antagonists derived from the natural product asperlicin [16] in order to describe the structural cores in ligands of more than one target receptor.

In a more general sense, PS were defined as a *single molecular framework able to provide ligands for diverse receptors* [16, 67]. Further publications reported the alleged existence of preferred molecular scaffolds, both target-family-specific and cross-target, with an intrinsic tendency toward biological activity [68]. Nicolaou stated that PS are *scaffolds positively correlated with favorable behavior meaning fragments present preferentially in compounds with a desired biological profile* [69]. Although they are intended to be target class-specific, it has been shown that separate molecular subunits also occurred in compounds active against diverse target families [68]. Thus, PS describe a strategy where we can identify common molecular fragments that are universal for a series of ligands binding to an individual receptor or to a particular receptor family. Since then, several reviews have dealt with the concept of privileged motifs [70-72], and numerous molecular fragments have been described as privileged (compare Table 2), e.g. benzazepinone [67], diphenylmethane [67], benzylpiperidine [67], biphenyl-tetrazole [67], indole [67, 73], biphenyl [67, 74], spiroindoline

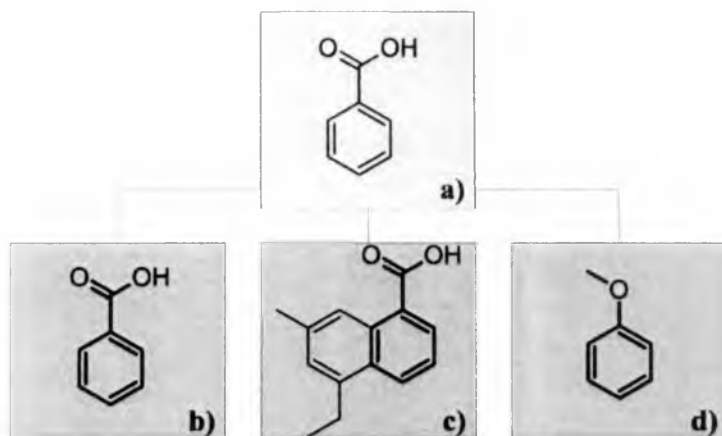


Fig. (4). Several aspects of molecular similarity evaluation. Original structure (a) corresponding to the exact (b), substructure (c) and minor partial (d) matches illustrate a relationship between two molecular bodies.

Table 2. Some of the Alleged PS Shown as Drug Substructures

Diazepam [67, 83]	Lorazepam [84]	Nimetazepam [85]	Midazolam [84]
Cyclizine [67, 86]	Buclizine [87]	Hydroxyzine [88]	Cetirizine [89]
Nifedipine [76, 90, 91]	Nicardipine [92]	Nilvadipine [81, 93]	Felodipine [94]

(Table 2). Contd.....

Valsartan [67, 95]	Irbesartan [96]	Candesartan [96]
Omeprazole [97]	Droperidol [98]	Emedastine [100]
Zolmitriptan [101]	Fludolac [102]	Ergonovine [103]

Benzodiazepine [67, 107] (a), benzylpiperidine [67] (b), 1,4-dihydropyridine [76] (c), bifenyltetrazole [67] (d), benzimidazole [107] (e), indole [67, 73] (f).

sulfonamide [67, 75], spiroindanyl piperidine [75], 1,4-dihydropyridine [76], 2,6-dichloro-9-thiabicyclo[3.3.1]nonane [77], benzopyran [78, 79], pyridazines [80], indoles and quinolines [81, 82].

However, the latter publications were not intended to carefully analyze privileged CS landscapes, but rather to illustrate the fact that there are many active compounds among certain categories of structures, i.e., these are usually drug-oriented reviews targeted at certain segments of the drug development audience. In this sense, PS mean that a certain scaffold is an interesting target under consideration for further synthetic modifications.

A single structural subunit could repeatedly be present in thousands of substances, including various natural products that exhibit miscellaneous pharmacological activities. From the beginning to the present, the original PS concept and definitions have evolved to describe not only those structures that ease ligand binding to an individual receptor or to a particular receptor family, but also those that modulate proteins lacking a strict target class relation.

For example, this is the case for two heterocyclic motifs, namely quinoline and acridine, which are particularly frequent in compounds active against significantly different pathologies [104].

Those kinds of privileged molecular fragments can thus be considered to be *generic drug-like motifs*. On the other hand, the so-called small common entities, i.e., certain linkers that appear frequently in drugs, have also been identified as privileged elements [58]. In contrast, PS should not be confused with promiscuous aggregators, e.g. unspecific behavior caused by the tendency to self-associate into colloids [105].

Privileged scaffolds were successfully used as the core structures for synthesis [106] and optimal starting points for the library design [107] of ligands with an affinity to certain molecular targets. There are several examples of privileged scaffold library synthesis, e.g. based on a 2-arylindole that resulted in the discovery of potent ligands for a variety of G-protein coupled receptors [108], purine derivatives [109] or tetrahydro-1,4-pyrazolodiazepin-8(2H)-ones [110]. In general, PS should typically exhibit good drug-like properties, which in turn lead to additional drug-like compound libraries and leads.

On the other hand, drugs are highly interrelated, e.g. Siegel and Vieth discovered that among 1,386 marketed drugs, 15% of molecules are embedded in the larger drug molecules [111]. Actually, many of the new medicines approved in recent years are in fact only slight modifications of existing marketed drugs. There

is no doubt that this seems to be a direct impact of implementing similarity approaches in drug discovery. Aside from the theoretical advantages, there is an important practical aspect that greatly encourages such approaches in drug design that are highly dependent on industrial and state regulations. Each new medicine has to conform to very strict requirements. In order to satisfy those high standards and to introduce a new drug to the market, it is more convenient to choose the structure of a known drug or any compound that has already passed some of the requirements as the starting material. One cannot forget there is also an economic factor and liability issues as well. The utilization of the structures of already known drugs in the development of new medicines can significantly lower the cost of *pharmaceutical innovation* [112] and therefore they are attractive for the industry as this reduces the risk of pharmaceutical liability. In turn, all of the reasons mentioned have somehow inspired efforts to invent new and completely different methods as a source of structurally distinctive compounds with desired pharmacological actions, e.g. the scaffold-hopping concept [113]. As opposed to privileged motifs, the main goal of scaffold-hopping techniques is to find structurally novel compounds by modifying the central core structure of the molecule but preserving ornamenting groups and still binding to the same molecular target [114].

The complex nature of drug design usually means that defining PS is not simple. Moreover by itself does not automatically induce a preferred bioactivity. In other words, the presence of a privileged scaffold in a molecule does not explicitly identify the target receptor and the desired binding behavior. The question that remains is how to discover new privileged scaffolds?

There are two possible scientific strategies in the search for new privileged motifs: experimental or computational methods. The first one was employed for example in the statistical analysis of the NMR-derived binding data of the biphenyl substructure to show their preferential interactions with a wide range of proteins [74]. The latter are mainly based on data extracted from structural chemical databases.

Fragment-Based Design

The PS concept can be classified as a part of the wider strategy in identifying the small building blocks required for the construction of new drugs. With the ability of NMR techniques, medicinal chemists were able to identify specific small molecules that were interacting with a protein receptor [115, 116]. This can be a valuable tool in building new molecular blocks and constructing new leads. Thus, although analogy and intuition are the foundations of the PS concept, it is a precise measurement that supports the

related *fragment-based design* that is becoming more and more widespread in a real practice [117].

In Fig. (5) we illustrate the substantial paradigm in *fragment-based design* or *frugonomics*. It is easier to find a smaller fragment of the drug than to find a complex drug [117, 118]. Analyses of the low molecular motifs offer a variety of novel and original methods in both computational and experimental investigations. In theoretical studies, the search for low-molecular weight fragments capable of binding with a target receptor can be less computationally demanding. In experimental researches, the higher probability of matching smaller and less complex fragments into the receptor also ensures the higher success ratio of the ligand receptor alignment [117]. Accordingly, the knowledge of small fragments that interact with a receptor seems to be useful for the construction of more complex structures, regardless of the relatively low affinity level of the starting motifs.

COMPUTATIONAL PROBLEMS IN PS IDENTIFICATION

Identifying a part of a molecule as privileged requires the fragmentation of a molecule. After all, PS is a fragment. Since fragments can contribute additively into chemical or biological effects, they form the theoretical basis for a number of approaches [119].

From Early Fragmental Concepts in Chemistry and Drug Design to Fragment Frequency Analyses

The Free-Wilson QSAR approach describes a series of molecules as a complex construct of a *core framework* and its substitution pattern. In this context, it can be described as an early example of fragment-based drug design [39]. In fact, fragmentation is an idea that inspired the Hammett equation, in which two molecular fragments are identified; one is a regular aromatic carboxylic acid *framework*, while the second is based on a series of its changing substituents. To give an analogy, Hansch's lipophilicity is constructed on a similar approach, and the Rekker hydrophobic fragmental approach [120] even more closely resembles the current fragmentation methods, 1D simplex [121], i.e., the collections of any atomic combination, can be cited here as the deepest fragmentation concept, while the neural network procedure can be classified as a fuzzy 1D or 2D simplex identification [122, 123]. The evolution of the early fragmental concepts is shown in Fig. (6). Todeschini distinguished substructure as a special kind of molecular description, for a precise definitions and discussion of the number of individual substructure (or fragment based) descriptors the reader is referred to the reference [35].

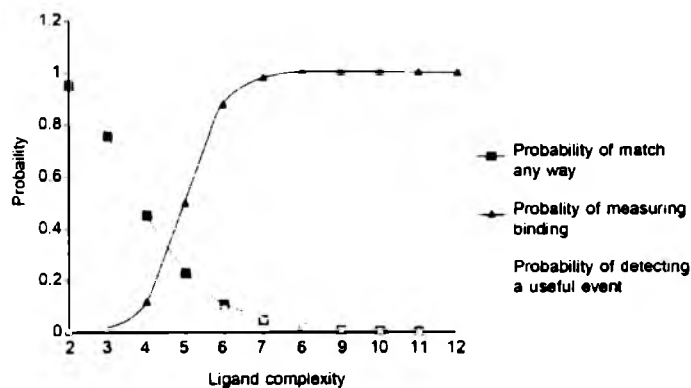


Fig. (5). Success landscape for a binding site of complexity 12. Reprinted with permission from [117]. Copyright 2005 Elsevier Ltd.

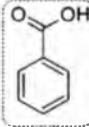
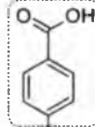
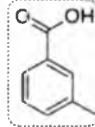
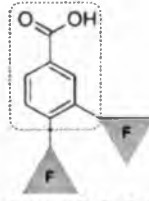
Equation/description		Fragmental representation																																																																									
Hammett	$\log \frac{k_F}{k} = \rho \sigma$																																																																										
Hansch	$\log \frac{P_F}{P} = \pi$	framework																																																																									
Free-Wilson	<table><thead><tr><th>ortho</th><th>para</th><th colspan="3">meta</th><th colspan="3">para</th></tr><tr><th>F</th><th>F</th><th>F₁</th><th>F₂</th><th>F₃</th><th>F₁</th><th>F₂</th><th>F₃</th></tr></thead><tbody><tr><td>H</td><td>F₁</td><td></td><td></td><td></td><td>1</td><td></td><td></td><td></td></tr><tr><td>H</td><td>F₂</td><td></td><td></td><td></td><td></td><td></td><td>1</td><td></td></tr><tr><td>F₁</td><td>H</td><td>1</td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>F₂</td><td>H</td><td></td><td>1</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>F₁</td><td>F₂</td><td>1</td><td></td><td></td><td></td><td>1</td><td></td><td></td></tr><tr><td>F₂</td><td>F₁</td><td></td><td></td><td></td><td></td><td></td><td>1</td><td></td></tr></tbody></table>					ortho	para	meta			para			F	F	F ₁	F ₂	F ₃	F ₁	F ₂	F ₃	H	F ₁				1				H	F ₂						1		F ₁	H	1							F ₂	H		1						F ₁	F ₂	1				1			F ₂	F ₁						1	
	ortho	para	meta			para																																																																					
F	F	F ₁	F ₂	F ₃	F ₁	F ₂	F ₃																																																																				
H	F ₁				1																																																																						
H	F ₂						1																																																																				
F ₁	H	1																																																																									
F ₂	H		1																																																																								
F ₁	F ₂	1				1																																																																					
F ₂	F ₁						1																																																																				
	Matrix for Free-Wilson analysis  F - fragment F₁, F₂, F₃ or H																																																																										

Fig. (6). Early fragmental concepts in Hammett, Hansch and Free-Wilson methods.

$\text{HO}-\text{C}(=\text{O})-\text{CH}_2-\text{CH}(\text{NH}_2)-\text{C}(=\text{O})-\text{OH}$		
size 3	size 4	size 5
$\text{C}''-\text{CH}_2-\text{CH}-$ $\text{C}''-\text{AH}-\text{CH}_2-$ $\text{NH}_2-\text{CH}-\text{CH}_2-$ $\text{NH}_2-\text{CH}-\text{C}-$ $\text{O}-\text{C}-\text{OH}$ $\text{OH}-\text{K}''-\text{CH}-$ $\text{OH}-\text{K}''-\text{CH}_2-$ $\text{O}-\text{K}-\text{CH}-$ $\text{O}-\text{K}-\text{CH}_2-$	$\text{C}''-\text{CH}_2-\text{CH}-\text{C}-$ $\text{C}''-\text{AH}-\text{CH}_2-\text{C}-$ $\text{NH}_2-\text{CH}-\text{CH}_2-\text{C}-$ $\text{OH}-\text{C}''-\text{AH}-\text{CH}_2-$ $\text{C}-\text{C}-\text{AH}-\text{CH}_2-$ $\text{OH}-\text{K}''-\text{CH}-\text{CH}_2-$ $\text{OH}-\text{K}''-\text{CH}_2-\text{CH}-$ $\text{O}-\text{K}-\text{CH}-\text{CH}_2-$ $\text{O}-\text{K}-\text{CH}_2-\text{CH}-$ $\text{OH}-\text{K}''-\text{CH}-\text{NH}_2$ $\text{O}-\text{K}-\text{CH}-\text{NH}_2$	$\text{OH}-\text{C}''-\text{CH}_2-\text{AH}-\text{C}-$ $\text{O}-\text{C}-\text{CH}_2-\text{AH}-\text{C}-$ $\text{OH}-\text{C}''-\text{AH}-\text{CH}_2-\text{C}-$ $\text{O}-\text{C}-\text{AH}-\text{CH}_2-\text{C}-$ $\text{OH}-\text{K}''-\text{CH}-\text{CH}_2-\text{C}-$ $\text{OH}-\text{K}''-\text{CH}_2-\text{CH}-\text{C}-$ $\text{O}-\text{K}-\text{CH}_2-\text{CH}_2-\text{C}-$ $\text{OH}-\text{K}''-\text{CH}_2-\text{CH}-\text{NH}_2$ $\text{O}-\text{K}-\text{CH}_2-\text{CH}-\text{NH}_2$

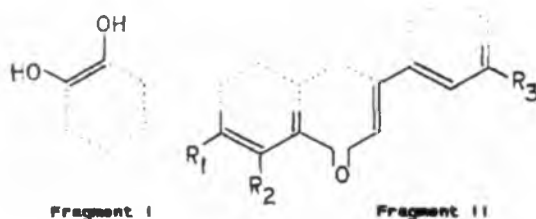
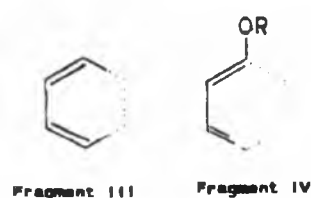
ACTIVATING FRAGMENTS**INACTIVATING FRAGMENTS**

Fig. (7). Structural subunits-based recognition of activating and inactivating molecular fragments. The idea of explained for aspartic acid (top); activating (privileged) and inactivating [unwanted] fragments revealed (down). Reprinted with permission from [124]. Copyright 1984 American Chemical Society. Reprinted with permission from [126]. Copyright 1988 IUPAC.

The Computer Automated Structure Evaluation (CASE) approach is a pioneering approach for recognizing and categorizing

active and non-active molecular structures based on their structural subunits positively or negatively correlated to biological activity

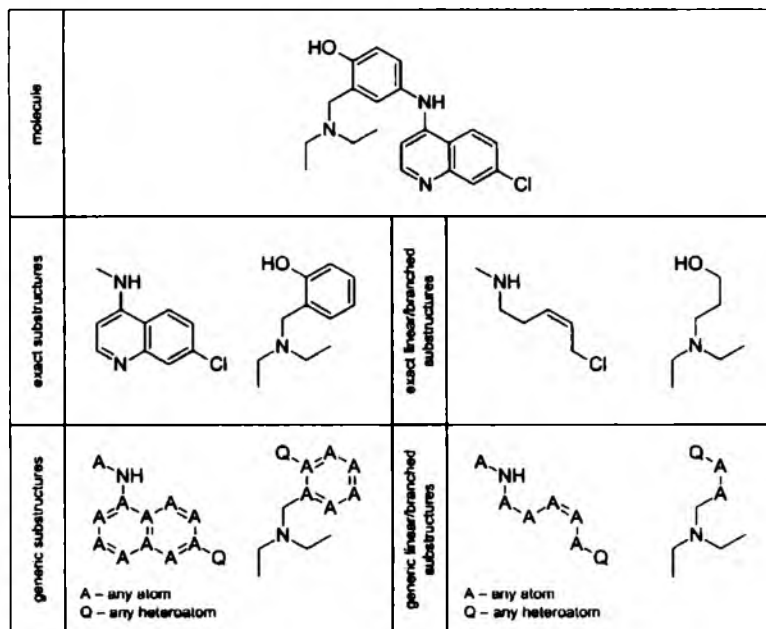


Fig. (8). Substructure definition modified from [22].

[124, 125]. Fig. (7: up) explains fragmenting aspartic acid structural subunits and in Fig. (7: down), we present a visualization of two categories of structural subunits, activating and inactivating flavonoid series as larval growth inhibitors [126]. Recently, the frequencies of fragments sampled in a molecular series have been used for analyses of kinase potency and selectivity, for example [127]. Other examples are discussed in review [39].

Fragmentation Methods

The arrangement of fragmentation is not an obvious pattern for molecules. Although molecules can be dissected while preserving rings and scaffolds, we can also define a fragment as a *particular chain or of carbon atoms. The definition follows from the breaking rules that are used* [128]. In such a definition, we can understand fragments as any atomic assemblies with identical molecular substructures. In Fig. (8), we illustrate this concept of substructure according to reference [22]. Table 1 provides a short overview of the terms used in molecular fragmentation, i.e. structure, substructure, fragment, scaffold, etc.

Formally, fragmentation can proceed according to two protocols. The first is to find all of the possible fragments through extensive analyses of the molecular substructures. In the second approach, we *a priori* define the breaking rule(s) that outline(s) the fragments under consideration [128]. Although finding all of the substructures in a molecular series seems impractical because of the explosion of their possible number, such methods are well known in QSAR, e.g. H-QSAR [129-131]. Mathematically speaking, a substructure is a subgraph of the connected and undirected molecule graph that directly reflect the features of the edges (atoms) and vertices (bonds) in the corresponding parental structure [22]. In a more general sense, the substructural approach can be related to the so-called frequent subgraph mining (FSM) or common substructure methods, both of which are based on graph theory. Readers should refer to the reference [128] for a more

detailed review of the computational problems of both approaches. More practical and complete FSM approaches to drug design are still a matter for future developments. In the second approach, we give up the analysis of the complete substructure patterns and concentrate instead on some building blocks, e.g. defined using the Bemis and Murcko method [11, 12].

Table 3. Different Fragmentation Methods

Category	References
knowledge-based	[16, 67, 132-138, 140-147]
hierarchical	[11, 12]
synthetically-oriented	[43, 115, 139, 144]
random	[145-150]

Formally, fragment design and generation can be classified into three major categories; knowledge-based, hierarchical and synthesis-oriented, as is summarized in Table 3. For instance, knowledge-based methods have been applied for the prediction of ADME properties [132] or the exclusion of compounds with unwanted structural motifs (reactive or toxic groups) from screening sets [133, 134] or for the design PS [16, 135, 136]. Systematic approaches to generate atom- and bond-centered fragments have provided the starting point for the design of fingerprints that capture strings of layered atom environments [137, 138]. The hierarchical method systematically divides molecules into core structures, linkers and side chains [11, 12]. Core structures can be represented by different levels of abstraction – as scaffolds or frameworks. Synthetic criteria are also taken into account to lead to fragment design [115, 139], including retrosynthetic fragmentation techniques to generate building blocks for library design.

Random approaches have only recently been introduced. These kinds of techniques broaden conventional fragmentation methods

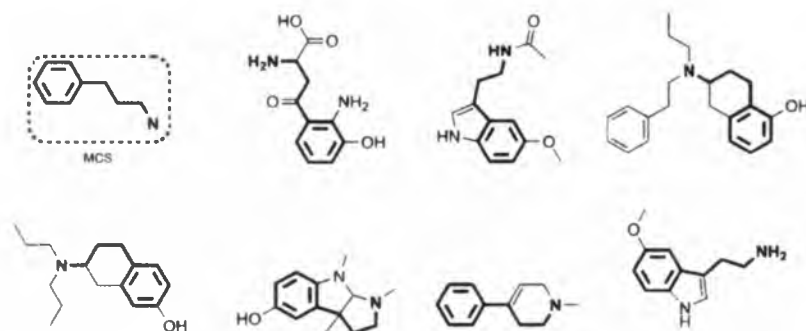


Fig. (9). Maximum common substructure (MCS) for a compound collection [50].

because they do not require any predefined fragment determinations and are thus a potential source of novel fragment descriptors. For example, a Brownian approach to fragment sampling through molecular graphs in order to evaluate the information content of molecules was reported [145].

Computational Approaches to PS Identification

Molecular similarity and diversity approaches are well known and have been used for database screening and library design with varying success. These concepts are reviewed in a number of available publications and monographs and will not be discussed here [61, 64, 65, 137, 143, 146, 149, 151-154].

In the maximum common substructure (MCS) method, the molecules are fragmented into smaller substructures (parts) as long as a common fragment can be identified among all members of the series, as is shown in Fig. (9). This approach, which is used for database screening [50], also seems to be especially useful for the discovery of a preferential landscape. The larger the MCS; the more specific interaction. In turn, the larger the series is; the smaller the MCS that can be found. A small MCS is not useful for virtual or database screening since it would provide too many hits [50]. Those types of investigations have shown that the diversity of known drugs is relatively low. Bemis and Murcko, who analyzed the common structural features present in commercially available drugs, discovered that about half of the drugs registered in the Comprehensive Medical Chemistry (CMC) database were described by the 32 most frequently occurring frameworks [115]. Furthermore, based on data retrieved from the BindingDB database, a systematic and comprehensive computational selectivity-centric analysis of public domain target-ligand interactions was conducted.

As a result, more than 200 molecular scaffolds were identified that are selective for individual targets among closely related ones and a majority of these scaffolds yield compounds that display a target-selective tendency. These scaffolds are underrepresented in approved drugs [155]. Consequently, the frequency of the occurrence of that kind of generic drug-like molecular fragments among drug populations and bioactive compounds ensembles could be a valuable index in the estimation of PS.

Gianti *et al.* developed a structure dissection procedure for the identification and selection of privileged fragments that is based on the physicochemical properties falling within the ranges of rule of three (Ro3) threshold and a topological shape similar to known drugs [31]. The reference set of the ZINC drug-like hits was decomposed, which resulted in a subset of substructures ordered according to an increasing level of complexity and belonging to distinct *decoration spheres* around the main core present in the parental molecule. Cluster analysis allowed the authors to group

and select a virtual representation of novel and original fragments while preserving the topological drug-like pattern that might be potentially useful for screening vendor's commercial libraries [42]. Apparently, the reduced PS patterns reveal greater potency within a class of targets [15]. This inspired scaffold-based classification in order to sort the ligands into compound classes and/or to measure their similarity to a certain scaffold [156]. Sutherland *et al.* showed that the fragment frequency in active molecules might specify the target similarity [127]. Massive database mining for fragment co-occurrences might potentially delve into the architecture of extensively or rarely explored regions of the chemical space. Naturally, questions arise about the preferable combinations of substructures comprising a molecule. Should it reflect only *chemical habits* or even *fashions* among chemists due to synthetic feasibility or is there any *natural force* that combines some *building blocks* together? Lamieijer *et al.* proposed the term *chemical clichés* for the fragments commonly found together in the NC1 repository [41]. The most/least frequently met substructural combinations were analyzed taking into account their expected and real co-occurrences, which revealed some fragment preferences, e.g. mutual attraction between phenyl and THF groups [41].

Vieth and Siegel compared fragment distribution within oral and injectable marketed drugs [111]. The assignment of fragment importance has been performed for substructures with frequencies greater than the standard deviation for each group and mutual frequencies have been compared using the difference function [157]:

$$Diff = 100 \frac{abs(freq_i^{oral} - freq_i^{injectable})}{\max(freq_i^{oral}, freq_i^{injectable})}$$

The difference is 100% when a fragment is found only in one group, whereas the difference is 0% when fragment is present in both groups. This kind of comparison might be helpful in understanding whether the *property in differences between these sets would translate into differences in nature* [111].

The *unprivileged substructures* as well as the *conditional privilege* rules that relate combinations of fragments to a biological affinity have been identified using the Tanimoto coefficient and decision tree based fragmentation method [3]. Decision trees were employed as classification models in order to specify substructures that are frequent in a set of active molecules and relatively rare in inactive ones. Finding discriminating fragments might be crucial in determining the potential activity of a compound. It was reported that many of the identified discriminating substructures possess a structural homology to naturally occurring molecules [158]. Among *unprivileged fragments* are highly flexible sp^3 -hybridized long carbon chains, but there is no necessity to exclude such motifs from

chemical libraries as they might provide binding specificity [3]. The *conditionally privileged substructures* did not show enhanced activity in individual assays but heightened activity was revealed only in the presence of other fragments.

A fragment-based strategy (FBS) for the efficient docking of the flexible ligands in the rigid binding/active site of the receptor/enzyme with the automatic decomposition and identification procedure (DAIM) has recently been proposed [36]. The fingerprint-based in-house-developed identification procedure with simple structural keys being generated for the resulting fragments was used to evaluate their suitability as anchors with highly favorable interactions with a target protein upon binding [159]. A large ZINC database collection has been fragmented resulting in a list of the most frequent cyclic/acyclic fragments that are similar to Murcko frameworks [36]. However, it is highly questionable whether the FBS technique is able to indicate specific fragments from a structurally diverse but non-target-directed fragment library. It was suggested that fragment-binding affinity is generally a target-dependent feature focusing more on the local fragment's adaptability to the diverse binding motifs present in different proteins [115]. The hydrogen bonding (HB) ability undeniably plays an important role in molecular recognition [160]. In fact, detailed knowledge of the hydrogen bond geometries, its directional preferences and inter/intramolecular motif formation is vital for any investigation of the fragment-receptor short-range non-covalent interactions. The number and strength of the non-bonded intermolecular contacts between the fragment or drug-size molecule and their complementary macromolecular target are of major importance for their mutual recognition process. Moreover, the alteration of the local hydrogen-bonding pattern is generally regarded as the primary procedure for the modification of ligand-receptor selectivity. Hence, the meticulous analysis of the hydrogen-bonding interactions is highly advisable in order to gain insight into the intermolecular geometry, including the favorable conformational preferences. However, it might be challenging to specify the strict probability criteria for the formation of the hydrogen bond. A favorable spatial distribution of the corresponding donor/acceptor atoms indicates some regularity in both the angular and distance locations, e.g. an oxygen in the carbonyl group carrying two lone pairs of electrons is potentially able to create the O $\cdots$ H O bond with a hydroxyl group donating a hydrogen located somewhere in distance 1.6 $\pm$ 2.2 Å and pseudo-angle 120 $\pm$ 180° [161].

In a series of publications, we have recently reported the application of the MoStBioDat platform for the mapping of the fragmental drug-likeness topology (FDT) and the intramolecular hydrogen bonded (IHB) motifs in diketo acid (DKA) related compounds as the DKA arrangement is a commonly observed structural feature that is of crucial importance for creating HIV-1 integrase (IN) inhibitors [162, 163]. The virtual 3D data retrieved from the PubChem and ZINC repositories have been compared to the PDB experimental and theoretical models (the cocrystal structures of the prototype foamy virus (PFV) complexed with RAL and ELG) acting as a benchmark moiety. A pharmacophore-based methodology coupled with three-dimensional (3D) database screening might enhance the probability of hit/lead discovery with novel structural scaffolds. Therefore, we compared the structure-guided DKA pharmacophore models against the freely available *on line* repositories of structurally diverse small molecules.

A number of the structurally diverse chemical compounds with functional DKA subunit(s) were revealed by combined *on line* and MoStBioDat 3D pharmacophore-guided ZINC and PubChem database screening. We used the structural data available from such a screening to analyze the similarities among compounds containing the DKA fragment. Generally, the analysis by PCA and the SOM neural network revealed four families of molecules that comply with the chemical constitution (aromatic, aliphatic) of the

compounds. From a practical point of view, similar studies can reveal potential bioisosters of known drugs, e.g. raltegravir/elvitegravir.

Fragment-based classification of protein targets into druggable and non-druggable subsets requires the identification of the binding site and a reliable quantitative measure for the evaluation of target druggability [2]. Huang and Jacobson demonstrated the discriminative power of virtual fragment screening, whereas Chen *et al.* compared the binding sites of the related targets and specified the cavity regions that are important for ligand specificity [164]. Target-family-specific substructure analysis indicated that a high frequency of some privileged fragments does not necessarily determine the target selectivity, e.g. the biphenyl scaffold was regarded as a privileged moiety for GPCRs, although it was detected in 5,658 bioeffectors associated with 311 activity classes [165].

From Natural Products to PS

Nature is an unlimited and constructive library of biologically active substances, and both nature and drug discovery approaches have contributed heavily to the modern pharmacopoeia. Approximately 50% of all drugs on the market have their roots in NP or NP analogues [166]. Moreover, a natural product-like chemical space that is populated with a diverse but biologically validated pool of structures (10^6 - 10^7) which specifically interact with a protein target is believed to contain scaffolds with the potentiality to be privileged structures [52]. Thus, the scaffold diversity and property distribution in the synthetic drug and natural product libraries is under intensive scrutiny [167, 168]. Despite some subtle variations among the findings obtained, some stable tendencies have been observed that depend mainly on the analyzed NP subset, e.g. pure, plant-derived or marine NP. The *in silico* screening of NP drug-like properties revealed that the number of Ro5-compliant molecules is high – only 10% of NP violated the Ro5 and most shared pharmacokinetic features with the orally bioavailable drugs [169]. However, NP are slightly more lipophilic than drugs containing more or less the same number of nitrogen atoms (HBA) but twice as many oxygen atoms (HBD) and 10 times less sulphur and halogens compared to other sets [170]. Genuine natural products generally have a higher and more broadly distributed Topological Polar Surface Area (TPSA) value in comparison to drug molecule [10]. As for the structural features, two more ring systems and three times as many stereogenic centers have been detected within the NP subset. Even though NP possess a similar number of unsaturated motifs as drugs, they incorporate less aromatic rings and a lower number of rotatable bonds [170]. A graph-based analysis showed that the six-membered ring is the most common scaffold in both drugs and natural products. Taking into account these findings, it was concluded that natural products are basically more rigid and drug-like structures than synthetic drugs [10]. Surprisingly, a closer look at NP-like compounds without a proven annotation as NP, namely *seminatural* compounds, clearly indicated a higher property resemblance to drugs than to NP especially in relation to the average number of chiral centers and aromatic rings [170]. The decomposition of NPs-based graph representation into the assembly of molecular scaffolds or distinctive skeletons inevitably reflects the general truth that 'great things are done by a series of small things brought together' (Vicent van Gogh, 1853-1890) [171]. The structural classification of natural products based on the underlying scaffolds with respect to the species of the origin of NP showed their restricted chemotype diversity [10]. It seems that NP compose a homogenous group of compounds with properties *forged in the crucible of evolution*; therefore, the biotope might reduce the availability of some elements, e.g. nitrogen in plants [172]. On the other hand, the NP that were examined comprise only a tiny fraction of structurally abundant nature, which reflects a significant gap in our knowledge.

For further analysis of the relation between NP and PS, readers should consult the excellent review and analyses of natural product oriented scaffold identification [173]

Enumeration of PS in a Virtual Chemical Space

Contemporary HTS methods often require a large number of virtual structures as potential drug candidates. We can use molecular databases as a source for such structures, e.g. those available online or our own generated databases. The methods for constructing such databases of virtual compounds are reviewed in reference [174]. In the same review, readers can find a discussion on the alternative method of so-called *chemical space travel* algorithms, which are used when an even larger number of molecules is needed. This can generate a focused library within a limited region of chemical space, i.e., the structures preferentially fitting a particular receptor.

Database Mining and Knowledge-Based Discovery for Fragment Identification

Knowledge discovery in databases is defined as the nontrivial extraction of implicit, previously unknown and potentially useful information from data [175, 176]. With the enormous increase in the amount of data, there is an urgent need for the invention and development of computational approaches to extract reliable information from all kind of databases, especially the rapidly growing molecular and structural chemical data resources. Typically, the screening procedure takes into account the entire structure or substructure as a search query for identifying the ensemble of compounds fulfilling the global similarity criteria. Substructure or pharmacophore searching in a database locates all of the compounds that share a common user-defined pattern of atoms or bonds in a specified topological or geometric arrangement [177, 178]. The algorithms that are essential for the three-dimensional substructural moiety specification are mainly based on subgraph-matching techniques [179]. The 2D-based procedure of finding a particular molecular motif can be performed using a regular expression approach (SMARTS) as well [180]. The transformation of the molecular (sub)graph representation into its canonically coded linear atoms/bonds specification enables massive substructure recognitions [181]. Typically, the substructures might be represented by exact and/or fuzzy linear/branched/cyclic motifs where nonhydrogen atoms are encoded symbolically with wildcards. The simplest substructure-based procedure that examines the (co)occurrence of fragments is based on the predefined library

of molecular motifs; however, similar to the structural keys, the main disadvantage is the inability to define all of the possible pharmacologically active substructures [22]. As an alternative, the MCS procedure is an intuitive approach where the largest shared component in the structurally related drugs might determine their biological activity [182]. The challenging issue of the restricted substructure dictionary can be circumvented by using various partitioning techniques.

PREFERENTIAL ARRANGEMENT OF AZANAPHTHALENES

The analysis of the problems of the identification of preferential landscapes presented above allows us to classify the PS approach as a special kind of building block analysis, where we often use more or less intuitive chemical knowledge for the identification and analysis of the activity rules within certain chemical scaffolds. Below, using a similar approach, but supplemented with the extensive database mining, we analyze a preferential arrangement of the azanaphthalene scaffold (AN) in drugs and natural products.

While mining databases, we found some highly interesting regularity. Chemical libraries of AN are dominated by monoaza-compounds, among which quinoline appears to be the most frequently investigated scaffold. The analysis performed on data stored in the MDL CrossFire Reilstein database revealed an alternating tendency of waxing and waning interest in a given group of compounds during the space of the last ninety years (Fig. (10)). Increasing interest in a certain class of chemical compounds (scaffold) triggers new investigations and after some years, the trend once again fades. The waves of interest superpose on the increasing number of registered compounds. Those fluctuations in the number of citations are present in each azanaphthalene ensemble labeled as ID = 1 to 12, in turn (Fig. (11)). The number of publications that refer to quinoline derivatives as a product in chemical reactions reached a peak in 2008. The revealed trends not only follow the common sense of the *me too* strategy in medicinal chemistry, and chemistry in general, but also provide an illustration of the fact of how sensitive database mining can be in revealing latent relationships and trends.

The analyses available in literature [104] indicated quinoline as a PS. Actually, quinoline antibiotics constructed of the quinoline ring can be the most spectacular example of the potential efficiency of this system in medicinal chemistry [183] while a quinine molecule that also contains a quinoline moiety proves Nature's preference for the system. The quinoline scaffold can also be found

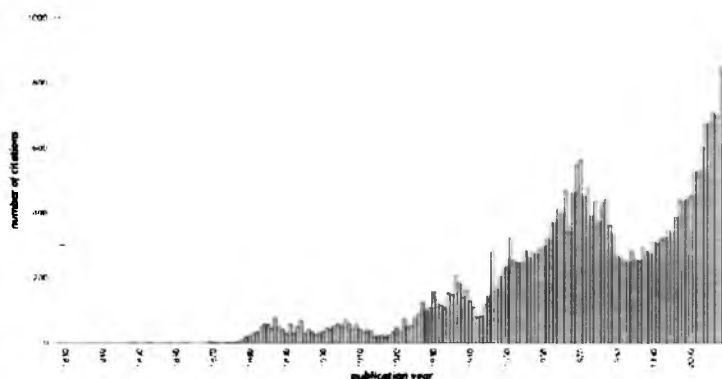


Fig. (10). Trends and interests in organic synthesis during the space of last ninety years. Plot of the number of citations in substructure searching for compounds with quinoline moiety as product in reaction. Data were extracted from Reilstein 2010/01 Database; access date: 14/05/2010.

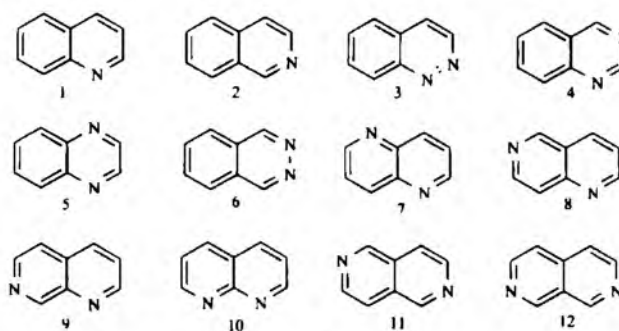


Fig. (11). Azanaphthalene (AN) scaffolds. All possible mono- and diANs scaffolds labelled as ID = 1 till 12.

in many classes of other biologically active compounds that are used as antifungals, antibacterials and antiprotozoic drugs as well as antituberculous agents. Some quinoline analogues also showed antineoplastic activity [81]. In Table 4 we compared the database frequencies of quinoline (q) vs. isoquinoline (iq) in the Beilstein (B), a synthetic-oriented database and DrugBank (D), a drug-oriented database.

Table 4. Database Hits of Quinoline and Isoquinoline Fragment Containing Molecules

	Beilstein (B)	DrugBank (D)	(D/B)
quinoline (q)	85857	32	$3.73 \cdot 10^{-4}$
isoquinoline (iq)	15391	1	$0.65 \cdot 10^{-4}$
parameter q/iq	5.58	32	-

A simple parameter of the *D* to *B* hit ratio clearly indicates the relative overrepresentation of the quinoline (q) systems vs. isoquinoline (iq) ones in the *D* vs. *B* data. One would wonder how these results compare to other AN scaffolds. In other words, is quinoline really a privileged structure among all AN? Fig. (12) illustrates the number of the PubChem database registered compounds having a given AN scaffold. Quinoline (Fig. (11), label 1) was also the most frequent hit in this database. What is the origin of this popularity: practical applications, synthetic availability or something else?

In Fig. (12a), we plotted two parameters: the number of compounds tested to all hits and the active to tested ratio (*b*-value), which are the simplest measures of attractiveness (tested to all) and a kind of drug-likeness (active to tested). Databases are updated regularly, which determines their usability; however, this also creates a problem for systematic mining. Fig. (12b) plots the results of systematic analyses over six months in order to evaluate the *b*-value changes during this time. As had been supposed, the calculated parameter values differed day to day; however, the previously observed trends in the attractiveness and drug-likeness among the twelve investigated AN scaffolds remained stable. If these simple descriptors were reliable, one would be able to distinguish between compounds that raise the highest interest simply because they are available as well as compounds that feature privileged motifs. However, because the amount of AN differs significantly for different scaffolds, very small for some AN, conclusive interpretations seem to lack credibility. In Table 5 we present a detailed frequency analysis on natural products having AN scaffolds that resulted from the searches through several databases. Thus, scaffold 1, i.e. quinoline is the most frequent not only in natural sources but also in Drugbank. Interestingly, scaffold structures 2, 5 and 12 that can be found in natural compounds are

much less frequent among drugs, while scaffold 4 constructs drug molecules with 2/3 of the frequency of quinoline. In general, the number of biologically active compounds can be more or less correlated to the number of all compounds separated from natural sources. Scaffold 12 is the only one for which not a single active compound has been registered in the Reaxys database. The comparison of the number of natural compounds (column 4, Table 5) vs. the number of compounds of the activity registered (column 2, Table 5) is worth further comment. This indicated how small a fraction of these compounds was tested, probably due to purity or availability issues. Below we briefly discuss several of the most prominent scaffolds among those analyzed in Table 5.

Quinolines

Quinine is probably the compound that has affected quinoline-related drugs throughout history. Its introduction in medicine is not known. Perhaps the Peruvian Quechua Indians were the first to use cinchona bark as a remedy for fever [184]. Extracts from cinchona bark were introduced in Europe and later purified quinine was broadly used in the treatment of malaria from the 17th century. It is still in use as an antiparasitic drug (mainly for treatment of babesiosis); however, other drugs have replaced it in general use [185, 186]. It is noteworthy that newer antimalarial drugs such as chloroquine, primaquine and mefloquine are also built on the quinoline moiety [187, 188]. Today, more than half of the hundreds of metric tons of the world's production of quinine is used in tonic water, a custom also borrowed from the Quechua people. Although quinine is no longer the drug of choice in malaria treatment, it remains a point of interest for its wide spectrum of activity, which encompasses antiviral [189], antimicrobial [190, 191] treatments as well as being used to prevent muscle cramps [192]. Although the side effects such as its influence on insulin secretion and hypoglycemia [193] have limited its direct use, quinine remains a leading structure.

Another closely related alkaloid derived from the bark of Cinchona tree is quinidine. Although structurally quinidine is stereoisomer of quinine, its biological activity differs slightly. It is best known as agent for reducing the excitability of cardiac and skeletal muscles by blocking sodium and potassium channels [194] and is used as an antiarrhythmic drug. Nevertheless, the influence of Cinchona on pharmacology is far more significant than it may appear to be from the uses of two above-mentioned drugs. By learning the lesson that Nature gave us with this plant, we were able to introduce probably the richest and most important group of drugs – drugs with a much wider spectrum of use than as an antiparasitic. The early modifications of the quinine structure were stimulated by the insufficiency or unavailability of natural sources of these alkaloids [195]. The only sources of cinchona bark were outside of Europe on plantations that were in strategic positions during the

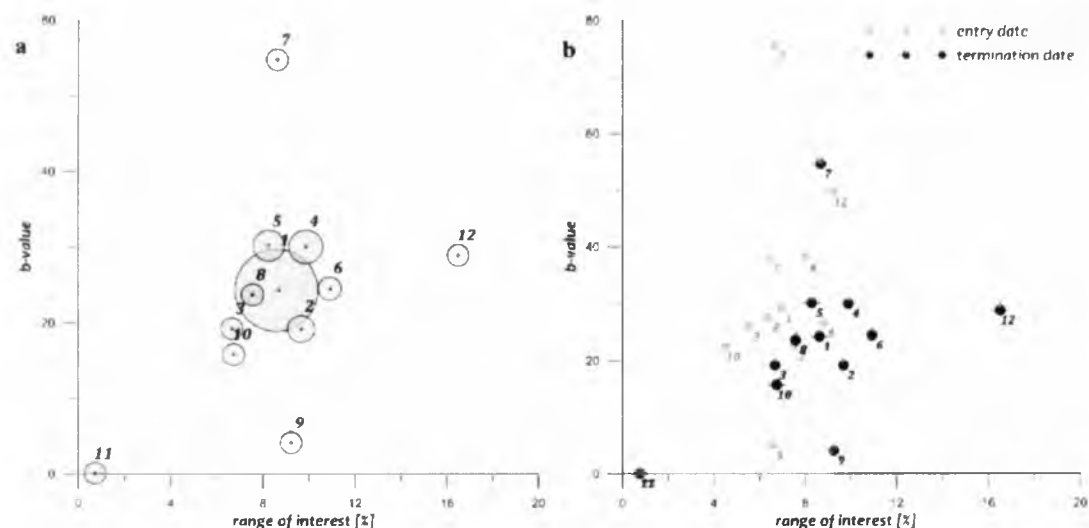


Fig. (12). Blnological activity among AN compounds. Findings on the basis of data obtain during PubChem database mining. PubChem active vs. tested rates for mono- and diANs (a). Bubble's diameter (a) reflects total ensemble's plentifulness; *b-value* is defined as percentage ratio of active compounds among tested compounds in each ensemble, whereas range of interest is the ratio of tested compounds to all compounds. Directional trends in *b-value* changes during the space of last six months (b), where 27/05/2010 was set as an entry date and 12/11/2010 as a termination date of analysis.

Table 5. AN Frequencies in Several Databases

AN scaffold	Number of Natural compounds of the known activity (Reaxys) ^a	Number of drugs (Drugbank) ^b	Number of natural compounds (Reaxys) ^c	MEGAbitte
1	289	112(34)	689	144
2	200	27(1)	490	2
3	0	2(1)	1	0
4	1	77(9)	1	9
5	43	22(4)	108	0
6	1	9(2)	1	0
7	0	1(0)	0	2
8	0	0	9	0
9	0	2(0)	0	0
10	1	4(4)	1	0
11	0	0	2	0
12	0	2(0)	51	0

Searching options

^apharmacological data available, available as a natural product isolated substance, substitution allowed on each atom of the scaffold

^bas substructure, all drugs; in parenthesis – approved

^cavailable as a natural product isolated substance, substitution allowed on each atom of the scaffold

times of colonization and the First World War. Attempts to find an independent and robust source of antimalarial drugs led to the discovery of the quinacrine. However, this drug was manufactured only by the Bayer Company in Germany; thus the outbreak of World War II brought an end to its worldwide availability creating a strong need to create drugs suitable for use in the prevention and treatment of malaria in soldiers fighting in tropical areas. Although American scientists were able to produce the quinacrine, the soldiers refused to take the pills for fear of side effects. In response, a new superior antiparasitic agent – chloroquine was introduced [3]. This drug is still in use [196, 197]. Later, during the development of chloroquine, a new compound with interesting properties was discovered as an impurity – nalidixic acid. This drug because of its

poor pharmacokinetics, relatively high active concentration and narrow spectrum of activity is mainly used in the treatment of bacterial infections of the urinary tract [198, 199]. Nevertheless, the research on nalidixic acid derivatives has led to a new era of fluoroquinolones [200]. Derivatization with fluoride at the C6 position of the quinoline skeleton resulted in the development of the second generation of quinolone antibiotics (Fig. (13)) and further modifications led to the next generations of improved drugs.

It is noticeable how quickly fluoroquinolones became popular in antibacterial therapy. A number of drugs have been developed in a relatively short period. Currently, more than twenty active compounds that belong to this class of drugs have revolutionized antimicrobial therapy [199, 201-203].

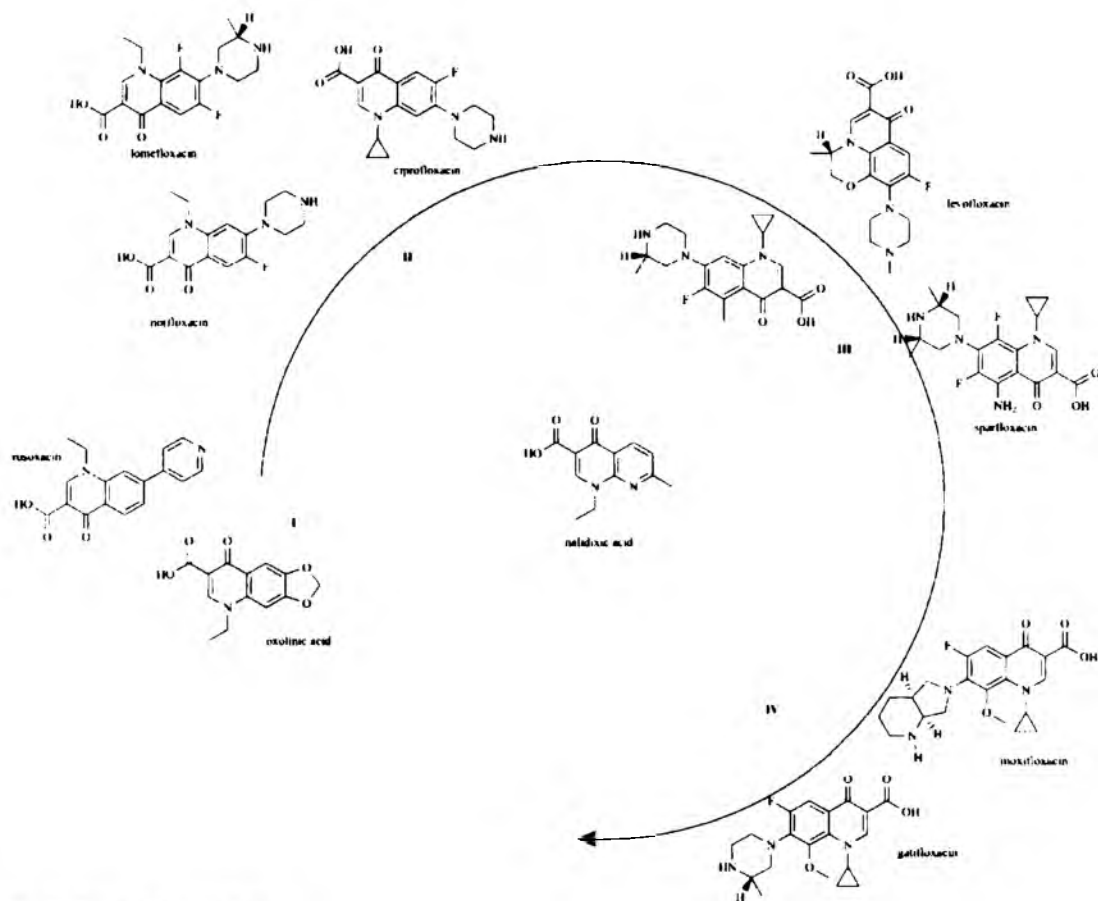
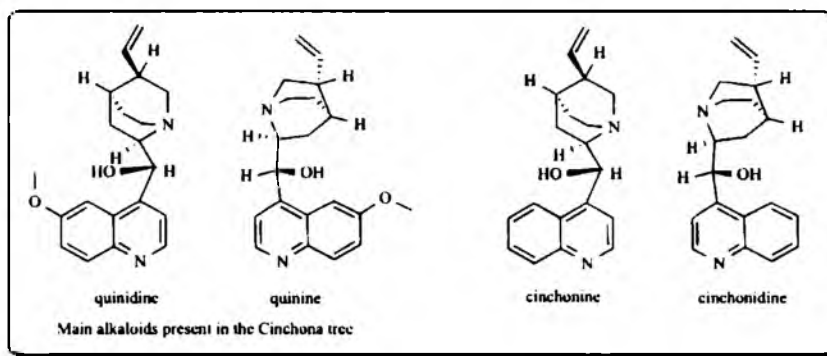


Fig. (13). History of fluoroquinolones.

Fluoroquinolones are rather well tolerated and have good pharmacokinetic parameters, which allows for oral administration although there are several serious adverse effects along with the resistance that is an effect of continuous overdosing, which have

had an impact on their use [204, 205]. Their broad spectrum of activity and applicability includes not only urinary tract infections but also respiratory or skin infections and sexually transmitted diseases, and it is still an effective trigger for further efforts that

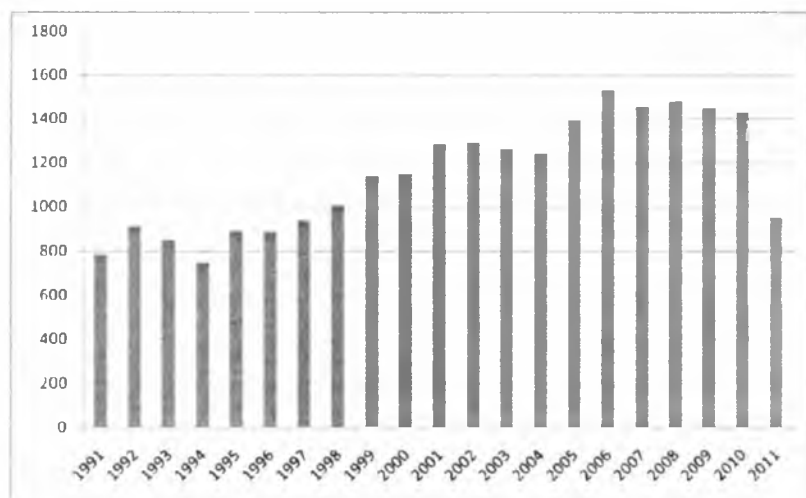


Fig. (14). Fluoroquinolones in last 20 years.

might lead to safer drugs. The introduction of fluoroquinolones to the second line treatment of tuberculosis and in other multidrug resistant infections is pointing towards further development of new promising drugs [206]. More recently, the antitumor and antiviral [207] activity of fluoroquinolones have been reported, which if proven will further extend their potential fields of application. Thus, the interest measured in terms of the number of papers published on fluoroquinolones during last twenty years is steadily growing (Fig. (14)). Mugnaini *et al.* reviewed the 4-quinolone-3-carboxylic scaffold as a core moiety for the preparation of targeted libraries in the search for antitumor, anxiolytic and antiviral agents [208].

Isoquinolines

Opium alkaloids may compete with quinine in terms of their importance in the history of medicine. Today, the use of opium poppy alkaloids is restricted by law; however, some of them still have an impact on the medicines. The most important marketed drug based on the isoquinoline moiety is papaverine; however, it is not closely related to other opium alkaloids and has no narcotic activity. It has a vasodilating activity and is used in the treatment of cerebral ischemia [209-212] and erectile dysfunction [213, 214].

The isoquinoline moiety can be identified in the antihypertension agents quinapril and moexipril [215]. Our analysis identified a considerable number of drug candidates with registered experimental data in the Drugbank that have the isoquinoline moiety.

Quinazolines

The quinazoline scaffold is the second most frequent scaffold among the drugs in the Drugbank. It is also virtually absent in natural sources. A closer look at the data in Table 5 and an analysis of the data from Drugbank led to some interesting conclusions. There are only 9 approved molecules that can be divided into two classes of drugs that are anticancer and adrenergic alpha antagonists. Only two drugs have been on the market for longer than 20 years (terazosin, prazosin). Thus, interest in this scaffold evidently appeared due to the economic success of a few other drugs that have been marketed recently. Undoubtedly, these are antitumor quinazolines that act as inhibitors of protein kinases. Currently, there are three drugs of this class: gefitinib (approved by the FDA in 2003), erlotinib (2004) and lapatinib (2007). See Fig. (15) [216-218].

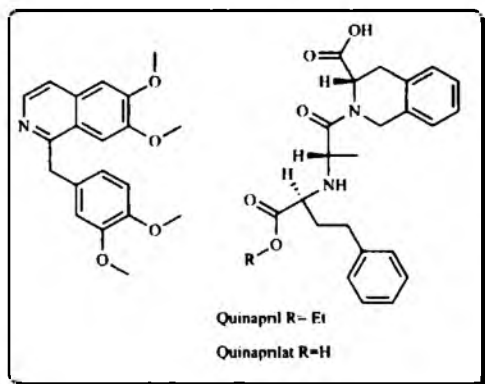
Noticeably, a major part of the potential drugs that have the quinazoline scaffold were designed to have antitumor activity.

Quinoxalines

This scaffold is represented by a relatively small group of drugs with a wide spectrum of applications. However, most of the experimental drugs were designed as inhibitors of tyrosine kinases [219]. Interestingly, the presence of the scaffold in natural sources is higher than those with scaffolds 3 or 4. There are also a considerable number of biologically active natural compounds in the Beilstein database. Echinomycin and related compounds are probably the best known [220-222].

Azennaphthalenes Under Clinical Trials

Clinical trials are essential for bringing drugs onto the market. They are also the last and decisive step in the extensive investigations of drug candidates, which comprises a lengthy period



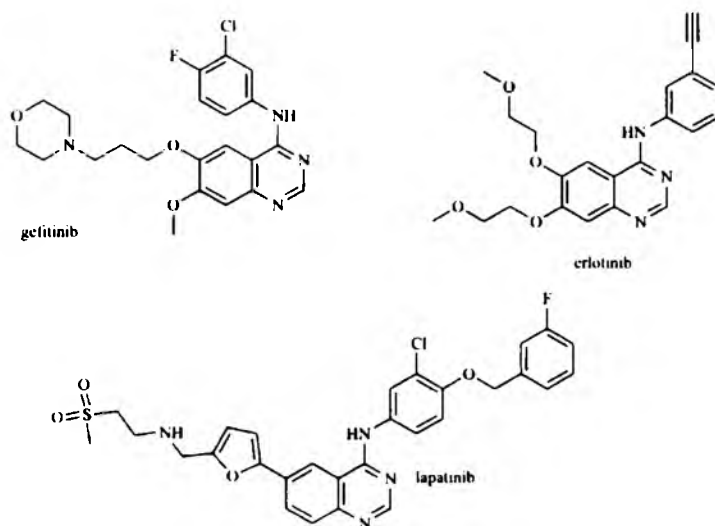
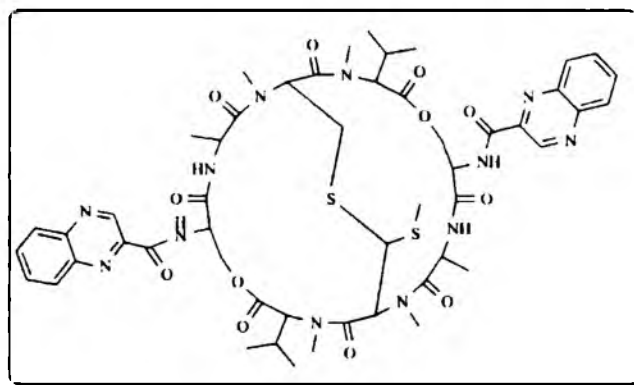
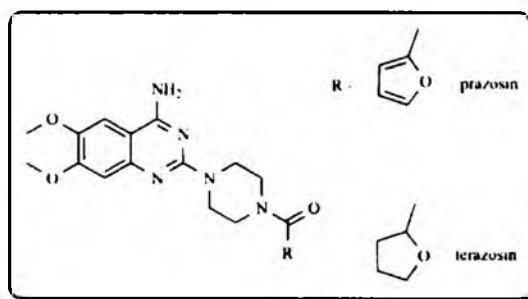
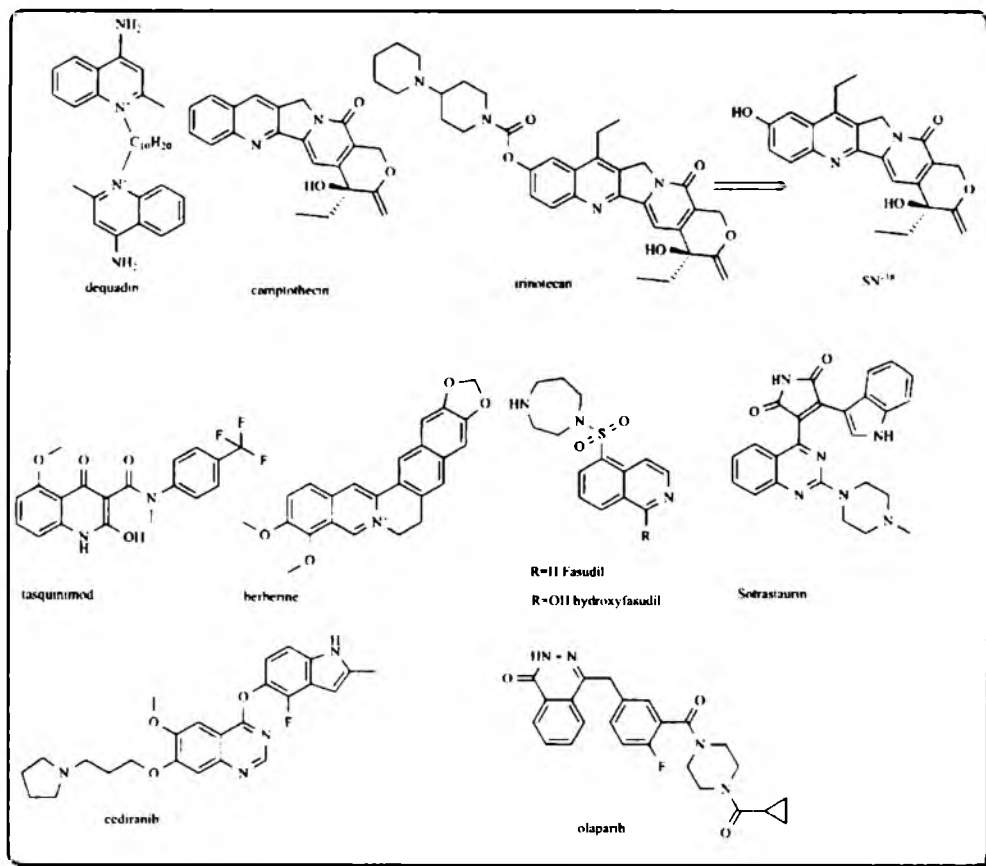


Fig. (15). Quinazoline based inhibitors of protein kinases



of research and often thousands of compounds being pre-evaluated. Thus, every entity under clinical trials stands for an entire class of active and potentially valuable compounds. In the case of me-too

drugs or drugs of subsequent generations, the group of tested structures is even larger. Thus, an insight into the most recent trends of clinical investigations of azanaphthalenes provides us with the



unique possibility to further evaluate the preferential organization of drug-like molecules discussed herein.

Garenoxacin and delafloxacin (RX-3341) are two quinolone antibiotics newly introduced into clinical trials. These drugs will expand the known library of the more than 30 marketed agents used for treating bacterial infections in humans and animals [199, 201, 203]. As was mentioned above, the number of described compounds from this class should be counted in thousands. Another antibacterial quinoline under clinical trial is dequadin (Fluomizine). The drug has recently finished the third phase investigation for the treatment of bacterial vaginosis [223, 224].

Camptothecin is a quinoline-related alkaloid that is isolated from the happy tree (*Camptotheca acuminata* – an Asian originated tree from the dogwood family). It is known for its high anticancer activity through the inhibition of DNA topoisomerase I [225-227]. Free camptothecin is toxic and has a poor pharmacokinetic profile making it unsuitable for oral administration. Nevertheless, due to its superior activity, it has been tested in numerous clinical trials, either being investigated alone or in various compositions or formulations of nanoparticles [228, 229], complexes [230, 231] and prodrugs [232].

A synthetic derivatisation of camptothecin has led to improved compounds and two of them (irinotecan and topotecan) have already been approved for the treatment of cancer [233, 234].

Irinotecan is metabolized in the system by means of hydrolysis to SN-38, which is a more potent inhibitor of topoisomerase than irinotecan [235]. Its direct use, however, is restricted by its poor solubility. Diverse approaches have been used to improve its usability for therapy and several SN-38 related drugs are currently under clinical examinations. These are SN-2310, an injectable emulsion of [236], EZN-2208 and NKTR-102 whose are pegylated SN-38 [237-240], LE-SN-38; drug encapsulated in liposomes [241]. Due to number of trials and positive results, compounds related to camptothecin are probably one of the most important anticancer agents that have been investigated recently. Another quinoline-related drug that is under clinical examination is tasquinimod (ABR-215050) – a potent antitumor agent that is in phase II of a study on the treatment of prostate cancer [242-244].

Berberine – an alkaloid isolated from plants of *Berberidaceae* family and present in numerous natural sources has its place in traditional medicine, and is believed to be a remedy for fungal and bacterial infections [245, 246] as well as for cancer [247, 248]. Current clinical trials are concentrated on the use of berberine in the treatment of polycystic ovary syndrome and on its effects on diabetes. Although a berberine-containing nutraceutical (Arnolipid) is currently on the market, the clinical trials are still under way in the hope that this interesting alkaloid can be introduced into medical treatment [249]. It is noteworthy that the mechanism of action of berberine may affect several important

metabolic pathways, e.g. reducing cholesterol and other triglycerides but also influencing glucose metabolism and reducing the HbA1c level [250-252]. The antimicrobial activity of berberine that changes the bacterial flora in the gastrointestinal tract [253] may inspire its further application in pharma treatment. The spectrum of activity of berberine is wide and further investigations are needed to explore its use in treating mental diseases or viral infections [254-256].

Another intensively investigated drug that is built on the isoquinoline moiety is Fasudil. It is a sulfodiazepane derivative of isoquinoline with potent activity against Rho-kinase and thus exhibits vasodilatory effects [257-260]. Current investigations are focused on the effects of Fasudil on the cardiovascular system and in the treatment of Raynaud's phenomenon. The latter is a vasospastic disorder manifested by discoloration of fingers or toes after internal or external factors as such cold or stress [261-262]. Vasodilators have been extensively used in the treatment of Raynaud's phenomenon so Fasudil may become effective as well. Fasudil is metabolised into hydroxyfasudil, which is also active vasodilator, in the human system. This compound is an experimental drug in the same area [263]. It has also been asserted improving learning working memory and cognitive processes [264].

Quinazolines are also represented by some compounds that are under clinical trials. Sotrastaurin (AEB071) is a potent inhibitor protein kinase C (PKC) that attenuates the activation of lymphocyte T immunoreactive cells [265, 266]. The drug is now under phase II trials for its safety, tolerability and pharmacokinetics in two large studies focused on renal and liver transplantations [267-269]. Another example of quinazoline inhibitors of protein kinases is cediranib (Recentin), which is a derivative of known anticancer inhibitors of kinases such as gefitinib or lapatinib [217, 270]. Its molecular target is the VEGF (vascular endothelial growth factor) receptor tyrosine kinases. There have been number of clinical trials undertaken with this drug in the treatment of various cancers as well as its pharmacokinetics [271-275]. The scope of the investigations conducted is broad and encompasses prostate, lung, colorectal, breast and ovarian cancers, glioblastomas and others. However, some issues have been reported such as the recurrence of a glioblastoma after the therapy [276] and some trials have been terminated due to slow or even no progress. Nevertheless, cediranib remains a promising agent in future anticancer therapy.

2,3-Diazanaphthalene (scaffold 6) is represented by one investigational drug – olaparib (AZD-2281 or KU-0059436). This is an inhibitor of poly (ADP-ribose) polymerase (PARP), which is built on the phthalazin-1-one moiety that was introduced into clinical trials by AstraZeneca [277, 278]. PARPs are a family of enzymes involved in several cellular mechanisms such as DNA repair, gene amplification and transcriptional regulation [279, 280]. After damage to DNA, these enzymes (mainly PARP-1) quickly bind to the breaks and synthesise poly(ADP-ribose) chains (PAR), which are essential in the repair process [279-281]. Olaparib was successfully tested for its anticancer activity in BRCA1 and 2 mutants as well as in other cancers as prostate or colorectal [282-287].

CONCLUSIONS

The PS concept evolved from a pragmatic tendency to simplify the complexity of drug design through fragmentation. After all, breaking molecules into fragments is a well-known procedure in chemistry, e.g. organic chemists discuss molecular reactivity in terms of functional groups, particularly small parts of molecules of a specific reactivity, while retro-synthetic analysis breaks molecules down into synthons that are useful in synthesizing target compounds. Similarly, the Hammett, Hansch or Free-Wilson approaches can be discussed in the categories of fragmental approaches to molecular reactivity or biological activity. The PS

method can be seen as a prototype and an offspring of fragonomics, which evolved from the experimental measurements of protein-ligand interactions. Formally, it is ill defined on the basis of identifying similarities between drugs. In reality, it is not easy to conclude whether a real preference exists and to describe this preference in numbers. Our attempt to analyze the preferential arrangement of azanaphthalenes based on database mining revealed some clear relationships, but large differences in the certain population of individual scaffolds allow for no clear conclusions.

In the language of fragonomics, PS can be defined as the identification of building blocks generally following the Murcko and Bemis fragmentation rules. It is often used in synthetic laboratories or combinatorial chemistry approaches where we need new but simple and clear design schemes. Despite all these limitations, the PS concept appears to be successful, thus proving the power of analogy. Inter-penetration of fragment-based design and PS methods can substantially contribute to this area by providing new ideas for drug design.

CONFLICT OF INTEREST

None declared.

ACKNOWLEDGEMENTS

AK and AB acknowledge a scholarship from the UPGOW projects co-financed by the European Social Fund.

REFERENCE

- [1] Kubinyi, H. *Privileged Structures and Analogue-Based Drug Discovery*. In *Analogue-Based Drug Discovery*; Fischer, J.; Gianellin, C.R., Eds.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, 2006; pp. 53-68.
- [2] Barelier, S.; Krimm, I. Ligand specificity, privileged substructures, and protein druggability from fragment-based screening. *Curr. Opin. Chem. Biol.* 2011, 15, 469-474.
- [3] Klekota, J.; Lundell, G. Chemical substructures that enrich for biological activity. *Bioinformatics* 2008, 24, 2518-2525.
- [4] Baum, R.M. The top pharmaceuticals that changed the world. *Chem. Eng. News* 2005, 83, p 5.
- [5] Lloyd, N.C.; Morgan, H.; W. Nicholson, B.K.; Rommus, R.S. The composition of Ehrlich's salvarsan: resolution of a century-old debate. *Angew. Chem. Int. Ed. Engl.* 2005, 44, 941-944.
- [6] Domagk, G. Further progress in chemotherapy of bacterial infections. In: *Nobel Lectures, Physiology or Medicine 1922-1941*, Nobel Lecture, December 12, 1947; Elsevier Publishing Company: Amsterdam, 1965; pp. 490-527.
- [7] Fialkowski, M.; Bishop, K.J.M.; Chubukov, V.A.; Campbell, C.J.; Grzybowski, B.A. Architecture and evolution of organic chemistry. *Angew. Chem. Int. Ed. Engl.* 2005, 44, 7263-7269.
- [8] Grzybowski, B.A.; Bishop, K.J.M.; Kowalczyk, B.; Wilmer, C.F. The "wired" universe of organic chemistry. *Nat. Chem.* 2009, 1, 31-36.
- [9] Chen, X.; Fang, Y.; Yao, L.; Chen, Y.; Xu, H. Does drug-target have a likeness? *Methods Inf. Med.* 2007, 46, 360-366.
- [10] Grabowski, K.; Schneider, G. Properties and architecture of drugs and natural products revisited. *Curr. Chem. Biol.* 2007, 1, 115-127.
- [11] Bemis, G.W.; Murcko, M.A. The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* 1996, 39, 2887-2893.
- [12] Bemis, G.W.; Murcko, M.A. Properties of known drugs. 2. Side chains. *J. Med. Chem.* 1999, 42, 5095-5099.
- [13] Agrafiotis, D.K.; Wiener, J.J. Scaffold explorer: an interactive tool for organizing and mining structure-activity data spanning multiple chemotypes. *J. Med. Chem.* 2010, 53, 5002-5011.
- [14] Schurer, S.; Tsinober, N. Screening Informatics. In *A Practical Guide to Assay Development and High-Throughput Screening in Drug Discovery*; Chen, T., Ed.; CRC Press, 2009; pp. 233-263.
- [15] de Sa Alves, F.R.; Barreiro, E.J.; Fraga, C.A.M. From nature to drug discovery: the indole scaffold as a "privileged structure". *Mini Rev. Med. Chem.* 2009, 9, 782-793.
- [16] Evans, B.; Rittle, K.; Buck, M.; DiPardo, R.; Freidinger, R.; Whitter, W.; Lundell, G.; Veber, D.; Anderson, P. Methods for drug discovery development of potent, selective, orally effective cholecystokinin antagonists. *J. Med. Chem.* 1988, 31, 2235-2246.
- [17] Roche, O.; Schneider, P.; Zuegg, J.; Guba, W.; Kansy, M.; Alanine, A.; Blecher, K.; Danel, F.; Grubnecht, E.M.; Rogers-Evans, M.; Neidhart, W.; Stalder, H.; Dillon, M.; Sjogren, E.; Fotouhi, N.; Gillespie, P.; Goodnow, R.; Harris, W.; Jones, P.; Taniguchi, M.; Tsuji, S.; von der Saal, W.

- Zimmermann, G.; Schneider, G.J. Development of a virtual screening method for identification of "frequent hitters" in compound libraries. *J. Med. Chem.* **2002**, *45*, 137-142.
- [18] Lipinski, C.; Hopkins, A. Navigating chemical space for biology and medicine. *Nature*, **2004**, *432*, 855-861.
- [19] Dobson, C.M. Chemical space and biology. *Nature*, **2004**, *7019*, 824-828.
- [20] Bohacek, R.S.; McMartin, C.; Guida, W.C. The art and practice of structure-based drug design: a molecular modeling perspective. *Med. Res. Rev.*, **1996**, *16*, 3-50.
- [21] Ganesan, A. The impact of natural products upon modern drug discovery. *Curr. Opin. Chem. Biol.*, **2008**, *12*, 306-317.
- [22] Merlot, C.; Domine, D.; Cleve, C.; Church, D.J. Chemical substructures in drug discovery. *Drug Discov. Today*, **2003**, *8*, 594-602.
- [23] Davies, T.G.; van Montfort, R.L.M.; Williams, G.; Jhoti, H. Pyramid: An Integrated Platform for Fragment-Based Drug Discovery. In: *Fragment-Based Approaches in Drug Discovery*; Jahnke, W.; Erlanson, D.A., Eds.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, **2006**; pp. 193-214.
- [24] Ajay, A.; Walters, W.P.; Murcko, M.A. Can we learn to distinguish between "drug-like" and "nondrug-like" molecules? *J. Med. Chem.*, **1999**, *41*, 3314-3324.
- [25] Sadowski, J.; Kubinyi, H. A scoring scheme for discriminating between drugs and nondrugs. *J. Med. Chem.*, **1998**, *41*, 3325-3329.
- [26] Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.*, **2001**, *46*, 3-26.
- [27] Veber, D.F.; Johnson, S.R.; Cheng, H.Y.; Smith, B.R.; Ward, K.W.; Kopple, K.D. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.*, **2002**, *45*, 2615-2623.
- [28] Fukunishi, Y.; Nakamura, H. Definition of drug-likeness for compound affinity. *J. Chem. Inf. Model.*, **2011**, *51*, 1012-1016.
- [29] Ursu, O.; Oprea, T.I. Model-free drug-likeness from fragments. *J. Chem. Inf. Model.*, **2010**, *50*, 1387-1394.
- [30] Hann, M.M.; Oprea, T.I. Pursuing the leadlikeness concept in pharmaceutical research. *Curr. Opin. Chem. Biol.*, **2006**, *10*, 255-263.
- [31] Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A "rule of three" for fragment-based lead discovery? *Drug Discov. Today*, **2003**, *8*, 876-877.
- [32] Lobell, M.; Hendrix, M.; Hünzler, B.; Keldenich, J.; Meier, H.; Schmeck, C.; Schoch-Loop, R.; Wunberg, T.; Hillisch, A. In silico ADMET traffic lights as a tool for the prioritization of HTS hits. *ChemMedChem*, **2006**, *1*, 1229-1236.
- [33] Cao, Y.; Jiang, T.; Girke, T. A maximum common substructure-based algorithm for searching and predicting drug-like compounds. *Bioinformatics*, **2008**, *24*, 1366-1374.
- [34] Vistoli, G.; Pedretti, A.; Testa, B. Assessing drug-likeness – what are we missing? *Drug Discov. Today*, **2008**, *13*, 285-294.
- [35] Todeschini, R.; Consonni, V. *Molecular Descriptors for Chemoinformatics*, Vol. 4; Wiley-VCH Verlag GmbH: Weinheim, **2009**.
- [36] Kolb, P.; Caffisch, A. Automatic and efficient decomposition of two-dimensional structures of small molecules for fragment-based high-throughput docking. *J. Med. Chem.*, **2006**, *49*, 7384-7392.
- [37] Ghose, A.K.; Herbertz, T.; Pippin, D.A.; Salvino, J.M.; Mallamo, J.P. Knowledge based prediction of ligand binding modes and rational inhibitor design for kinase drug discovery. *J. Med. Chem.*, **2008**, *51*, 5149-5171.
- [38] Erlanson, D.A.; McDowell, R.S.; O'Brien, T. Fragment-based drug discovery. *J. Med. Chem.*, **2004**, *47*, 3463-3482.
- [39] Rabal, O.; Urbano-Cuadrado, M.; Oyarzabal, J. Computational medicinal chemistry in fragment-based drug discovery: what, how and when. *Future Med. Chem.*, **2011**, *3*, 95-134.
- [40] Borgelt, C.; Hofer, H.; Berthold, M. Finding Discriminative Molecular Fragments. In: *Workshop Information Mining - Navigating Large Heterogeneous Spaces of Multimedia Information*, Proceedings of the 26th German Conference on Artificial Intelligence, Hamburg, Germany, September 15-18, **2003**.
- [41] Lameyer, E.W.; Kok, J.N.; Back, T.; Ijzerman, A.P. Mining a chemical database for fragment co-occurrence: discovery of "chemical clichés". *J. Chem. Inf. Model.*, **2006**, *46*, 553-562.
- [42] Gianti, E.; Santori, L. Identification and selection of "privileged fragments" suitable for primary screening. *J. Chem. Inf. Model.*, **2008**, *48*, 2129-2139.
- [43] Lowell, X.Q.; Judd, D.B.; Watson, S.P.; Hann, M.M. 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.*, **1999**, *39*, 511-522.
- [44] Polanski, J. Receptor dependent multidimensional QSAR for modeling drug-receptor interactions. *Curr. Med. Chem.*, **2009**, *16*, 3243-3257.
- [45] Doweyko, A.M. QSAR: dead or alive? *J. Comput. Aided Mol. Des.*, **2000**, *22*, 81-89.
- [46] Tropsha, A. Recent Advances in Development, Validation and Exploitation of QSAR Models. In: *Burger's Medicinal Chemistry, Drug Discovery and Development*, John Wiley & Sons, **2010**; pp. 1-30.
- [47] Wagener, M.; van Geerestein, V.J. Potential drugs and nondrugs: prediction and identification of important structural features. *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 280-292.
- [48] Chen, H.; Yang, Y.; Engkvist, O. Molecular topology analysis of the differences between drugs, clinical candidate compounds, and bioactive molecules. *J. Chem. Inf. Model.*, **2010**, *50*, 2141-2150.
- [49] Oprea, T.I. Property distribution of drug-related chemical databases. *J. Comput. Aided Mol. Des.*, **2000**, *14*, 251-264.
- [50] Ghose, A.K.; Herbertz, T.; Salvino, J.M.; Mallamo, J.P. Knowledge-based chemoinformatic approaches to drug discovery. *Drug Discov. Today*, **2006**, *11*, 1107-1114.
- [51] Leeson, P.D.; Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discov.*, **2007**, *6*, 881-890.
- [52] Costantino, L.; Barlicco, D. Privileged structures as leads in medicinal chemistry. *Curr. Med. Chem.*, **2006**, *13*, 65-85.
- [53] Clark, A.M.; Labute, P. Detection and assignment of common scaffolds in project databases of lead molecules. *J. Med. Chem.*, **2009**, *52*, 469-483.
- [54] Schulz, M.N.; Hubbard, R.E. Recent progress in fragment-based lead discovery. *Curr. Opin. Pharmacol.*, **2009**, *9*, 615-621.
- [55] Duarte, C.D.; Barreiro, E.J.; Fraga, C.A.M. Privileged structures: a useful concept for the rational design of new lead drug candidates. *Mini Rev. Med. Chem.*, **2007**, *7*, 1108-1119.
- [56] Oprea, T.I. Current trends in lead discovery: are we looking for the appropriate properties? *J. Comput. Aided Mol. Des.*, **2002**, *16*, 325-334.
- [57] Oprea, T.I.; Allu, T.K.; Fara, D.C.; Rad, R.F.; Ostropovici, L.; Bologna, C.G. Lead-like, drug-like or "Pub-like": how different are they? *J. Comput. Aided Mol. Des.*, **2010**, *21*, 113-119.
- [58] Fattori, D. Molecular recognition: the fragment approach in lead generation. *Drug Discov. Today*, **2004**, *9*, 229-238.
- [59] Byvatov, E.; Schneider, G. SVM-based feature selection for characterization of focused compound collections. *J. Chem. Inf. Comput. Sci.*, **2004**, *44*, 993-999.
- [60] Beroza, P.; Damodaran, K.; Lum, R.T. Target-related affinity profiling: Telik's lead discovery technology. *Curr. Top. Med. Chem.*, **2005**, *5*, 371-381.
- [61] Johnson, M.A.; Maggiora, G.M. *Concepts and applications of molecular similarity*, John Wiley & Sons: New York, **1990**.
- [62] Bajorath, J. Virtual screening: methods, expectations and reality. *Curr. Drug Discov.*, **2002**, *2*, 24-28.
- [63] Bajorath, J. Integration of virtual and high-throughput screening. *Nat. Rev. Drug Discov.*, **2002**, *1*, 882-894.
- [64] Kubinyi, H. Chemical similarity and biological activities. *J. Braz. Chem. Soc.*, **2002**, *13*, 717-726.
- [65] Jaworski, J.; Nikolova-Jeliazkova, N. How can structural similarity analysis help in category formation? *SAR QSAR Environ. Res.*, **2007**, *18*, 195-207.
- [66] DeSimone, R.W.; Currie, K.S.; Mitchell, S.A.; Darrow, J.W.; Pippin, D.A. Privileged structures: applications in drug discovery. *Comb. Chem. High Throughput Screen.*, **2004**, *7*, 473-494.
- [67] Patchett, A.A.; Nargund, R.P. Privileged structures – an update. *Annu. Rev. Med. Chem.*, **2000**, *35*, 289-298.
- [68] Schuur, D.M.; Hermameier, M.A.; Tebben, A.J. Are target-family-privileged substructures truly privileged? *J. Med. Chem.*, **2006**, *49*, 2000-2009.
- [69] Nicolaou, C.A.; Patchett, C.S. Molecular Substructure Mining Approaches for Computer-Aided Drug Discovery: A Review. In: *Proceedings of ITAB: Ioannina, Greece, October 26-28, 2006*.
- [70] Bondensgaard, K.; Andersen, M.; Thøgersen, H.; Hansen, B.S.; Wulff, B.S.; Bywater, R.P. Recognition of privileged structures by G-protein coupled receptors. *J. Med. Chem.*, **2004**, *47*, 888-899.
- [71] Horton, D.A.; Bourne, G.T.; Smythe, M.L. The combinatorial synthesis of bicyclic privileged structures or privileged substructures. *Chem. Rev.*, **2003**, *103*, 893-930.
- [72] Che, Y.; Marshall, G.R. Privileged scaffolds targeting reverse-turn and helix recognition. *Expert Opin. Ther. Targets*, **2008**, *12*, 101-114.
- [73] Austin, J.F.; MacMillan, D.W.C. Enantioselective organocatalytic indole alkylations. Design of a new and highly effective chiral amine for iminium catalysis. *J. Am. Chem. Soc.*, **2002**, *124*, 1172-1173.
- [74] Hajduk, P.J.; Bures, M.; Prasadgaard, J.; Fesik, S.W. Privileged molecules for protein binding identified from NMR-based screening. *J. Med. Chem.*, **2000**, *43*, 3443-3447.
- [75] Barakat, K.J.; Cheng, K.; Chan, W.W.; Butler, B.S.; Jacks, T.M.; Schleim, K.D.; Hora, D.F.; Hickey, G.J.; Smith, R.G.; Patchett, A.A.; Nargund, R.P. Synthesis and biological activities of phenyl piperazine-based peptidomimetic growth hormone secretagogues. *Bioorg. Med. Chem. Lett.*, **1998**, *8*, 1431-1436.
- [76] Jacobson, K.A.; Kim, Y.C.; Kim, B.F. In search of selective P2 receptor ligands: interaction of dihydropyridine derivatives at recombinant rat P2X(2) receptors. *J. Auton. Nerv. Syst.*, **2000**, *81*, 152-157.
- [77] Converso, A.; Burrow, K.; Martinzik, A.; Sharpless, K.B.; Finn, M.G. 2,6-Dichloro-9-*thiabicyclo[3.3.1]nonane*: a privileged bivalent scaffold for the display of nucleophilic components. *J. Org. Chem.*, **2001**, *66*, 4386-4392.
- [78] Nicolaou, K.C.; Pfefferkorn, J.A.; Roecker, A.J.; Cao, G.Q.; Barluenga, S.; Mitchell, H.J. Natural product-like combinatorial libraries based on privileged structures. I. General principles and solid-phase synthesis of benzopyrans. *J. Am. Chem. Soc.*, **2000**, *122*, 9939-9953.
- [79] Nicolaou, K.C.; Roecker, A.J.; Barluenga, S.; Pfefferkorn, J.A.; Cao, G.Q. Discovery of novel antibacterial agents active against methicillin-resistant *Staphylococcus aureus* from combinatorial benzopyran libraries. *ChemBiochem*, **2001**, *2*, 460-465.
- [80] Wermuth, C.G. Are pyridazines privileged structures? *Medchemcomm*, **2011**, *2*, 915-941.
- [81] Musiol, R.; Serda, M.; Hensel-Bielowska, S.; Polanski, J. Quinoline-based

- antifungals. *Curr Med Chem*, **2010**, *17*, 1960-1973.
- [82] Sokunm, V.R.; Lee, H. Quinoline as a privileged scaffold in cancer drug discovery. *Curr Med Chem*, **2011**, *18*, 1488-1508.
- [83] Rosenthal, J. Nifedipine: profile of a new calcium antagonist. An overview. *J Cardiovasc Pharmacol*, **1994**, *24* Suppl 2, S92-107.
- [84] Gathwala, G.; Goel, M.; Singh, J.; Mittal, K. Intravenous Diazepam, Midazolam and Lorazepam in acute seizure control. *Indian J Pediatr*, **2011**, Epub ahead of print.
- [85] Fukunaga, M.; Ishizawa, K.; Karret, C. Anticonvulsant properties of 1,4-benzodiazepine derivatives in amygdaloid-kindled seizures and their chemical structure-related anticonvulsant action. *Pharmacology*, **1998**, *57*, 233-241.
- [86] Dodd, T.; Doyle, T. Cyclozine. *Emerg Med J*, **2003**, *20*, 302.
- [87] Babu, T.A. Bucizone is back again! This time as a pediatric appetite stimulant. *Indian J Pharmacol*, **2011**, *43*, 219.
- [88] Guanana, G.; Barbu, C.; Cipriani, A. Hydroxyzine for generalised anxiety disorder. *Cochrane Database Syst Rev*, **2010**, *8*, CD006815.
- [89] Bachert, C.; Maspero, J. Efficacy of second-generation antihistamines in patients with allergic rhinitis and comorbid asthma. *J Asthma*, **2011**, *48*, 965-973.
- [90] Taira, N. Nifedipine: a novel vasodilator. *Drugs*, **2006**, *66*, 1-3.
- [91] Kuwajima, I. Antihypertensive effects of nifedipine and changing for modified-release formula. *Drugs*, **2006**, *66*, 7-9.
- [92] Peacock, W.F.; Hilleman, D.E.; Levy, P.D.; Rhoney, D.H.; Varon, J. A systematic review of nifedipine vs labetalol for the management of hypertensive crises. *Am J Emerg Med*, **2011**, Epub ahead of print.
- [93] Matsuda, H.; Araki, N.; Kuji, I.; Okubo, T.; Imabayashi, E.; Shimazu, K. Effect of nifedipine on regional cerebral blood flow in a patient with early Alzheimer disease. *Clin Nucl Med*, **2008**, *33*, 34-35.
- [94] Gradman, A.H. The evolving role of calcium channel blockers in the treatment of angina pectoris: focus on felodipine. *Can J Cardiol*, **1995**, *11* Suppl B, 14B-21B.
- [95] Maksimov, M.L.; Mochkin, I.A.; Starodubtsev, A.K. Application of AT1-angiotensin II receptor blocker valsartan in clinical practice. *Kardiologia*, **2011**, *51*, 77-84.
- [96] Collaboration, I.A.T. Effects of telmisartan, irbesartan, valsartan, candesartan, and losartan on cancers in 15 trials enrolling 138,769 individuals. *J Hypertens*, **2011**, *29*, 623-635.
- [97] Vachhani, R.; Olds, G.; Velanovich, V. Esomeprazole: a proton pump inhibitor. *Expert Rev Gastroenterol Hepatol*, **2009**, *3*, 15-27.
- [98] Dorman, S.; Perkins, P. Droperidol for treatment of nausea and vomiting in palliative care patients. *Cochrane Database Syst Rev*, **2010**, *6*, CD006938.
- [99] Canete, R.; Escobedo, A.A.; Almirall, P.; Gonzalez, M.E.; Brito, K.; Cimerman, S. Mebendazole in parasitic infections other than those caused by soil-transmitted helminths. *Trans R Soc Trop Med Hyg*, **2009**, *103*, 437-442.
- [100] Murota, H.; Katayama, I. Emedastine difumarate: a review of its potential ameliorating effect for tissue remodeling in allergic diseases. *Expert Opin Pharmacother*, **2009**, *10*, 1859-1867.
- [101] Dowson, A.J.; Charlesworth, B. Review of zolmitriptan and its clinical applications in migraine. *Expert Opin Pharmacother*, **2002**, *3*, 993-1005.
- [102] Dvorsnik, D.M. Tissue selective inhibition of prostaglandin biosynthesis by endothelium. *J Rheumatol Suppl*, **1997**, *47*, 40-47.
- [103] Foote, A.; Hannon, D.L.; Strickland, J.R.; Bush, L.P.; Klotz, J.L. Effect of ergot alkaloids on contractility of bovine right ruminal artery and vein. *J Anim Sci*, **2011**, *49*, 2944-2949.
- [104] Bongarzoni, S.; Bolognesi, M.L. The concept of privileged structures in rational drug design: focus on acridine and quinoline scaffolds in neurodegenerative and protozoan diseases. *Expert Opin Drug Discov*, **2011**, *6*, 251-268.
- [105] McGovern, S.L.; Heltand, B.T.; Feng, B.; Shoichet, B.K. A specific mechanism of nonspecific inhibition. *J Med Chem*, **2003**, *46*, 4265-4272.
- [106] Liu, Z.; Chen, X.; Yu, L.; Zhen, X.; Zhang, A. Synthesis and pharmacological investigation of novel 2-aminothiazole-privileged aporphines. *Bioorg Med Chem*, **2008**, *16*, 6675-6681.
- [107] Welsch, M.E.; Snyder, S.A.; Stockwell, B.R. Privileged scaffolds for library design and drug discovery. *Curr Opin Chem Biol*, **2010**, *14*, 347-361.
- [108] Willoughby, C.A.; Hutchins, S.M.; Rosauer, K.G.; Dhar, M.J.; Chapman, K.T.; Checchi, G.G.; Sadowski, S.; Weinberg, D.H.; Patel, S.; Malkowitz, L.; Di Salvo, J.; Pacholik, S.G.; Cheng, K. Combinatorial synthesis of 3-(amidoalkyl) and 3-(aminoalkyl)-2-arylindole derivatives: discovery of potent ligands for a variety of G-protein coupled receptors. *Bioorg Med Chem Lett*, **2002**, *12*, 93-96.
- [109] Chang, Y.T.; Gray, N.S.; Rosania, G.R.; Sutherland, D.P.; Kwon, S.; Norman, T.C.; Sarabia, R.; Leost, M.; Meyer, L.; Schultz, P.G. Synthesis and application of functionally diverse 2,6,9-trisubstituted purine libraries as CDK inhibitors. *Chem Biol*, **1999**, *6*, 361-375.
- [110] Lee, J.Y.; Kim, Y.C. Combinatorial library synthesis and biological evaluation of pyrazolo[4,3-e][1,4]diazepine as a potential privileged structure. *ChemMedChem*, **2009**, *4*, 733-737.
- [111] Siegel, M.G.; Vieti, M. Drugs in other drugs: a new look at drugs as fragments. *Drug Discov Today*, **2007**, *12*, 71-79.
- [112] DiMasi, J.A.; Hansen, R.W.; Grabowski, H.G. The price of innovation: new estimates of drug development costs. *J Health Econ*, **2003**, *22*, 151-185.
- [113] Bohm, H.; Flohr, A.; Stahl, M. Scaffold hopping. *Drug Discov Today Technol*, **2004**, *1*, 217-224.
- [114] Garattini, S. Are me-too drugs justified? *J Nephrol*, **1997**, *10*, 283-294.
- [115] Hajduk, P.J.; Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat Rev Drug Discov*, **2007**, *6*, 211-219.
- [116] Sun, C.; Petros, A.M.; Hajduk, P.J. Fragment-based lead discovery: challenges and opportunities. *J Comput Aided Mol Des*, **2011**, *25*, 607-610.
- [117] Zartler, E.R.; Shapiro, M.J. Fragonomics: fragment-based drug discovery. *Curr Opin Chem Biol*, **2005**, *9*, 366-370.
- [118] Hans, M.M.; Leach, A.R.; Harper, G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *J Chem Inf Comput Sci*, **2001**, *41*, 856-864.
- [119] Kubinyi, H. *QSAR: Hansch Analysis and Related Approaches*; Wiley-VCH Verlag GmbH: Weinheim, **1993**.
- [120] Rekker, R.F.; Mannhold, R. *Calculation of Drug Lipophilicity - The Hydrophobic Fragmental Constant Approach*; Wiley-VCH Verlag GmbH: Weinheim, **1992**.
- [121] Kuz'min, V.F.; Artemenko, A.G.; Muratov, E.N.; Polischuk, P.G.; Ognichenko, L.N.; Lishovsky, A.V.; Hromov, A.I. Virtual Screening and Molecular Design Based on Hierarchical QSAR Technology. In: *Recent Advances in QSAR Studies: Methods and Applications*; Puzyn, T.; Leszczynski, J.; Cronin, M., Eds.; Springer, **2010**; pp 127-176.
- [122] Polanski, J. Neural nets for the simulation of molecular recognition within MS-Windows environment. *J Chem Inf Comput Sci*, **1996**, *36*, 694-705.
- [123] Polanski, J. The receptor-like neural network for modeling corticosteroid and testosterone binding globulins. *J Chem Inf Comput Sci*, **1997**, *37*, 553-561.
- [124] Klopman, G. Artificial intelligence approach to structure-activity studies. Computer automated structure evaluation of biological activity of organic molecules. *J Am Chem Soc*, **1984**, *106*, 7315-7321.
- [125] Klopman, G.; Kalos, A.N. Quantitative structure-activity relationships of beta-adrenergic agents. Application of the computer automated structure evaluation (CASE) technique of molecular fragment recognition. *J Theor Biol*, **1986**, *118*, 199-214.
- [126] Klopman, G.; Venegas, R.E. Computer automated structure evaluation (CASE) of flavonoids as larval growth inhibitors. *Pure Appl Chem*, **1988**, *60*, 265-270.
- [127] Sutherland, J.J.; Higga, R.E.; Watson, I.; Vieti, M. Chemical fragments as foundations for understanding target space and activity prediction. *J Med Chem*, **2008**, *51*, 2689-2700.
- [128] van der Horst, E.; Ijzerman, A.P. Computational Approaches to Fragment and Substructure Discovery and Evaluation. In: *Fragment-Based Drug Discovery: A Practical Approach*, 1st ed.; Zartler, E.R.; Shapiro, M.J., Eds.; John Wiley & Sons, **2008**; pp 199-222.
- [129] Hurst, T.; Heritage, T. HQSAR - a highly predictive QSAR technique based on molecular holograms. In: *Book of Abstracts*, 213rd ACS National Meeting, San Francisco, the USA, April 13-17, **1997**.
- [130] Rodriguez, C.R.; Flaherty, T.M.; Springer, C.; McKerrow, J.H.; Cohen, F.E. CoMFA and HQSAR of acylhydrazide cruzan inhibitors. *Bioorg Med Chem Lett*, **2002**, *12*, 1537-1541.
- [131] Waller, C.L. A comparative QSAR study using CoMFA, HQSAR, and FRED/SKEYS paradigms for estrogen receptor binding affinities of structurally diverse compounds. *J Chem Inf Comput Sci*, **1999**, *44*, 758-765.
- [132] Clark, M. Generalized fragment-substructure based property prediction method. *J Chem Inf Model*, **2005**, *45*, 30-38.
- [133] Matter, H.; Baringhaus, K.H.; Naumann, T.; Klabunde, T.; Pirard, B. Computational approaches towards the rational design of drug-like compound libraries. *Comb Chem High Throughput Screen*, **2001**, *4*, 453-475.
- [134] Oprea, T.I.; Davis, A.M.; Teague, S.J.; Leeson, P.D. Is there a difference between leads and drugs? A historical perspective. *J Chem Inf Comput Sci*, **2001**, *41*, 1308-1315.
- [135] Aronov, A.M.; Bemis, G.W. A minimalist approach to fragment-based ligand design using common rings and linkers: application to kinase inhibitors. *Proteins*, **2004**, *57*, 36-50.
- [136] Aronov, A.M.; McClain, B.; Moody, C.S.; Murcko, M.A. Kinase-likeness and kinase-privileged fragments: toward virtual polypharmacology. *J Med Chem*, **2008**, *51*, 1214-1222.
- [137] Bender, A.; Musa, H.Y.; Glen, R.C.; Reiling, S. Molecular similarity searching using atom environments, information-based feature selection, and a naive Bayesian classifier. *J Chem Inf Comput Sci*, **2004**, *44*, 170-178.
- [138] Xing, L.; Glen, R.C. Novel methods for the prediction of logP, pK(a), and logD. *J Chem Inf Comput Sci*, **2002**, *42*, 796-805.
- [139] Schneider, G.; Fechner, U. Computer-based de novo design of drug-like molecules. *Nat Rev Drug Discov*, **2005**, *4*, 649-663.
- [140] Carhart, R.E.; Smith, D.H.; Venkataraghavan, R. Atom pairs as molecular features in structure-activity studies: definition and applications. *J Chem Inf Comput Sci*, **1985**, *25*, 64-73.
- [141] Schuur, D.; Beno, B.R.; Good, A.; Tebben, A. Approaches to target class combinatorial library design. *Methods Mol Biol*, **2004**, *275*, 355-378.
- [142] Tan, D.S. Diversity-oriented synthesis: exploring the intersections between chemistry and biology. *Nat Chem Biol*, **2005**, *1*, 74-84.
- [143] Kearsley, S.K.; Sallamack, S.; Fluder, E.M.; Andose, J.D.; Mosley, R.T.; Sheridan, R.P. Chemical similarity using physicochemical property

- descriptors. *J. Chem. Inf. Comput. Sci.*, **1996**, *36*, 118-127.
- [144] Gillet, V.J.; Myatt, G.; Zaidos, Z.; Johnson, A.P. SPROUT, HIPPO and CAESA: tools for de novo structure generation and estimation of synthetic accessibility. *Perspect. Drug Discov.*, **1998**, *3*, 34-50.
- [145] Graham, D.J.; Malarkey, C.; Schmelicher, M.V. Information content in organic molecules: quantification and statistical structure via Brownian processing. *J. Chem. Inf. Comput. Sci.*, **2004**, *44*, 1601-1611.
- [146] Batista, J.; Godden, J.W.; Bajorath, J. Assessment of molecular similarity from the analysis of randomly generated structural fragment populations. *J. Chem. Inf. Model.*, **2006**, *46*, 1937-1944.
- [147] Batista, J.; Bajorath, J. Mining of randomly generated molecular fragment populations uncovers activity-specific fragment hierarchies. *J. Chem. Inf. Model.*, **2007**, *47*, 1405-1413.
- [148] Lounkine, E.; Batista, J.; Bajorath, J. Mapping of activity-specific fragment pathways isolated from random fragment populations reveals the formation of coherent molecular cores. *J. Chem. Inf. Model.*, **2007**, *47*, 2133-2139.
- [149] Batista, J.; Bajorath, J. Chemical database mining through entropy-based molecular similarity assessment of randomly generated structural fragment populations. *J. Chem. Inf. Model.*, **2007**, *47*, 59-68.
- [150] Lounkine, E.; Batista, J.; Bajorath, J. Random molecular fragment methods in computational medicinal chemistry. *Curr. Med. Chem.*, **2008**, *15*, 2108-2121.
- [151] Maggiora, G.M.; Shanmugasundaram, V. Molecular Similarity Measures. *Methods Mol. Biol.*, **2004**, *275*, 1-50.
- [152] Bender, A.; Glen, R.C. Molecular similarity: a key technique in molecular informatics. *Org. Biomol. Chem.*, **2004**, *2*, 3204-3218.
- [153] Schuffenhauer, A.; Brown, N. Chemical diversity and biological activity. *Drug Discov. Today Technol.*, **2006**, *3*, 387-395.
- [154] Krier, M.; Bret, G.; Roggan, D. Assessing the scaffold diversity of screening libraries. *J. Chem. Inf. Model.*, **2006**, *46*, 512-524.
- [155] Hu, Y.; Wassermann, A.M.; Lounkine, E.; Bajorath, J. Systematic analysis of public domain compound potency data identifies selective molecular scaffolds across druggable target families. *J. Med. Chem.*, **2010**, *53*, 752-758.
- [156] Schuffenhauer, A.; Brown, N.; Fritl, P.; Jenkins, J.L.; Selzer, P.; Hamon, J. Clustering and rule-based classifications of chemical structures evaluated in the biological activity space. *J. Chem. Inf. Model.*, **2007**, *47*, 325-336.
- [157] Vieth, M.; Siegel, M.G.; Higgins, R.E.; Watson, J.A.; Robertson, D.H.; Savin, K.A.; Durst, G.L.; Lipskind, P.A. Characteristic physical properties and structural fragments of marketed oral drugs. *J. Med. Chem.*, **2004**, *47*, 224-232.
- [158] van Rhee, A.M. Use of recursion forests in the sequential screening process: consensus selection by multiple recursive trees. *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, 941-948.
- [159] Budin, N.; Majeux, N.; Caillach, A. Fragment-based flexible ligand docking by evolutionary optimization. *Biol. Chem.*, **2001**, *382*, 1365-1372.
- [160] Bilton, C.; Allen, F.H.; Shields, G.P.; Howard, J.A. Intramolecular hydrogen bonds: common motifs, probabilities of formation and implications for supramolecular organization. *Acta Crystallogr. B*, **2000**, *56*, 849-856.
- [161] Pitchford, N.A.; Taylor, R. Crystallographic Databases and Their Use for Studying Intramolecular Interactions. In: *Designing Bioactive Molecules*, Martin, Y.C.; Willet, P., Eds.; ACS: Washington, DC, **1998**; pp. 19-46.
- [162] Dayam, R.; Sanchez, T.; Clement, O.; Shoemaker, R.; Sei, S.; Neamat, N. β -Diketo acid pharmacophore hypothesis. I. Discovery of a novel class of HIV-1 integrase inhibitors. *J. Med. Chem.*, **2005**, *48*, 111-120.
- [163] Bacchi, A.; Carcelli, M.; Compari, C.; Piscaro, E.; Pala, N.; Rispoli, G.; Rogolino, D.; Sanchez, T.W.; Sechi, M.; Neamat, N. HIV-1 IN strand transfer chelating inhibitors: a focus on metal binding. *Mol. Pharm.*, **2011**, *8*, 507-519.
- [164] Huang, N.; Jacobson, M.P. Binding-site assessment by virtual fragment screening. *PLoS One*, **2010**, *5*, e10109.
- [165] Chen, J.Z.; Wang, J.; Xie, X.Q. GPCR structure-based virtual screening approach for CB2 antagonist search. *J. Chem. Inf. Model.*, **2007**, *47*, 1626-1637.
- [166] Lachance, H.; Wetzel, S.; Waldmann, H. Role of natural products in drug discovery. In: *Lead Generation Approaches in Drug Discovery*; Rankovic, Z.; Morphy, R., Eds.; John Wiley & Sons: New York, **2010**; pp. 187-229.
- [167] Saklani, A.; Kuty, S.K. Plant-derived compounds in clinical trials. *Drug Discov. Today*, **2008**, *13*, 161-171.
- [168] Ojima, I. Modern natural products chemistry and drug discovery. *J. Med. Chem.*, **2008**, *51*, 2587-2588.
- [169] Lee, M.L.; Schneider, G. Scaffold architecture and pharmacophoric properties of natural products and trade drugs: application in the design of natural product-based combinatorial libraries. *J. Comb. Chem.*, **2001**, *3*, 284-289.
- [170] Fcher, M.; Schmidt, J.M. Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, 218-227.
- [171] Trigg, D.J. The chemist as astronaut: searching for biologically useful space in the chemical universe. *Biochem. Pharmacol.*, **2009**, *78*, 217-223.
- [172] Fritl, P.; Schuffenhauer, A. Cheminformatics analysis of natural products: lessons from nature inspiring the design of new drugs. *Prog. Drug Res.*, **2008**, *66*, 217-219-235.
- [173] Grabowski, K.; Baringhaus, K.H.; Schneider, G. Scaffold diversity of natural products: inspiration for combinatorial library design. *Nat. Prod. Rep.*, **2008**, *25*, 892-904.
- [174] Reymond, J.L.; van Deursen, R.; Hlum, L.C.; Ruddigkeit, L. Chemical space as a source for new drugs. *Mechchemcomm*, **2010**, *1*, 30-38.
- [175] Frawley, W.J.; Platetsky-Shapiro, G.; Matheus, C.J. Knowledge discovery in databases: an overview. *AI Magazine*, **1992**, *13*, 57-70.
- [176] Fayyad, U.; Platetsky-Shapiro, G.; Smyth, P. From data mining to knowledge discovery in databases. *AI Magazine*, **1996**, *17*, 37-54.
- [177] Bak, A.; Magdziarz, T.; Kurczyk, A.; Polanski, J. Mapping fragmental drug-likeness in the MolSiftBioDot environment: intramolecular hydrogen bonding motifs in β -ketonols. *Comb. Chem. High Throughput Screen.*, **2011**, *14*, 560-569.
- [178] Podolyan, Y.; Karypis, G. Common pharmacophore identification using frequent clique detection algorithm. *J. Chem. Inf. Model.*, **2009**, *49*, 13-21.
- [179] Schomburg, K.; Ehrlich, H.C.; Sierand, K.; Rarey, M. From structure diagrams to visual chemical patterns. *J. Chem. Inf. Model.*, **2010**, *50*, 1529-1535.
- [180] Weininger, D.; Weininger, A.; Weininger, J.L. SMILES 2: Algorithm for generation of unique SMILES notation. *J. Chem. Inf. Comput. Sci.*, **1999**, *29*, 97-101.
- [181] Satoh, H.; Koshin, H.; Funatsu, K.; Nakata, T. Novel canonical coding method for representation of three-dimensional structures. *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 622-630.
- [182] Lisurek, M.; Rupp, B.; Wichard, J.; Neuenchwander, M.; von Kries, J.P.; Frank, R.; Rademann, J.; Kuhne, R. Design of chemical libraries with potentially bioactive molecules applying a maximum common substructure concept. *Mol. Divers.*, **2010**, *14*, 401-408.
- [183] Petersen, U. Quinolone Antibiotics: The Development of Moxifloxacin. In: *Analogous-Based Drug Discovery*; Fischer, J.; Gianelli, C.R., Eds.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, **2006**; pp. 315-370.
- [184] Bertani, S.; Bourdy, G.; Landau, I.; Robinson, J.C.; Esterre, P.; Deharo, E. Evaluation of French Guiana traditional antimalarial remedies. *J. Ethnopharmacol.*, **2005**, *98*, 45-54.
- [185] Kumura, N.; Izumi, M.; Nakajima, S.; Shimizu, S.; Kim, H.S.; Waiyaya, Y.; Baba, N. Synthesis and biological activity of fatty acid derivatives of quinine. *Biosci. Biotechnol. Biochem.*, **2005**, *69*, 2250-2253.
- [186] Butler, A.; Khan, S.; Ferguson, E. A brief history of malaria chemotherapy. *J. R. Coll. Physicians Edinb.*, **2010**, *40*, 172-177.
- [187] Burgome, K.L.; Bancone, G.; Nosten, F. The reality of using primaquine. *Malar. J.*, **2010**, *9*, 376.
- [188] Biot, C. Ferroquine: a new weapon in the fight against malaria. *Curr. Med. Chem. Anti-Infect. Agents*, **2004**, *3*, 135-147.
- [189] Wolf, R.; Baroni, A.; Greco, R.; Corrado, F.; Ruocco, E.; Tufano, M.A.; Ruocco, V. Quinine sulfate and HSV replication. *Dermatol. Online J.*, **2003**, *9*, 3.
- [190] Rojas, J.J.; Ochoa, V.J.; Ocampo, S.A.; Munoz, J.F., Screening for antimicrobial activity of ten medicinal plants used in Colombian folkloric medicine: a possible alternative in the treatment of non-nosocomial infections. *BMC Complement. Altern. Med.*, **2006**, *6*, 2.
- [191] Mussol, R.; Magdziarz, T.; Kurczyk, A. Quinolone scaffold as a privileged substructure in antimicrobial drugs. In: *Science against microbial pathogens communicating current research and technological advances*; Mender-Vilas, A., Ed.; Formates Research Center, **2011**; pp. 72-83.
- [192] Reichel, G. Muscle cramps – differential diagnosis and therapy. *Med. Monatsschr. Pharm.*, **2009**, *32*, 80-86.
- [193] Limburg, P.J.; Katz, H.; Grant, C.S.; Service, F.J. Quinine-induced hypoglycemia. *Ann. Intern. Med.*, **1993**, *119*, 218-219.
- [194] Wu, L.; Guo, D.; Li, H.; Hackett, J.; Yan, G.X.; Jiao, Z.; Antzelevitch, C.; Shryock, J.C.; Belardinelli, L. Role of late sodium current in modulating the proarrhythmic and antiarrhythmic effects of quinidine. *Heart Rhythm*, **2008**, *5*, 1726-1734.
- [195] Kaufman, T.S.; Ruveda, E.A. Die Jagd auf Chinin: Etappenfolge und Gesamtsiege. *Angew. Chem.*, **2005**, *117*, 876-907.
- [196] Hand, C.C.; Meshnick, S.R. Is chloroquine making a comeback? *J. Infect. Dis.*, **2011**, *203*, 11-12.
- [197] Escobedo, A.A.; Almirall, P.; Alfonso, M.; Salazar, Y.; Avila, I.; Cimerman, S.; Nunez, F.A.; Dawkins, I.V. Hospitalization of Cuban children for giardiasis: a retrospective study in a pediatric hospital in Havana. *Ann. Trop. Med. Parasitol.*, **2011**, *105*, 47-56.
- [198] Wise, R. Norfloxacin – a review of pharmacology and tissue penetration. *J. Antimicrob. Chemother.*, **1984**, *13* Suppl. B, 59-64.
- [199] Emmerson, A.M.; Jones, A.M. The quinolones: decades of development and use. *J. Antimicrob. Chemother.*, **2003**, *51* Suppl. 1, 13-20.
- [200] Przybyla, B.; Georgii, A.; Bergner, T.; Ring, J. Demonstration of quinolone phototoxicity in vitro. *Dermatologica*, **1990**, *181*, 98-103.
- [201] Blondeau, J.M. Expanded activity and utility of the new fluoroquinolones: a review. *Clin. Ther.*, **1999**, *21*, 3-40.
- [202] Bolon, M.K. The newer fluoroquinolones. *Infect. Dis. Clin. North Am.*, **2009**, *23*, 1027-1051.
- [203] Martinez, M.; McDermott, P.; Walker, R. Pharmacology of the fluoroquinolones: a perspective for the use in domestic animals. *Vet. J.*, **2006**, *172*, 10-28.
- [204] Owens, R.C.; Ambrose, P.G. Antimicrobial safety: focus on fluoroquinolones. *Clin. Infect. Dis.*, **2005**, *41* Suppl. 2, S144-S157.
- [205] Bradley, J.S.; Jackson, M.A. The use of systemic and topical fluoroquinolones. *Pediatrics*, **2011**, *128*, e1034-1045.

- [206] O'Brien, R.J. Development of fluoroquinolones as first-line drugs for tuberculosis – at long last! *Am J Respir Crit Care Med*, **2003**, *168*, 1266-1268.
- [207] Tabarrini, O.; Massan, S.; Cecchetti, V. 6-desethylfluoroquinolones as HIV-1 Tat-mediated transcription inhibitors. *Future Med Chem*, **2010**, *2*, 1161-1180.
- [208] Mugnaini, C.; Pasquini, S.; Corelli, F. The 4-quinolone-3-carboxylic acid motif as a multivalent scaffold in medicinal chemistry. *Curr Med Chem*, **2009**, *16*, 1746-1767.
- [209] Hoh, B.L.; Ogilvy, C.S. Endovascular treatment of cerebral vasospasm: transluminal balloon angioplasty, intra-arterial papaverine, and intra-arterial nicardipine. *Neurosurg Clin. N. Am.*, **2005**, *16*, 501-516.
- [210] Liu, J.K.; Coudwell, W.T. Intra-arterial papaverine infusions for the treatment of cerebral vasospasm induced by aneurysmal subarachnoid hemorrhage. *Neurocrit Care*, **2005**, *2*, 124-132.
- [211] Mudea, S.A.; Yang, B.P.; Bendok, B.R.; Miller, J.W.; Bajaj, H.H. Endovascular treatment strategies for cerebral vasospasm. *Neurosurg Focus*, **2006**, *21*, F13.
- [212] Brismar, J.L.; Eskridge, J.M.; Newell, D.W. Neurointerventional treatment of vasospasm. *Neural Res*, **2006**, *28*, 769-776.
- [213] Dinsmore, W.W. Available and future treatments for erectile dysfunction. *Clin Cornerstone*, **2005**, *7*, 37-45.
- [214] Hatzimouratidis, K.; Hatzichristou, D.G. A comparative review of the options for treatment of erectile dysfunction: which treatment for which patient? *Drugs*, **2005**, *65*, 1621-1650.
- [215] Kieback, A.G.; Felix, S.B.; Rellmann, T. Quinapril: a review of its pharmacokinetics, pharmacodynamics, toxicological data and clinical application. *Expert Opin. Drug Metab. Toxicol.*, **2009**, *5*, 1337-1347.
- [216] Costanzo, R.; Piccirillo, M.C.; Sandomenico, C.; Carilli, G.; Montanino, A.; Daniele, G.; Giordano, P.; Bryce, J.; Feo, G.D.; Di Mauro, M.; Rocco, G.; Normanno, N.; Perrone, F.; Morabito, A. Gefitinib in non small cell lung cancer. *J. Biomed. Biotechnol.*, **2011**, *2011*, 815269.
- [217] Scherfner, M.; Gion, P.D.; Doroshenko, O.; Wolf, J.; Fuhr, U. Clinical pharmacokinetics of tyrosine kinase inhibitors: focus on 4-anilinoquinazolinones. *Clin Pharmacokinet*, **2011**, *50*, 371-403.
- [218] Shao, Y.Y.; Lin, C.C.; Yang, C.H. Gefitinib or erlotinib in the treatment of advanced non-small cell lung cancer. *Discov. Med.*, **2010**, *9*, 538-545.
- [219] Nishimura, N.; Siegmund, A.; Liu, L.; Yang, K.; Bryan, M.C.; Andrews, K.L.; Ho, Y.; Booker, S.K.; Caenepeel, S.; Freeman, D.; Liao, H.; McCarter, J.; Mullady, E.L.; San Miguel, T.; Subramanian, R.; Tamayo, N.; Wang, L.; Whittington, D.A.; Zalameda, L.; Zhang, N.; Hughes, P.E.; Norman, M.H. Phosphoinositide 3-kinase (PI3K)/mammalian target of rapamycin (mTOR) dual inhibitors: discovery and structure-activity relationships of a series of quinoline and quinazoline derivatives. *J. Med. Chem.*, **2011**, *54*, 4735-4751.
- [220] Socha, A.M.; Laplante, K.L.; Russell, J.D.; Rowley, D.C. Structure-activity studies of echinomycin antibiotics against drug-resistant and biofilm-forming *Staphylococcus aureus* and *Enterococcus faecalis*. *Bioorg. Med. Chem.*, **2009**, *19*, 1504-1507.
- [221] Sandoe, J.A.T.; Wyszynski, J.; West, A.P.; Heritage, J.; Wilcox, M.H. Measurement of ampicillin, vancomycin, linezolid and gentamicin activity against enterococcal biofilms. *J. Antimicrob. Chemother.*, **2006**, *57*, 767-770.
- [222] Kim, J.B.; Lee, G.S.; Kim, Y.B.; Kim, S.K.; Kim, Y.H.; In vitro antibacterial activity of echinomycin and a novel analogue, YK2000, against vancomycin-resistant enterococci. *Int. J. Antimicrob. Agents*, **2004**, *24*, 613-615.
- [223] Weissenbacher, E.R.; Dondera, G.; Unzeitig, V.; Martinez de Tejada, B.; Gerber, S.; Halaska, M.; Spack, J. A Comparison of Dequalinum Chloride Vaginal Tablets (Fluomizimik) and Clindamycin Vaginal Cream in the Treatment of Bacterial Vaginosis: A Single-Blind, Randomized Clinical Trial of Efficacy and Safety. *Gynecol. Obstet. Invest.*, **2012**, Epub ahead of print.
- [224] Frey, I.M. B. Antimicrobial topical agents used in the vagina. *Curr. Probl. Dermatol.*, **2011**, *40*, 36-47.
- [225] Manita, D.; Toba, Y.; Takakusagi, Y.; Matsumoto, Y.; Kuwayanagi, T.; Takakusagi, K.; Tsukuda, S.; Takada, K.; Kanai, Y.; Kamakura, S.; Sakaguchi, K.; Sugawara, F. Camptothecin (CPT) directly binds to human heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) and inhibits the hnRNP A1 topoisomerase I interaction. *Bioorg. Med. Chem.*, **2011**, *19*, 7690-7697.
- [226] Pommer, Y. Camptothecins and topoisomerase I: a foot in the door. Targeting the genome beyond topoisomerase I with camptothecins and novel anticancer drugs: importance of DNA replication, repair and cell cycle checkpoints. *Curr. Med. Chem. Anticancer Agents*, **2004**, *4*, 429-434.
- [227] Pommer, Y. Topoisomerase I inhibitors: camptothecins and beyond. *Nat. Rev. Cancer*, **2006**, *6*, 789-802.
- [228] Young, C.; Schlup, T.; Hwang, J.; Eliasof, S. CRLX101 (formerly IT-101): a novel nanopharmaceutical of camptothecin in clinical development. *Curr. Biotech. Compd.*, **2011**, *7*, 8-14.
- [229] Soukasene, S.; Toft, D.J.; Moyer, T.J.; Lu, H.; Lee, H.K.; Standley, S.M.; Cryns, V.L.; Stupp, S.I. Antitumor activity of peptide amphiphilic nanolipid-encapsulated camptothecin. *ACS Nano*, **2011**, *5*, 9113-9121.
- [230] Gaur, S.; Chen, L.; Yen, T.; Wang, Y.; Zhou, B.; Davis, M.; Yen, Y. Preclinical study of the cyclodextrin-polymer conjugate of camptothecin CRLX101 for the treatment of gastric cancer. *Nanomedicine*, **2011**, Epub ahead of print.
- [231] Chang, J.; Khin, K.T.; Davis, M.E. Antitumor activity of beta-cyclodextrin polymer-camptothecin conjugates. *Mol. Pharm.*, **2004**, *1*, 183-193.
- [232] Scott, L.C.; Yao, J.C.; Benzon, A.B.; Thomas, A.L.; Falk, S.; Mena, R.R.; Picus, J.; Wright, J.; Mulcahy, M.F.; Ajani, J.A.; Evans, T.R.J. A phase II study of pegylated-camptothecin (pegamotecan) in the treatment of locally advanced and metastatic gastric and gastro-oesophageal junction adenocarcinoma. *Cancer Chemother. Pharmacol.*, **2009**, *63*, 363-370.
- [233] Ullakan, H.; Swaan, P.W. Camptothecins: a review of their chemotherapeutic potential. *Drugs*, **2002**, *62*, 2039-2057.
- [234] Zunino, F.; Dallavalle, S.; Laccabue, D.; Beretta, G.; Merlini, L.; Pratesi, G. Current status and perspectives in the development of camptothecins. *Curr. Pharm. Des.*, **2002**, *8*, 2505-2520.
- [235] Czejka, M.; Kiss, A.; Kocsner, C.; Terkula, R.; Eitlinger, D.; Schueller, J. Metabolic activation of irinotecan during intra-arterial chemotherapy of metastatic colorectal cancer. *Anticancer Res.*, **2011**, *31*, 3573-3578.
- [236] Warner, J.F.; Pheng, L.; Trinh, M.M.; Burris, H.A.; Jones, S.; Anderson, K.; Warner, S.; Porubek, D. Pharmacokinetics of SN3310, an injectable emulsion that incorporates a novel derivative of SN-38 in patients with advanced solid tumors. *J. Pharm. Sci.*, **2011**, Epub ahead of print.
- [237] Sapra, P.; Kraft, P.; Mehlig, M.; Malaby, J.; Kraft, P.; Greenberger, L.M.; Horak, I.D. Marked therapeutic efficacy of a novel polyethylene glycol-SN38 conjugate, EZN-2208, in xenograft models of B-cell non-Hodgkin's lymphoma. *Haematologica*, **2009**, *94*, 1456-1459.
- [238] Zhao, H.; Rubio, B.; Sapra, P.; Wu, D.; Reddy, P.; Sai, P.; Martinez, A.; Gao, Y.; Lozano, Y.; Longley, C.; Greenberger, L.M.; Horak, I.D. Novel prodrugs of SN38 using multiarm poly(ethylene glycol) linkers. *Bioconjug. Chem.*, **2008**, *19*, 849-859.
- [239] Sapra, P.; Zhao, H.; Mehlig, M.; Malaby, J.; Kraft, P.; Longley, C.; Greenberger, L.M.; Horak, I.D. Novel delivery of SN38 markedly inhibits tumor growth in xenografts, including a camptothecin-11-refractory model. *Clin. Cancer Res.*, **2008**, *14*, 1888-1896.
- [240] Gale, S.; Crosswell, G. 28th Annual JPMorgan Healthcare Conference – Exelixis and Nektar Therapeutics. *J. Drugs*, **2010**, *13*, 139-141.
- [241] Zhang, J.A.; Xuan, T.; Pernar, M.; Ma, L.; Ugwu, S.; Ali, S.; Ahmad, I. Development and characterization of a novel liposome-based formulation of SN-38. *Int. J. Pharm.*, **2004**, *270*, 93-107.
- [242] Agarwal, N.; Sonpavde, G.; Sternberg, C.N. Novel molecular targets for the therapy of castration-resistant prostate cancer. *Eur. Urol.*, **2011**, Epub ahead of print.
- [243] Olsson, A.; Bjork, A.; Vallon-Christersson, J.; Isaacs, J.T.; Leanderson, T. Tasquinimod (ABR-215050), a quinoline-3-carboxamide anti-angiogenic agent, modulates the expression of thrombospondin-1 in human prostate tumors. *Mol. Cancer*, **2010**, *9*, 107.
- [244] Bratt, O.; Hagman, M.; Ahlgren, G.; Nordle, O.; Bjork, A.; Danberg, J.E. Open-label, clinical phase I studies of tasquinimod in patients with castration-resistant prostate cancer. *Br. J. Cancer*, **2009**, *101*, 1233-1240.
- [245] Cybulski, P.; Thakur, S.D.; Foster, B.C.; Scott, I.M.; Ledue, R.I.; Arnason, J.T.; Dillon, J.A.R. Extracts of Canadian first nations medicinal plants, used as natural products, inhibit neisseria gonorrhoeae isolates with different antibiotic resistance profiles. *Sex. Transm. Dis.*, **2011**, *38*, 667-671.
- [246] Yu, H.H.; Kim, K.J.; Cha, J.D.; Kim, I.K.; Lee, Y.E.; Choi, N.Y.; You, Y.O. Antimicrobial activity of berberine alone and in combination with ampicillin or oxacillin against methicillin-resistant *Staphylococcus aureus*. *J. Med. Food*, **2005**, *8*, 454-461.
- [247] Piyanch, R.; Sukhtankar, M.; Wande, G.; Back, S.J. Berberine, a natural isoquinoline alkaloid, induces NAG-1 and ATF3 expression in human colorectal cancer cells. *Cancer Lett.*, **2007**, *258*, 230-240.
- [248] Letavova, S.; Janova, S.; Muckova, M.; Theisova, M. Antiproliferative activity of berberine in vitro and in vivo. *Biomed. Pap. Med. Fac. Univ. Palacky Olomouc Czech Repub.*, **2005**, *149*, 461-463.
- [249] Izzo, R.; de Simone, G.; Giudice, R.; Chinali, M.; Trimarco, V.; De Luca, N.; Trimarco, B. Effects of nutraceuticals on prevalence of metabolic syndrome and on calculated Framingham Risk Score in individuals with dyslipidemia. *J. Hypertens.*, **2010**, *28*, 1482-1487.
- [250] Vuddanda, P.R.; Chakraborty, S.; Singh, S. Berberine: a potential phytochemical with multispectrum therapeutic activities. *Expert Opin. Investig. Drugs*, **2010**, *19*, 1297-1307.
- [251] Cok, A.; Plaisier, C.; Saïte, M.J.; Orum, D.S.; Cheng, J.; Louters, L.L. Berberine acutely activates the glucose transport activity of GLUT1. *Biochimie*, **2011**, *93*, 1187-1192.
- [252] Chen, C.; Zhang, Y.; Huang, C. Berberine inhibits PTP1B activity and mimics insulin action. *Biochem. Biophys. Res. Commun.*, **2010**, *397*, 543-547.
- [253] Han, J.; Lin, H.; Huang, W. Modulating gut microbiota as an anti-diabetic mechanism of berberine. *Med. Sci. Monit.*, **2011**, *17*, RA164-167.
- [254] Kulkarni, S.K.; Dhir, A. Berberine: a plant alkaloid with therapeutic potential for central nervous system disorders. *Phytother. Res.*, **2010**, *24*, 317-324.
- [255] Habtemariam, S. The therapeutic potential of Berberis darwinii stem-bark: quantification of berberine and in vitro evidence for Alzheimer's disease therapy. *Nat. Prod. Commun.*, **2011**, *6*, 1089-1090.
- [256] Wu, Y.; Li, J.Q.; Kim, Y.J.; Wu, L.; Wang, Q.; Hao, Y. In vivo and in vitro antiviral effects of berberine on influenza virus. *Chin. J. Integr. Med.*, **2011**, *17*, 444-452.
- [257] Li, Q.; Chen, Y.; Sun, L.; Fu, G.; Guo, L. Vasodilatation produced by Fasudil Mesylate in vivo and in vitro. *Vascul. Pharmacol.*, **2011**, *55*, 121-126.

- [258] Soga, J.; Noma, K.; Hata, T.; Hidaka, T.; Fujii, Y.; Idei, N.; Fujimura, N.; Mikami, S.; Marubashi, T.; Kohara, Y.; Chayama, K.; Kato, H.; Liao, J.K.; Higashi, Y. Rho-associated kinase activity, endothelial function, and cardiovascular risk factors. *Arterioscler. Thromb. Vasc. Biol.*, **2011**, *31*, 2353-2359.
- [259] Fukumoto, Y.; Shimokawa, H. Recent progress in the management of pulmonary hypertension. *Circ. J.*, **2011**, *75*, 1801-1810.
- [260] Satoh, K.; Fukumoto, Y.; Shimokawa, H. Rho-kinase: important new therapeutic target in cardiovascular diseases. *Am. J. Physiol. Heart Circ. Physiol.*, **2011**, *301*, H287-296.
- [261] O'Connor, C.M. Raynaud's phenomenon. *J. Vasc. Nurs.*, **2001**, *19*, 87-92.
- [262] Garcia-Carrasco, M.; Jimenez-Hernandez, M.; Escarceja, R. O.; Mendoza-Pinto, C.; Pardo-Santos, R.; Levy, R.; Maldonado, C.G.; Chavez, G.P.; Cervera, R. Treatment of Raynaud's phenomenon. *Autoimmun. Rev.*, **2008**, *8*, 62-68.
- [263] Yada, T.; Shimokawa, H.; Kajiya, F. Cardioprotective effect of hydroxyfasudil as a specific Rho-kinase inhibitor on ischemia-reperfusion injury in canine coronary microvessels *in vivo*. *Clin. Hemorheol. Microcirc.*, **2006**, *34*, 177-183.
- [264] Huentelman, M.J.; Stephan, D.A.; Talboom, J.; Corneveaux, J.J.; Reiman, D.M.; Gerber, J.D.; Barnes, C.A.; Alexander, G.E.; Reiman, E.M.; Bimonte-Nelson, H.A. Peripheral delivery of a ROCK inhibitor improves learning and working memory. *Behav. Neurosci.*, **2009**, *123*, 218-223.
- [265] Yamashita, K.; Todo, S. Sotrastaurin, a new selective protein kinase C inhibitor, on the way. *Transplantation*, **2011**, *93*, 146-147.
- [266] Evmon, J.P.; Wagner, J.; Zenke, G.; Brinkmann, V.; Wagner, K.; Kovarik, J.; Welzenbach, K.A.; Weitz-Schmidt, G.; Guntermann, C.; Towbin, H.; Cottens, S.; Kaminski, S.; Letschka, T.; Lutz-Nicoladoni, C.; Gruber, T.; Hermann-Kleiter, N.; Thümler, N.; Baier, G. The potent protein kinase C-selective inhibitor AEB071 (sotrastaurin) represents a new class of immunosuppressive agents affecting early T-cell activation. *J. Pharmacol. Exp. Ther.*, **2009**, *330*, 792-801.
- [267] Friman, S.; Arns, W.; Nashed, B.; Vincenti, F.; Banas, B.; Budde, K.; Cibrik, D.; Chan, L.; Klempnauer, J.; Mulgaonkar, S.; Nicholson, M.; Wahlberg, J.; Wissing, K.M.; Abrams, K.; Witte, S.; Woodle, E.S. Sotrastaurin, a novel small molecule inhibiting protein-kinase C: randomized phase II study in renal transplant recipients. *Am. J. Transplant.*, **2011**, *11*, 1444-1455.
- [268] Kovarik, J.M.; Neuhaus, P.; Cillo, U.; Weber, M.; Ströhl, S.; Gatlif, E.; Meiser, K.; Slade, A. Sotrastaurin single-dose pharmacokinetics in de novo liver transplant recipients. *Transpl. Int.*, **2011**, *24*, 276-283.
- [269] von Hahn, T.; Schulze, A.; Chicano Wust, I.; Heidrich, B.; Becker, T.; Steinmann, E.; Helfritz, F.A.; Rohrmann, K.; Urban, S.; Manns, M. P.; Pietschmann, T.; Ciesek, S. The novel immunosuppressive protein kinase C inhibitor sotrastaurin has no pro-viral effects on the replication cycle of hepatitis B or C virus. *PLoS One*, **2011**, *6*, e24142.
- [270] Brave, S.R.; Ratzliff, K.; Wilson, Z.; James, N.H.; Ashton, S.; Wainwright, A.; Kendrew, J.; Dudley, P.; Broadbent, N.; Sproat, G.; Taylor, S.; Barnes, C.; Silva, J.C.; Farnsworth, C.L.; Hennequin, L.; Ogilvie, D.J.; Jurgensmeyer, J.M.; Shibuya, M.; Wedge, S.R.; Barry, S.T. Assessing the activity of cediranib, a VEGFR-2/3 tyrosine kinase inhibitor, against VEGFR-1 and members of the structurally related PDGFR family. *Mol. Cancer Ther.*, **2011**, *10*, R61-R73.
- [271] Mitchell, C.L.; O'Connor, J.P.B.; Roberts, C.; Watson, Y.; Jackson, A.; Cheung, S.; Evans, J.; Spicer, J.; Harris, A.; Kelly, C.; Rudman, S.; Middleton, M.; Fielding, A.; Tessier, J.; Young, H.; Parker, G.J.M.; Jayson, G.C. A two-part phase II study of cediranib in patients with advanced solid tumours: the effect of food on single-dose pharmacokinetics and an evaluation of safety, efficacy and imaging pharmacodynamics. *Cancer Chemother. Pharmacol.*, **2011**, *68*, 631-641.
- [272] Shanafelt, T.; Zent, C.; Byrd, J.; Erlichman, C.; Laplant, B.; Ghosh, A.; Call, T.; Villalona-Calero, M.; Jelinek, D.; Bowen, D.; Laumann, K.; Wu, W.; Hanson, C.; Kay, N. Phase II trials of single-agent anti-VEGF therapy for patients with chronic lymphocytic leukemia. *Leuk. Lymphoma*, **2010**, *51*, 2222-2229.
- [273] Lorusso, P.; Shields, A.F.; Gadgeel, S.; Vaishampayan, U.; Guthrie, T.; Puchalski, T.; Xu, J.; Liu, Q. Cediranib in combination with various anticancer regimens: results of a phase I multi-cohort study. *Invest. New Drugs*, **2011**, *29*, 1395-1405.
- [274] Ramalingam, S.S.; Belani, C.P.; Mack, P.C.; Vokes, E.E.; Longmate, J.; Govindan, R.; Koczywas, M.; Ivy, S.P.; Gundara, D.R. Phase II study of Cediranib (AZD 2171), an inhibitor of the vascular endothelial growth factor receptor, for second-line therapy of small cell lung cancer (National Cancer Institute #7097). *J. Thorac. Oncol.*, **2010**, *5*, 1279-1284.
- [275] Najj, A.J.; Jung, Y.S.; Won, J.J.; Conley-Lacomb, M.K.; Saliganan, A.; Kim, C.J.; Heath, E.; Cher, M.L.; Bonfil, R.D.; Kim, H.R.C. Cediranib inhibits both the intrasosseous growth of PDGF D-Positive prostate cancer cells and the associated bone reaction. *Prostate*, **2011**, Epub ahead of print.
- [276] di Tomaso, E.; Stauder, M.; Kamoun, W.S.; Duda, D.G.; Auluck, P.K.; Fazlollahi, L.; Androneni, O.C.; Frosch, M.P.; Wen, P.Y.; Plotkin, S.R.; Hedley-Whyte, E.T.; Sorensen, A.G.; Hatchelor, T.T.; Jain, R.K. Glioblastoma recurrence after cediranib therapy in patients: lack of "rebound" revascularization as mode of escape. *Cancer Res.*, **2011**, *71*, 19-28.
- [277] Dean, E.; Middleton, M.R.; Pwint, T.; Swaisland, H.; Carmichael, J.; Goodege-Kunwar, P.; Ranson, M. Phase I study to assess the safety and tolerability of olaparib in combination with bevacizumab in patients with advanced solid tumours. *Br. J. Cancer*, **2012**, *106*, 468-474.
- [278] Glendonning, J.; Tutt, A. PARP inhibition—current status and the walk towards early breast cancer. *Breast*, **2011**, *20* Suppl 1, S12-19.
- [279] D'Amours, D.; Demoyers, S.; D'Silva, I.; Poirier, G.G. Poly(ADP-ribose)ylation reactions in the regulation of nuclear functions. *Biochem. J.*, **1998**, *342*, 249-268.
- [280] d'Adda di Fagagna, F.; Hande, M.P.; Tong, W.M.; Lansdorp, P.M.; Wang, Z.Q.; Jackson, S.P. Functions of poly(ADP-ribose) polymerase in controlling telomere length and chromosomal stability. *Nat. Genet.*, **1999**, *23*, 76-80.
- [281] Narod, S.A.; Foulkes, W.D. BRCA1 and BRCA2: 1994 and beyond. *Nat. Rev. Cancer*, **2004**, *4*, 665-676.
- [282] Gartner, E.M.; Burger, A.M.; Lorusso, P.M. Poly(ADP-ribose) polymerase inhibitors: a novel drug class with a promising future. *Cancer J.*, **2010**, *16*, 83-90.
- [283] Khan, O.A.; Gore, M.; Longan, P.; Stone, J.; Greytstock, A.; Burke, W.; Carmichael, J.; Watson, A.J.; McGown, G.; Thorncroft, M.; Margison, G.P.; Califano, R.; Larkin, J.; Wellman, S.; Middleton, M.R. A phase I study of the safety and tolerability of olaparib (AZD2281, KU0059436) and dacarbazine in patients with advanced solid tumours. *Br. J. Cancer*, **2011**, *104*, 750-755.
- [284] Fong, P.C.; Boss, D.S.; Yap, T.A.; Tutt, A.; Wu, P.; Mergui-Roelvink, M.; Mortimer, P.; Swaisland, H.; Lau, A.; O'Connor, M.J.; Ashworth, A.; Carmichael, J.; Kaye, S.B.; Schellens, J.H.M.; de Bono, J.S. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N. Engl. J. Med.*, **2009**, *361*, 123-134.
- [285] Gelmon, K.A.; Tsiacklowitz, M.; Mackay, H.; Swenerton, K.; Robidoux, A.; Tonkin, K.; Hirte, H.; Huntsman, D.; Clemons, M.; Gilks, B.; Yerushalmi, R.; Macpherson, E.; Carmichael, J.; Oza, A. Olaparib in patients with recurrent high-grade serous or poorly differentiated ovarian carcinoma or triple-negative breast cancer: a phase 2, multicentre, open-label, non-randomised study. *Lancet Oncol.*, **2011**, *12*, 852-861.
- [286] Audeh, M.W.; Carmichael, J.; Penson, R.T.; Friedlander, M.; Powell, B.; Bell-McGuinn, K.M.; Scott, C.; Weitzel, J.N.; Oaknin, A.; Loman, N.; Lu, K.; Schmutzler, R.K.; Matulonis, U.; Wickens, M.; Tutt, A. Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: a proof-of-concept trial. *Lancet*, **2010**, *376*, 245-251.
- [287] Kruse, V.; Rottey, S.; De Backer, O.; Van Belle, S.; Coquyt, V.; Denys, H. PARP inhibitors in oncology: a new synthetic lethal approach to cancer therapy. *Acta Clin. Belg.*, **2011**, *56*, 2-9.

Załącznik 2b. Kserokopia publikacji naukowej:

Bak A., Magdziarz T., Kurczyk A., Serafin K., Polanski J.; Probing a Chemical Space for Fragmental Topology-Activity Landscapes (FRAGTAL): Application for Diketo Acid and Catechol HIV Integrase Inhibitor Offspring Fragments., *Comb. Chem. High Throughput Screen.*, 2013, 16(4), 274-287. (IF₂₀₁₁ = 1,785)

Probing a Chemical Space for Fragmental Topology-Activity Landscapes (FRAGTAL): Application for Diketo Acid and Catechol HIV Integrase Inhibitor Offspring Fragments

Andrzej Bak*, Tomasz Magdziarz, Agata Kurczyk, Katarzyna Serafin and Jaroslaw Polanski*

Department of Organic Chemistry, Institute of Chemistry, University of Silesia, PL-40-006 Katowice, Poland

Abstract: Fragmental topology-activity landscapes (FRAGTAL), a new concept for encoding molecular descriptors for fragonomics into the framework of the molecular database records is presented in this paper. Thus, a structural repository containing biological activity data was searched in a substructure mode by a series of molecular fragments constructed in an incremental or decremental manner. The resulted series of database hits annotated with their activities construct FRAGTAL descriptors encoding a frequency of the certain fragments among active compounds and/or their activities. Actually, this method might be interpreted as a simplified adaptation of the frequent subgraph mining (FSM) method. The FRAGTAL method reconstructs the way in which medicinal chemists are used to designing a prospective drug structure intuitively. A representative example of the practical application of FRAGTAL within the ChemDB Anti-HIV/OITB database for disclosing new fragments for HIV-1 integrase inhibition is discussed. In particular, FRAGTAL method identifies ethyl malonate amide (EMA) as the diketo acid (DKA) related arrangement. Since new molecular constructs based on the EMA fragment are still a matter of future investigations we referred to this as the DKA offspring.

Keywords: Virtual ligand screening, druglikeness, fragment-based design, molecular scaffolds, HIV-1 integrase inhibitors.

1. INTRODUCTION

The nature of the drug-target interactions ruled by inter- and/or intra-molecular forces is a tremendously complex and divergent process that still lacks a universal approach. Hence, several approaches to drug discovery have been developed in the quest to identify the differences between drug- and non-drug molecules (chemicals). Most often, these methods are related to the drug-likeness concept originating from the idea that drug compounds differ from other molecules in their physicochemical properties [1]. The *a priori* calculation of molecular descriptors that is mainly derived from the molecular structure is critical for the compound's bioavailability and its ADMET properties [2]. Drug-likeness can be seen as the absolute prerequisite for the compound to survive through the expected clinical trials and should be estimated before the synthesis of the compounds is rationalised.

In reality, molecular descriptors quantifying drug-like properties (DP) are easily available from molecular formulae, e.g. the number of hydrogen bond donors (HBD) or acceptors (HBA). In practice, even those that refer to molecular properties can nowadays be estimated for virtual molecules by performing a calculation, e.g. logP [3]; however, the reliability of the resulting DP values is still questionable. Moreover, the DPs are generally extensive types of parameters describing a whole molecule and therefore they do not permit discriminating between diversified molecular structures easily and efficiently. The

empirical analysis of the marketed drug-size library provided the heuristic guide that relates the calculable physicochemical properties with the complex *in vivo* pharmacokinetic parameter that comprises aqueous solubility and oral bioavailability [4]. The emergence of the quantitative property-based Lipinski Ro5 'sieve' is focused on the importance of restricting the chosen *in silico* molecular parameters of preclinical drug candidates in the pharmacokinetic-friendly property space [5]. The discrimination Ro5 model restrains the 'drug-like property space' through a set of threshold values which are necessary but are not sufficient boundary guidelines of the preferred 'drug architecture' that is especially useful in the computer-driven process of filtering chemical libraries; however, this concept is still elusive and ambiguous. It is clear that *a good drug-like score does not make a drug* since 80% of the chemicals recorded in the ACD database are Ro5-compliant, while the molecular weight (MW) distribution of drugs vs other chemicals were found to be convergent [6]. Therefore, the drug-likeness filter (DLF) is less applicable in discriminating drugs from non-drugs but more likely to quickly eliminate those chemicals that are rich in nondrug-like fragments. Despite these uncertainties, the drug-likeness concept appears to be a great success and has contributed a great deal to our understanding of medicinal chemistry.

The similarity tenet of molecular design claims that structurally related compounds display a similar biological activity profile [7]. Despite many exceptions the similar-property principle is deeply rooted not only in the art of the drug discovery, but more basically, in chemistry where analogy is crucial, especially when we do not yet fully understand a given process or effect [8]. The high-speed examination of a large number of molecules combined with the metric quantifying the similarity with a specified pattern

*Address correspondence to this author at the Department of Organic Chemistry, Institute of Chemistry, University of Silesia, PL-40-006 Katowice, Poland; Tel: 0048 32 359 136 99; E-mails: abak@us.edu.pl, polanski@us.edu.pl

normally requires the abstract molecule representations in form of fix-sized binary fingerprint denoting the presence and/or absence of the predefined object feature, sometimes augmented with the weighting factors [9]. The set of various similarity measures as well as the pattern matching protocols have been developed for estimation of the structural relatedness within a molecular repository with preferably Hamming, Euclidean, Tanimoto or Tversky coefficient [10].

Fragment-based drug design or fragonomics, which has recently been getting more and more widespread appeared to offer an alternative way for identifying advantageous drug structures [11-13]. This methodology assumes the existence of a preferred molecular 'core', namely a framework that works as an 'anchor' for the 'ornamenting' blocks – privileged motifs [14]. The analysis of low-weighted molecular motifs offers a variety of novel and original methods in both computational and experimental investigations [15]. In theoretical studies, the search for suitable molecular fragments capable of binding to a specified target might be less computationally demanding, which allows for efficient high-throughput-screening (HTS) procedures, e.g. fragmental docking [16]. Privileged scaffolds have been used successfully as 'core' structures for synthesis and the optimal starting point for the library design of fragments which interact with a given receptor in experimental analyses. They appear to be useful for the construction of more complex structures, regardless of the relatively low affinity level of the initial motifs [17]. It is well known that the identified privileged motifs are highly dependent on the amount and diversity of the accessible structural data [18].

Moreover, recently published studies of drug-size libraries confirmed that clinical candidates with MW<400 possess a 50% higher probability of being approved as drugs [19]. In this context reasonable is question about 'how big is too big' or even more radical philosophy that 'the smaller the better'? [20]. The principal challenge of fragment-based ligand design method is addressed to issues of providing the diverse set of fragments with physicochemical properties which might be optimized to have greater potency. It is currently well-established procedure for effective lead specification with lower molecular complexity compared to drug molecules derived solely on the basis of the Ro5 criteria. Congreve et al. adjusted the Ro5 rule for fragments to so called 'rule of three' restricting the boundary limits to MW 300, logP 3, NRB 3, HBD 3, HBA 3 and PSA 60Å² [19]. The ensemble of Ro3-consistent property thresholds following the 'reduced complexity' paradigm in fragment library for lead generation have been proposed by Oprca et al. (MW 350, NRB 6, HBD 3, HBA 8, logP 2.2 and heavy atoms NHA 22) [21, 22].

The topological aspect of drug-like molecules can explain the preferential architecture of drug molecules, e.g. Murcko and Bemis analyzed the frameworks [23] and side chains [24] of some known drugs. More recently, new descriptors were developed for the analysis of the topology of fragments, e.g. ToFI [25]. A detailed discussion of the application of fragment-based methodology for finding privileged motifs can be found elsewhere [26]. Although the identification of the structural elements and functions of the drug-like molecules and/or the recognition of privileged

scaffolds is a method of potential interest for differentiation drugs vs non-drugs, topological approaches were not often used for these purposes [27]. Briefly, the idea of privileged structures was originally inspired by Evans [28] who argued that we can reveal *certain structural features which produce biological effects more often than others* [29]. A number of privileged structures can be found elsewhere, e.g. in reference [30]. Surprisingly, an experimental search for new privileged structures through the HTS screening of a large library of 10 000 compounds using NMR-based binding assays with 11 different protein targets compounds has not expanded the record of the scaffolds that are already known [31, 32]. Moreover, fragment based drug design using fluorescence correlation spectroscopy techniques has been proposed [33].

However, massive database mining for fragment (co)-occurrence might potentially delve into the architecture of extensively or rarely explored regions of the chemical space (CS). Questions about the preferable combinations of substructures comprising a molecule naturally appear. Does this only reflect a 'chemical habit' or even 'fashion' among chemists due to the feasibility of a synthetic drug or is there a 'natural force' that combines some 'building blocks' together? Although the concept of privileged structures is still controversial since it is not easy to determine conclusively if an overrepresentation of a certain structure is not a minor consequence of its abundance in chemical libraries [34, 35], a number of computational approaches have been aimed at their identification. Moreover, their experimental verification using HTS approaches has also been attempted [36].

Generally, the problem of identifying relevant molecular fragments is a complex issue that is related to the so-called frequent subgraph mining (FSM) [37]. Conceptually, the subgraph mining algorithms identify and retrieve from a repository all compounds sharing a common user-defined pattern of atoms/bonds arranged in the specified topological/geometrical manner [38]. Then each selected subgraph might be transformed into a corresponding canonically-coded linear atoms/bonds pattern well suited for the high-speed database screening. The procedure of finding a particular motif in molecule might be regarded as string regular expressions describing the search criteria [39]. In the maximum common substructure (MCS) method, the molecules are fragmented into smaller parts as long as a common fragment might be identified within the analyzed series of compounds [40]. More practical FSM approaches to drug design are still a matter of future developments that need some sophisticated algorithms for massive *in-silico* database mining and data processing [41]. The massive screening of an entire CS is computationally infeasible; however, it might be a valuable source for gaining a completely new understanding of chemistry and drug discovery. In this context, a novel molecular and structural database management system (MoStBioDat) has been developed as a free software package designed as a dual purpose storage/extraction platform providing a high degree of data reliability as well as software-based solutions for massive protocols that are capable of parallel ligand-target analysis [42]. A detailed description of the package and its practical applications for structural knowledge discovery can be found elsewhere [43-45].

In practice, the design of a candidate compound in organic synthesis is still of key importance both during searches for new hit structures and their optimization into suitable pharmaceutical leads. Thus, success depends upon the knowledge and experience of organic and medicinal chemists. In particular, intuition is of key importance here and this can be developed by training and an awareness of the mass of chemical information that is available in databases, and can be sharpened by the visualization of these data by their structure-activity or structure-property relationships. Intuition helps chemists to find suitable small fragments and to connect them and find interesting target molecules. The concept of organic chemistry architecture suggests that one way to explore the information from the molecular repositories in order to model is 'preferential' activity landscapes (ALs) [46]. The graphical representation of the easily interpretable biological response surface in a CS which integrates the molecular structure and potency relationship allows for the massive mining of the SAR information as well. Consequently, synthetic chemists have realised the importance of SAR visualisation, which might be used as *intuitive roadmaps for molecular design and optimization* [47]. Thus, the SAR landscapes or fitness landscapes approaches have gained importance recently, e.g. Schneider *et al.* presented a ligand-induced structure-activity relationship display (LiSARD) method for the neighborhood-preserving visualisation of adaptive SAR landscapes based on the compression of multidimensional compound representations using principal component analysis (PCA) or related methods [47].

In this paper we present a new concept for encoding molecular database records into molecular descriptors which can be used for the visualization of the activity landscapes for incrementally growing or decreasing fragments, namely, fragmental topology-activity landscapes (FRAGTAL). The structural ChemDB and Ligand.Info repositories containing biological activity data was investigated in a substructure mode through a series of molecular fragments that were constructed in an incremental or decremental manner. The retrieved database hits were subsequently applied to determine the frequency and to perform activity analyses of the constructed molecular fragments. In reality, this method might be interpreted as a simplified adaptation of the frequent subgraph mining (FSM) method. FRAGTAL reconstructs the way in which medicinal chemists are used to designing a prospective drug structure intuitively. A representative example of the practical application of FRAGTAL within the ChemDB Anti-HIV/OI/TB database for disclosing new fragments for HIV-1 integrase inhibition is discussed.

Freely accessible *on-line* search tools and the MoStBioDat package for the molecular database analysis were used in order to present a new concept for encoding the records of a biological activity database into a molecular descriptor for fragonomics. A database is screened in a substructure mode through a series of incrementally increasing or decreasing molecular subfragments. Database hits identify the substructures bearing these particular subfragments and analyses of these subfragments define their frequency and activity FRAGTAL descriptors. In this paper, we show that this method can provide a rationale that can be used to enhance the intuitive design of potential drugs

by discussing an example of the practical application of FRAGTAL for disclosing new fragments for HIV-1 integrase inhibition within the ChemDB Anti-HIV/OI/TB database.

2. THEORETICAL BACKGROUND

2.1. Overview of the MostBioDat Architecture

An exploration of chemical libraries provides a challenging object of interest for contemporary molecular modeling as it is mainly addressed to new software approaches. The MoStBioDat package is a consistent, dual-purpose deposit/retrieval chemical information system that provides integrated software-based solutions for massive *in silico* protocols which enables the simultaneous analysis of ligand and receptor data, respectively. The developed set of the procedures might be combined with the external pythonic packages e.g. OEChem or Pybel in a fully programmable environment that automates the challenging issue of the virtual database mining and structural knowledge screening [48, 49]. In fact, the conjunction of the expressive scripting language with the specified functionalities of the high-performance database system enables the subsequent sampling of the virtually infinite chemical space.

Conceptually, the system architecture currently comprises the following components: *Storage* and *User Layer*, respectively. The *Storage Layer* is responsible for data storing/preserving with simple methods for data querying and retrieving while the post-processing component in the *User Layer* called the *User Interface Component* provides procedures for data analysis. In general, the relational database system with *Ligand* and *Macromolecule* parts enables the storage/search and analysis of 1D (biological and physical properties), 2D (topological) and 3D (topographical) data. The *Macromolecule* organization imitates the conventional topology of macromolecular systems following the typical PDB flat-file hierarchy [50]. The *Data Management Component* forms the abstract middleware with tools for structure validation, deposition and extraction. The fully programmable text mode access combined with a set of mature Python libraries and programs is a preferable utility for the post-processing analysis, e.g. massive database mining protocols. Basically, the database system has been formatted to become searchable using the standard algorithmic techniques needed for molecular (sub)structure and similarity recognition.

2.2. Fragmental Topology-Activity Landscapes (FRAGTAL)

A knowledge-discovery-based molecular descriptor is proposed for encoding the data retrieved from a mining factual molecular database(s), in particular, a biologically annotated database. The construction of the descriptor is a multistep procedure in which a key step is the database mining.

1. A series of queries Q_i , $i = 1, \dots, n$ are defined by the construction or fragmentation, e.g. the incremental addition or subtraction of any atom (group of atoms) to/from the molecular fragment.

2. A database is queried (in a substructure mode) which provides a series of superstructures S_i that are available for each query Q_i .
3. Each set of superstructures S_i is analyzed according to its properties and the number of occurrences in the database (database frequency). Afterward, the fragmental topology (FT) descriptor encoding database frequency F_i and maximal, minimal or mean property value P_i for each fragment is calculated.

2.3. FRAGTAL for a Small Series of Carboxylic Acids Annotated with pKa Values

In order to provide a better understanding of the FRAGTAL concept, a schematic workflow in Fig. (1) and an illustrative example of the small database defined by formula 1 (Table 1) and annotated by its pKa values is given below. Our database registers a series of 29 carboxylic acids, starting from acetic to caproic acids which were supplemented by all of the possible terminally substituted *mono-* *di-* and *tri-* halogen acids. For simplicity, in these series only the analogues that were substituted with the same halogen were included. The database is presented in Table 1. The values of pKa were estimated using the ACD Labs pKa DB software [51]. Starting from the lowest analogue, a series of queries Q_1 - Q_5 was obtained by the incremental addition of a carbon atom as shown in Fig. (2). Next, the queries Q_1 - Q_5 were used for probing the database presented in Table 1 in order to analyze the frequency and the distribution of the pKa values from the hits, respectively. Thus, querying in the substructure mode provides the frequencies given by the FRAGTAL_n specified in Table 2. At the same time, the Q_1 - Q_5 hits can be described by the pKa values which means that the maximal and minimal values are given in Table 2 (rows 2-4), respectively. Additionally, the FRAGTAL descriptors from Table 2 are graphically illustrated in the histogram shown in Fig. (3), where a bar at each carbon atom illustrates the frequency (black) and mean pKa (grey) FRAGTAL descriptors for a substructure terminating at this specific atom, respectively. A detailed analysis of the frequency of a FRAGTAL_n descriptor shows that the successive addition of carbon atoms from down to up results in a decrease in the number of hits. The lowest substructure S_1 includes all of the 29 analogues used for the construction of the database, while the highest S_5 one finds only 4 caproic acid derivatives in the database. The increase of the FRAGTAL_{pKa} values from the S_1 to S_5 nicely illustrates the known electronic effect of the acid substitution with halogen. As this effect rapidly decreases with the increasing distance of the substituting halogen from the carboxylic function, the lowest value of FRAGTAL_{pKa} was obtained for the Q_1 query which included all of the halogenated acetic acids (C^1). Conversely, the Q_5 hits which included only (halogenated) caproic acids (C^6) had the highest pKa values. Moreover, while the mean, maximal and minimal FRAGTAL_{pKa} values were very different for the Q_1 hits, they differed only slightly for the Q_5 hits, which illustrates that the substitution did not influence the pKa values in the latter case. A further analysis of the FRAGTAL_{pKa} (mean, max and min values) revealed the subtle structure-activity rules of the acid substitution with halogens.

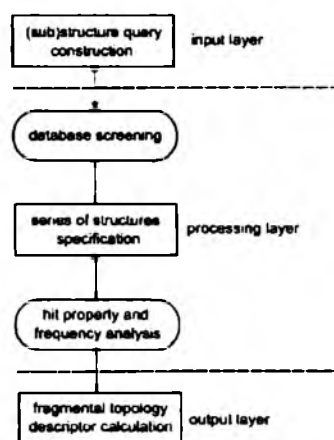
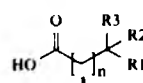


Fig. (1). The workflow of the FRAGTAL descriptor calculation.

Table 1. The Illustrative Database of Carboxylic Acids Defined by Formula 1 Annotated by their pKa Values



Nr	n	R1	R2	R3	pKa Value
1	0	H	H	H	4.792
2	1	H	H	H	4.791
3	2	H	H	H	4.758
4	3	H	H	H	4.778
5	4	H	H	H	4.781
6+17	0	X	H	H	0.046-3.175
		X	X	H	
		X	X	X	
where X = F, Cl, Br, I					
18-29	1-4	F	H	H	3.163+4.725
		F	F	H	
		F	F	F	

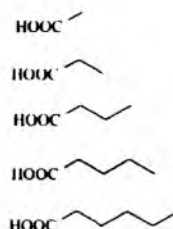
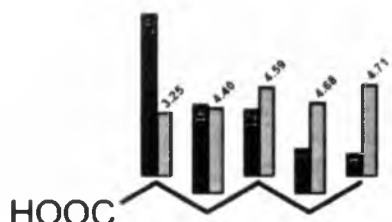


Fig. (2). The incremental addition of a carbon atom to the base carboxylic acid subunit.

Table 2. The FRAGTAL Descriptors for a Carboxylic Acid Substructure Terminated at the Given Carbon Atom

Descriptor Type	C ¹	C ²	C ³	C ⁴	C ⁵
FRAGTAL _n ^a	29	16	12	8	4
mean(FRAGTAL _{pKa}) ^b	3.25	4.40	4.59	4.68	4.71
max(FRAGTAL _{pKa}) ^c	4.79	4.79	4.78	4.78	4.78
min(FRAGTAL _{pKa}) ^d	0.05	3.16	4.12	4.56	4.66

^aFrequency descriptor.^bMean value of the FRAGTAL pKa value.^cMaximal value of the FRAGTAL pKa value.^dMinimal value of the FRAGTAL pKa value.**Fig. (3).** Histogram of FRAGTAL frequency (black) and pKa (grey) for carboxylic acid structures terminating at a given carbon atom.

2.4. Molecular Descriptors

The values of the chosen molecular descriptors were calculated for the Ligand.Info Anti-HIV NCI subset as well as for the 6 reference ChemDB Anti-HIV/OI/TB subsets shown in Table 3. The structural graph representation was converted into the canonically-coded linear atom/bond specification preserving the molecular topology and subsequently processed into a SMILES/SMARTS string notation. A series of 186 descriptors that are available in the MOE 2010 suite were evaluated and augmented with 4 Ro5 descriptors (MW, HBA, HBD, ClogP).

Table 3. The Number of High/Low Active Compounds in the ChemDB Anti-HIV/OI/TB Database

Nr	Enzyme/Activity	Number of Compounds
1	RT/low	10
2	RT/high	10
3	P/low	17
4	P/high	10
5	IN/low	11
6	IN/high	11

^aIntegrase: IN; protease: P; reverse transcriptase: RT.

Moreover, the set of the produced 190 descriptors was supplemented with the activity IC₅₀ value and 7 flags denoting its adherence to the subsets corresponding to a seven-step incremental construction of the full FRAGTAL superstructure of the DKA query - a particular FRAGTAL flag contains all of the compounds of the consecutive step. At this stage, the

Lipinski Ro5 filter was directly applied within the MoStBioDat environment. Thus, the entire ensemble of descriptors for the particular compound counts equally 198. A comprehensive list of the applied descriptors is given in the APPENDIX.

2.5. Mapping FRAGTAL Descriptors Using Self-Organizing Neural Networks

A self-organizing neural network (SOM) is comprised of a single layer of neurons typically arranged as a hexagonal or rectangular array of nodes using the unsupervised learning rules initially proposed by Kohonen [52]. The two-dimensional topology of the neural grid with the defined winning and neighborhood distances between individual neurons directly specify the mutual relations among the neurons. The presented m -dimensional input vector $x_i = (x_{i1}, \dots, x_{im})$ is distributed between neurons according to the similarity/correlation weight criteria - similar inputs are located in the same or proximal nodes. The typical competitive Kohonen procedure is based on the comparison of the input vector to the corresponding m -element weigh vectors $w_j = (w_{j1}, \dots, w_{jm})$ that describe each neuron in order to detect the winning one (out_c) into which the particular input will be projected. Practically, the out_c is selected by the optimization of the Euclidean distance between the input (x) and a weigh vector (w) according to the formula:

$$out_c = \min (x_i - w_{ij})^2 \quad (1)$$

where:

$j = 1, \dots, n$ - refers to a particular neuron, n -refers to number of neurons, m - is the number of weights per neuron and i - indicates a particular input.

The weights of the winning and the neighboring neurons are then adapted in order to attract any similar input vectors. While the subsequent input vector is being processed, the entire algorithm is recurred. Consequently, the multidimensional space is transformed into a two- or three-dimensional neural space that maintains the topological relationships among the input objects. The KNN forms the base for the projection of the selected molecular feature prescribed to the input vector, thus producing a color-coded planar clustering pattern. The implementation of the SOM for the classification, visualization and compression of the structural data has been widely reported [53, 54].

The Kohonen color-coded mapping of the property distribution pattern of FRAGTAL descriptors was conducted in the MATLAB programming environment using the SOM Toolbox 2.0 [55]. The unsupervised SOM learning procedure was completed for a set of 198 normalized descriptor data in the batch training mode for a hexagonal 20x20 map using the Gaussian neighborhood function.

3. RESULTS AND DISCUSSION

3.1. Anti-HIV Database Queries

By applying the FRAGTAL procedure, the population of the HIV-1 integrase inhibitors that are accessible in the ChemDB Anti-HIV/OI/TB database that was focused on diketo acid (DKA) and catechol derivatives was investigated. It is a

well-established and reliable repository which stores structural and pharmacological information about anti-HIV inhibitors and contains references to literature as well [56]. When not specified otherwise, all of the results of the *on-line* searches are limited to the HIV-1 integrase (IN) inhibitors obtained using a chemical search with the Boolean operator of conjunction (target class = HIV-inhibition and individual structure query). Unfortunately, the *on-line* graphical user interface (GUI) of the ChemDB Anti-HIV/OI/TB database is not adapted for the batch processing of multiple queries; therefore, we used the Ligand.Info repository to collect anti-HIV NCI subset data as well [57]. The Ligand.Info hits were extracted using the MoStBioDat storage protocols, thereby preserving the high-standards of data integrity, which resulted in the creation of an *in-house* data collection. The implemented library of the procedures integrated with the external pythonic modules automated the complicated task of screening the molecular diversity and exploring the structural knowledge. In fact, the conjunction of the expressive scripting language with the specified functionalities of the high-performance database system makes the subsequent sampling of the virtual molecules with an activity analysis or calculation of Ro5 descriptors possible. The subset of the chosen structures was limited to compounds having the specified IC_{50} cell activity.

The number of structures enumerated by the highest and lowest activity data against individual enzymes registered in the *on-line* Anti-HIV/OI/TB database is listed in Table 3.

3.2. Probing FRAGTAL for Diketo Acid Analogues

Since we applied the databases for the more or less intuitive design of the HIV integrase (IN) inhibitors, our strategy can be formulated relatively simply as: *finding possibly active fragments of drugs, modify them and put together* [58]. At the same time, the modification should provide a new class and cannot be too extensive in order not to completely lose the analogy and in turn the activity. On the basis of chemoinformatics, this strategy can be connected to the total occurrence of a certain fragment namely the hit frequency within the ChemDB Anti-HIV/OI/TB database.

Integrase is an important target for anti-AIDS therapy, which also attracts quinoline-based drug candidates. The main point here is not only the key role of IN in viral replication, but also the fact that this enzyme is a unique target that does not have a host cell counterpart [59, 60]. IN catalyses the insertion of reverse-transcribed viral DNA into the host cell genome in two distinct steps, namely, 3'-processing (3'P) and strand transfer (ST). The serendipitous discovery of the diketo acid (DKA) pharmacophore seemed to provide selective ST inhibitors that have a high activity against IN. However, DKAs needed some modifications in order to produce promising drug candidates like S-1360, L-870,810 that could hold up to the clinical trial level or to provide the first commercial anti-IN drug, raltegravir. Although styrylquinolines (SQ) may look similar, the mechanism of action appears to be quite different (Fig. 4) [61].

In Fig. (5) we reconstructed the way in which we disclosed an offspring DKA compound, ethyl malonate amide **2**, by finding synthetic targets by using database searches. Since we were engaged in optimizing the SQs, we were especially interested in the importance of the arrangement of the carboxylic function vs quinoline nitrogen. Thus, we started from the N-substituted aromatic ring that was differently ornamented with carboxylic functionality. In fact, we found that the arrangement of the carboxylic function in quinoline is important for the activity level of the starting fragment [62, 63]. Fig. (5a) illustrates an incremental construction of the molecular fragment starting from the base core (bold lines) through the successive addition of atoms from up to down, which is presented on the virtual network (dotted lines). In Fig. (5b) the same procedure is presented (in the horizontal line) using chemical structures. The bars labeling the atoms of the incrementally growing fragment illustrate the hit frequency of the compounds that contain this specific fragment. The detailed frequency data for this example is given in Table 4. In the context of the frequency of the hits registered, the probed fragment appeared to be relatively stable. It is, however, only the last increment of the carboxylic O atom that suddenly

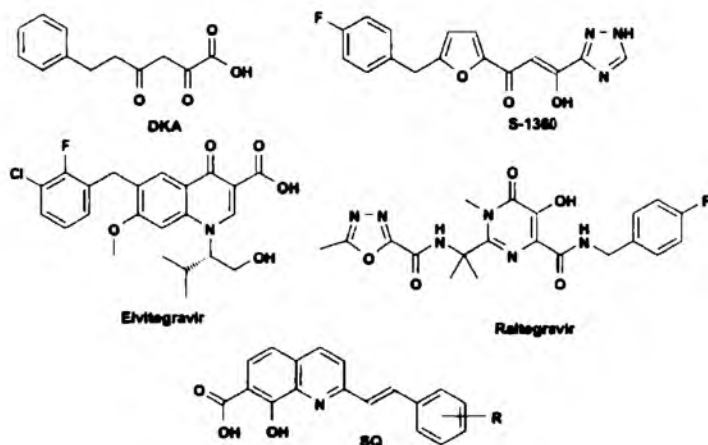


Fig. (4). The structure of DKA-based IN inhibitors.

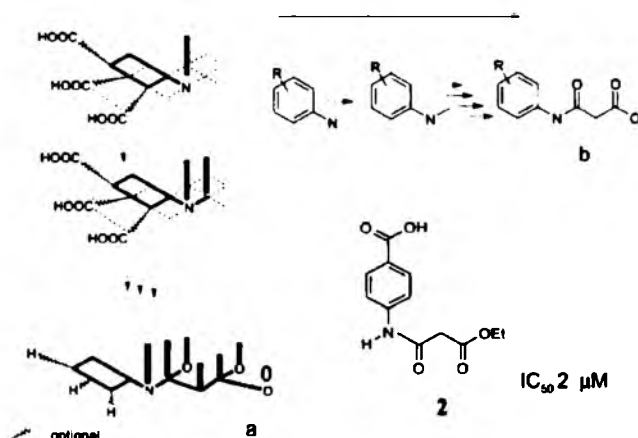


Fig. (5). An incremental construction of the FRAGTAL superfragment on a virtual grid (a) illustrated by chemical structures (b). The bars in (a) illustrate the frequency of the incremental FRAGTAL descriptors schematically as probed in ChemDB Anti-HIV/OI/TB database starting from left to right. Table 4 and Fig. (7) give the detailed FRAGTAL data. Compound 2 was synthesized and evaluated in our labs on the basis of similar database mining [63].

diminished the FRAGTAL value, e.g. not a single compound (0 hits) that has this motif is registered in the database. Thus, although this addition is risky, at the same time this provides a new unregistered target, which is an important goal of drug development. In fact, we synthesized compound 2 and found a new lead structure [63].

Table 4. The FRAGTAL Frequency (IN ChemDB Anti-HIV/OI/TB Inhibition Data) for Fragments Shown in Fig. (5b) Listed Incrementally in Rows for Different Aromatics, Respectively

Aromatics	N ¹	C ¹	O ¹	C ²	C ³	O ²	O ³
All H	1234	1192	794	752	663	633	0
4-COOH	96	82	82	82	80	77	0
3-COOH	68	57	43	43	38	30	0
2-COOH	11	5	5	5	1	0	0

3.3. Probing FRAGTAL for Caffeic Analogues

In Fig. (6) we present the tested FRAGTAL pattern for the caffeic acid (also referred as catecholics) that was derived from the fragments. In this case, the main reason why we synthesized compound 3 (synthetic data available in APPENDIX) is not only as a model compound but we had also found a number of caffeic acid amides with a relatively high activity in the databases. Now, the extension by a single carbocyclic-membered ring provides a fragment of two registered compounds (one of which is shown in Fig. 6). Active compound 3 indicates a fair similarity to the database hit. The heterocyclic quinoline ring provides an isosteric carboxylic group replacement; however, unlike compound 3, the database hit includes a double caffeic acid moiety.

3.4. Off-Line Database Search

All procedures related to the FRAGTAL concept described so far can be performed by *off-line* database searches as well. In Tables 5 and 6, and Fig. (7), we present the FRAGTAL landscape for compound 2 based on the data extracted from an *in-house* Anti-HIV NCI repository after the Ligand.Info Anti-HIV subset was downloaded and subsequently analyzed automatically in the batch mode using the MoStBioDat programming package. The database copy which was available for downloading (Ligand.Info Anti-HIV NCI) did not flag the specific target of the recorded anti-HIV agents, i.e. integrase (IN), protease (PR) or reverse-transcriptase (RT), which included similar molecular fragments. Thus, the activity value is the only parameter that can be extracted and analyzed. A polypharmacological study was then performed, e.g. all compounds were included regardless of the affected anti-HIV target and all types of anti-HIV activity were overlapped as registered in the database. At first, the results might look quite surprising because the lower the fragment is, the higher the activity and the lower the standard deviation (error-like bars in the histogram shown in Fig. 7), respectively. This trend is especially well preserved for the densely populated all-H subunit. One might expect quite a reverse relationship, since with each atom that is added, we are closer to the high activity DKA IN inhibitors. Nevertheless, in comparison with other targets, e.g. reverse transcriptase, the numerical values of enzyme inhibition for IN were relatively high (lower values indicate higher activity and *vice versa*). Moreover, the closer we are to the IN specific DKA then we should observe the lower specificity towards other enzymes, i.e., RT and PR. *Vice versa* among lower fragments RT and/or protease activity level dominates over that of IN which results in higher activity values (compare Table 7).

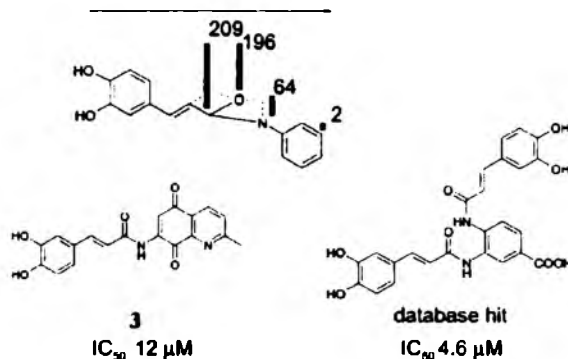


Fig. (6). FRAGTAL analysis of a caffeic fragment. Details in text.

Apparently, the disclosed effect illustrates the discriminative power of the fragment under investigation. Incrementally growing fragments that form favorable integrase inhibitors are not necessarily advantageous for the construction of RT ones for which the preferred fragment should be designed differently. If so, following the preferential IN FRAGTAL, we decreased the activity assayed vs other enzyme targets.

Table 5. Polypharmacological FRAGTAL Frequency (All Enzyme Ligand.Info Anti-HIV NCI Data) for the Fragments Shown in Fig. (5b) Listed Incrementally in Rows for Different Aromatics, Respectively

Aromatics	N ¹	C ¹	O ¹	C ²	C ³	O ²	O ³
All H	9654	6461	3920	2972	2093	359	9
4-COOH	141	71	31	18	8	3	1
3-COOH	134	36	17	7	1	1	0
2-COOH	84	28	18	15	8	0	0

Table 6. Polypharmacological Mean FRAGTAL Activity (All Enzyme log IC_{50} Ligand.Info Anti-HIV NCI Data) for the Fragments Shown in Fig. (5) Listed Incrementally in Rows for Different Aromatics, Respectively; the Standard Deviation is Given in Parentheses

Aromatics	N ¹	C ¹	O ¹	C ²	C ³	O ²	O ³
All H	-4.397 (0.793)	-4.374 (0.786)	-4.259 (0.672)	-4.251 (0.647)	-4.256 (0.638)	-4.157 (0.620)	-3.941 (0.410)
4-COOH	-4.452 (0.706)	-4.646 (0.778)	-4.326 (0.531)	-4.315 (0.460)	-4.325 (0.410)	-3.907 (0.292)	-4.320 (0)
3-COOH	-4.226 (0.677)	-4.340 (0.512)	-3.913 (0.381)	-3.784 (0.289)	-3.700 (0)	-3.700 (0)	N/A (N/A)
2-COOH	-4.286 (0.810)	-4.215 (0.659)	-4.176 (0.605)	-4.221 (0.599)	-4.486 (0.649)	N/A (N/A)	N/A (N/A)

Compounds 2 and 3 were also synthesized as model compounds mimicking N vs COOH arrangement in SQ

because derivatives of 6-quinolinecarboxylic acid appeared more active than those from 8-quinolinecarboxylic acid [63]. In the aspect of the optimization of compounds related to 2, the main conclusion from the current analysis is that unsubstituted (all-H) analogue 2 seems to be an interesting target for further synthesis. FRAGTAL nicely illustrates this rule. Moreover, the activity remains relatively stable within the COOH series with the increasing length of the FRAGTAL superfragment, which indicates its IN affinity.

Table 7. EC_{50} and IC_{50} Activity Ranges for IN and Reverse Transcriptase (RT)

Enzyme	Number of Hits of the Activity Level Below			
	10^{-4}	10^{-3}	10^{-2}	10^{-1}
IN EC_{50}	1	8	245	578
IN IC_{50}	1	1	21	70
RT EC_{50}	60	626	2009	4616
RT IC_{50}	3	13	28	99

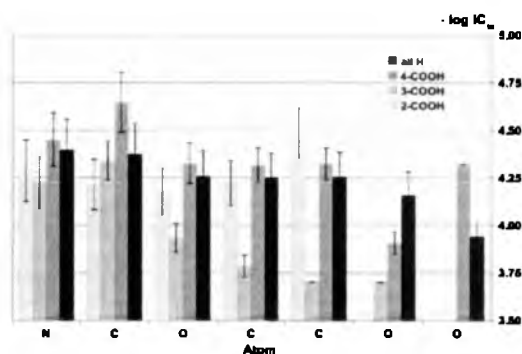


Fig. (7). Polypharmacological FRAGTAL activity (fragments defined in Fig. 5b) for different aromatics, respectively. Overall, activity decreases with an increasing number of fragments. Details in text.

3.5. Self-Organizing Maps for the Visualization of FRAGTAL Descriptors

The polypharmacological activity of the offspring DKA compounds (all-H subunit) identified during the FRAGTAL visualization of an anti-HIV subset was further analyzed using the self-organizing neural network technique. Initially, each compound was annotated by a set of 186 molecular descriptors that had been calculated in the MOE 2010 suite and supplemented by 4 Ro5 descriptors (MW, IIBA, IIBD, ClogP). Additionally, the ensemble of the resulting 190 molecular descriptors was augmented with the activity IC_{50} value and 7 flags denoting its adherence to the subsets corresponding to a seven-step incremental construction of the full FRAGTAL superstructure of the DKA query. A particular FRAGTAL flag contains all of the compounds of the consecutive step, which means that e.g. 9654 molecules have a flag 1 set to 1, flag 2 set to 1 for 6461 hits etc. (see Table 5). The winning neuron of the map was identified taking into account its similarity to the ensemble of 190 descriptors; however, the entire 198 element weigh vector attributed to each neuron was adjusted at the training stage. Thus, the 8-digit weight vector corresponding to the IC_{50} activity and the FRAGTAL 1-7 membership group constitutes an 'image' of the distribution of these elements on the SOM map. Regarding the Gaussian neighborhood function, the weight-color-coded map might be interpreted as a rough distribution of the molecular population density specified by the corresponding FRAGTAL steps.

The color-coded projection of the activity and flag descriptors on the hexagonal 20x20 map trained by the normalized set of 198 descriptors specified for each structure enabled the analysis of the distribution of the fragmental-based property.

Fig. (8) shows the distribution of the IC_{50} activity over the whole set of DKA derivatives. The higher values indicated by the bright color scale represent lower activity while the lower values marked by the dark color range represent high activity. The crude separation of active compounds from inactive ones is noticeable. It seems that the three clusters which contain more active molecules are more distinctive due to the lower number of less structurally diverse compounds in contrast to the groups formed by non-active ones.

The distribution of the molecular population identified in the consecutive FRAGTAL stages indicated a tendency to form several clusters as shown in Fig. (9a-f). The initial addition of the carbon atom to the base fragment resulted in tiny changes in the distribution of the molecular population as illustrated in Fig. (9a). Not surprisingly, the addition of one carbon atom to the 'core' fragment generated a simple (sub)structural motif that is present in a wide range of anti-HIV compounds and other drugs as well. Several important regularities were revealed by the analysis of the maps. Fig. (9b) illustrates the emergence of three main clusters related to the step 3 FRAGTAL. This distribution was basically preserved in subsequent incremental fragments that increased in steps 4 and 5 as shown in Fig. (9c, d), respectively. The mapping of the DKA hit ensemble produced by FRAGTAL step 6 slightly changed the cluster pattern as indicated in Fig. (9e). In this case, the number of

identified compounds decreased from 2093 to 359 (see Table 5), which is reflected in two distinctive groups where only one corresponded well to the clusters formed by the previous FRAGTALs. Regarding the chemotype of the complexing two-metal DKA divalent model, it is clear that the addition of a carbonyl oxygen atom of the carboxylic functional group in step 6 forms a scaffold which is crucial for the anti-HIV IN moiety. Finally, the increment of the carboxylic pattern with a hydroxylic oxygen atom (step 7) dramatically decreased the number of hits specified in both the ChemDB Anti-HIV/OI/TB subset and in the Ligand.Info Anti-HIV NCI subset, respectively (see Tables 4 and 5).

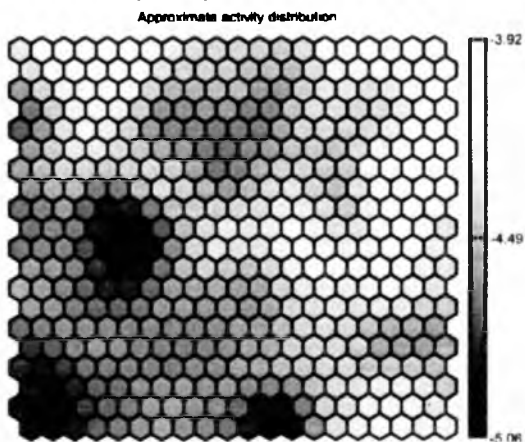


Fig. (8). SOM of the approximate anti-HIV $\log IC_{50}$ activity distribution in the chemical space constituted by the Ligand Info Anti-HIV NCI subset database.

Another interesting issue that should be considered carefully is the mutual distribution of well-known highly active HIV-1 inhibitors on the neural map. Is it possible to indicate the densely populated regions of the map by the concentration of high affinity compounds? Is there any direct correspondence to the molecular fragments specified in the consecutive FRAGTAL database survey? The Kohonen mapping of the IC_{50} distribution pattern coupled with identifying the neurons that attract the high activity representatives of the reverse transcriptase, protease and integrase inhibitors is given in Fig. (10). Surprisingly, only the protease inhibitors, which are marked by x hexagons, are located relatively close to the high active regions of the map whereas the P and IN ones are evenly scattered over the SOM network. Roughly speaking, the Ligand.Info Anti-HIV NCI database subset was last updated in 2006 while the ChemDB Anti-HIV/OI/TB is relatively up-to-date; therefore, the P inhibitors considerably outnumber the RT and IN inhibitors, respectively.

The color-coded comparison of the molecular SOM projection of DKA offspring, which are identified by consecutive FRAGTALs conjugated with the distribution of previously described HIV-1 inhibitors, is illustrated in Fig. (11a, b). The structural classification by FRAGTAL of step 3 exhibited a decent coherence with high active IN inhibitors (Fig. 11a); however, this of the step 6 illustrates the

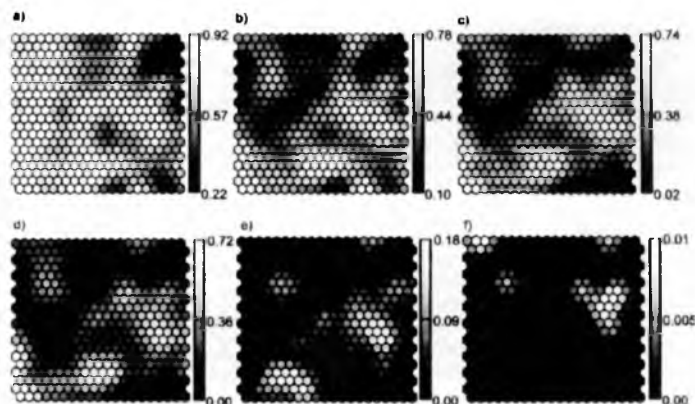


Fig. (9). SOMs of the distribution of the molecular population density identified by FRAGTAL steps 2-7 of compound 2 obtained by querying the Ligand.Info Anti-HIV NCI subset database. a-f correspond to steps 2-7, respectively. The ranges displayed show the molecular population density relative to the FRAGTAL step 1.

Table 8. The Representative Topologies of Compounds with a Different Location of a DKA-like Moiety within a Molecular Pattern

Cluster 1		Cluster 2	
SMILES	ID	SMILES	ID
<chem>CC(C)OC1=C(C)=C2C(=C1)N(C=C(C2=O)C(=O)O)C(=O)C(=O)C(C)C(C)=C(C)=CC=C3C(=O)F</chem>	5277123	<chem>C1CCSi(=O)(=O)N(C1)C2=NC(=C(C3=C2C=CC=N3)OC(=O)NC(=O)C=CC=C(C=C4)F</chem>	457930
<chem>CC(C)C(C(=O)N1C(=O)C(=O)C2=CC(=CC(=C2)O)C(C)C(C)C3=C(C)=CC=C3C(=O)F)C(=O)O</chem>	5277128	<chem>CC1(C2=N(C(=C(NCC3=CC=C(C=C3)F)O)C1=O)C(=O)N2CCN1CC4=NNN=N4)C</chem>	11948113
<chem>CC(C)OC1=C2C(=CC(=C1)CC3=C(C(=CC=C3)C(F)C(=O)C(=O)N2)C(=O)C(C)C(C)C(=O)O</chem>	5277129	<chem>CC1(C2=N(C(=C(NCC3=CC=C(C=C3)F)O)C1=O)C(=O)N2CCN1CC4=CN(N=N4)C)C</chem>	16078898
<chem>CC(C)C(C(=O)N1C(=O)C(=O)C2=CC(=CC(=C2)O)C(C)C(C)C3=C(C)=CC=C3C(=O)F)C(=O)O</chem>	5277136		
<chem>CCOC1=C2C(=CC(=C1)CC3=C(C(=CC=C3)C(F)C(=O)C(=O)N2)C(=O)C(C)C(C)C(=O)O</chem>	5277140		

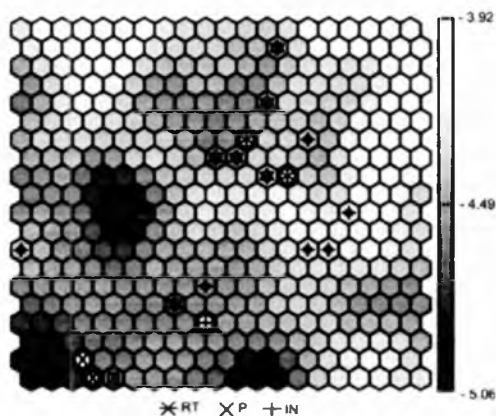


Fig. (10). SOM of the approximate activity distribution as shown on Fig. (8) supplemented by an indication of the neurons attracting the high activity representatives of the reverse transcriptase (RT - * hexagons), protease (P - x hexagons) and integrase (IN - + hexagons) inhibitors. The size of the hexagons reflects the number of hits attracted by a particular neuron

distribution changes shown in Fig. (9e). Surprisingly, a large number of high active IN inhibitors in this case are situated in the 'poorly' occupied neural area, which is shown in Fig. (11b). A detailed analysis of compounds distributed among two distinctive clusters revealed the main structural difference in the location of the DKA-like moiety within the molecular pattern. The first one is dominated by molecules where the DKA functional subunit is located 'terminally', whereas the second one is 'hidden internally' within the given structure. Some representative structural topologies that form the above groups are presented in Fig. (12a, b), whereas the corresponding SMILES codes are presented in Table 8.

4. CONCLUSIONS

In conclusion, only a better knowledge of the CS can provide us with more efficient chemistry and drug design. Many problems were encountered when probing the CS. Hence, we presented a new method based on database mining, namely, the fragmental topology-activity landscape (FRAGTAL), which is designed to mimic the way in which active compounds are designed in the ligand-based mode by a medicinal chemist searching in database. It should be

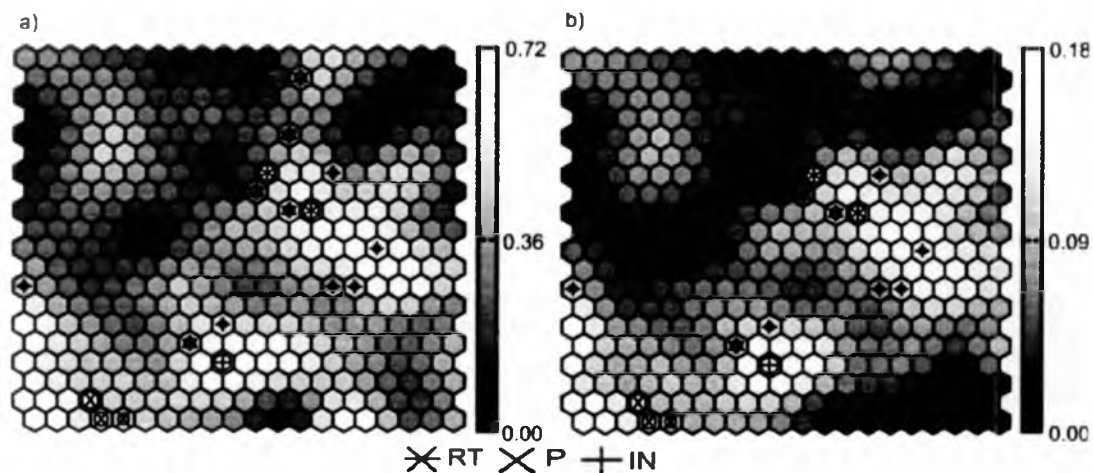


Fig. (11). SOMs of the distribution of the molecular population density as shown on Fig. (9d, e) supplemented by an indication of the neurons attracting the high activity representatives of the three HIV enzyme inhibitors (see Fig. 10). Figures a and b correspond to Fig. (9d, e), respectively.

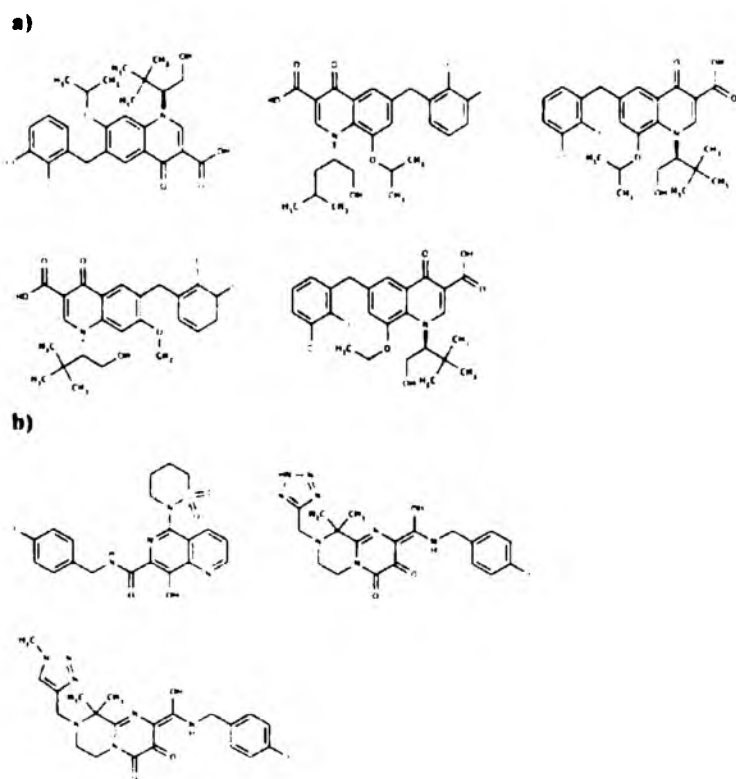


Fig. (12). The representative structures of compounds with a different location of the DKA-like moiety within a molecular pattern.

indicated here that we are not pretending that we have precisely designed the compounds presented. Instead, we have tried to reconstruct the way in which our synthetic targets were found in the hope of increasing the chances of finding actives while at the same time exploring new CS regions. In reality, our approach can be seen as an adaptation of the frequent subgraph mining (FSM) method. FRAGTAL can be performed in a simple *on-line* mode but can also include more sophisticated analyses that demand massive database screening which can be efficiently performed by the application of the MoStBioDat package. Moreover, we show here how FRAGTAL descriptors can be used, analyzed, visualized and interpreted, e.g. by using the SOM mapping as a tool for the interpretation of polypharmacological data. We are now testing the extensive application of the FRAGTAL/MoStBioDat system for the analyses of polypharmacological problems involving the parallel analysis of

small molecules and target protein macromolecular receptors.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

We would like to acknowledge the OpenEye and OpenBabel Scientific Software for the free software academic license. Dr. Andrzej Bak would also like to thank the Foundation for Polish Science for his individual grant, Dr. Stian Soiland-Reyes for the ForgetSQL2 python module and Dr. Ian Bicking for the SQLBuilder python module. This work was partially financed by Polish Ministry of Science grant N NS19 575638.

APPENDIX

1. The List of the MOE Descriptors Used for Training SOM Maps

Apol, a_acc, a_acid, a_aro, a_base, a_count, a_don, a_heavy, a_hyd, a_IC, a_ICM, a_nB, a_nBr, a_nC, a_nCl, a_nF, a_nH, a_nI, a_nN, a_nO, a_nP, a_nS, balabanJ, BCUT_PEOE_0, BCUT_PEOE_1, BCUT_PEOE_2, BCUT_PEOE_3, BCUT_SLOGP_0, BCUT_SLOGP_1, BCUT_SLOGP_2, BCUT_SLOGP_3, BCUT_SMR_0, BCUT_SMR_1, BCUT_SMR_2, BCUT_SMR_3, bpol, b_1rotN, b_1rotRb_ar, b_count, b_double, b_heavy, b_rotN, b_rotR, b_single, b_triple, chi0, chi0v, chi0v_C, chi0_C, chi1, chi1v, chi1v_C, chi1_C, Chirac, chiral_u, density, diameter, FCharge, GCUT_PEOE_0, GCUT_PEOE_1, GCUT_PEOE_2, GCUT_PEOE_3, GCUT_SLOGP_0, GCUT_SLOGP_1, GCUT_SLOGP_2, GCUT_SLOGP_3, GCUT_SMR_0, GCUT_SMR_1, GCUT_SMR_2, GCUT_SMR_3, Kier1, Kier2, Kier3, KierA1, KierA2, KierA3, KierFlex, lip_acc, lip_don, lip_druglike, lip_violation, logP(o/w), logS, mr, mutagenic, nmol, opr_brigid, opr_leadlike, opr_nring, opr_nrot, opr_violation, PC+, PC-, PEOE_PC+, PEOE_PC-, PEOE_RPC+, PEOE_RPC-, PEOE_VSA+0, PEOE_VSA+1, PEOE_VSA+2, PEOE_VSA+3, PEOE_VSA+4, PEOE_VSA+5, PEOE_VSA+6, PEOE_VSA-0, PEOE_VSA-1, PEOE_VSA-2, PEOE_VSA-3, PEOE_VSA-4, PEOE_VSA-5, PEOE_VSA6, PEOE_VSA_FHYD, PEOE_VSA_FNEG, PEOE_VSA_FPNEG, PEOE_VSA_FPOL, PEOE_VSA_FPOS, PEOE_VSA_FPPOS, PEOE_VSA_HYD, PEOE_VSA_NEG, PEOE_VSA_PNEG, PEOE_VSA_POL, PEOE_VSA_POS, PEOE_VSA_PPOS, petitjean, petitjeanSC, Q_PC+, Q_PC-, Q_RPC+, Q_RPC-, Q_VSA_FHYD, Q_VSA_FNEG, Q_VSA_FPNEG, Q_VSA_FPOL, Q_VSA_FPOS, Q_VSA_FPPOS, Q_VSA_HYD, Q_VSA_NEG, Q_VSA_PNEG, Q_VSA_POL, Q_VSA_POS, Q_VSA_PPOS, radius, reactive, rings, RPC+, RPC, rsynth, SlogP, SlogP_VSA0, SlogP_VSA1, SlogP_VSA2, SlogP_VSA3, SlogP_VSA4, SlogP_VSA5, SlogP_VSA6, SlogP_VSA7, SlogP_VSA8, SlogP_VSA9, SMR, SMR_VSA0, SMR_VSA1, SMR_VSA2, SMR_VSA3, SMR_VSA4, SMR_VSA5, SMR_VSA6, SMR_VSA7, TPSA, VAdjEq, VAdjMa, VDistEq, VDistMa, vdw_area, vdw_vol, vsa_acc, vsa_acid, vsa_base, vsa_don, vsa_hyd, vsa_other, vsa_pol, WeightweinerPath, weinerPol, zagreb

2. The NMR Data of Synthesized Compound 3.

¹H NMR (DMSO-d₆) 1: 1.2 (t, J= 7.1, 3H, CH₃); 3.48 (s, 2H, CH₂); 4.1 (m, 2H, O-CH₂); 7.6 (d, J=8.7, 2H, Ar-H); 7.8 (d, J=8.6, 2H, Ar-H); 10.5 (s, 1H, N-H); 12.7 (s, 1H, OH); 2: 2.89 (s, 3H, CH₃-Ar) 5.8 (s, 1H, Ar-H); 6.25 (d, 1H J=15.8, vinyl); 7.1 (d, J=8.2, 2H, Ar-H); 6.82 (d, J= 8.13, 1H, Ar-H) 6.98 (d, 1H J= 8.18; Ar-H); 7.06 (s, 1H, Ar-H); 7.48 (d, 1H, J=15.8, vinyl); 7.62 (d, 1H, J=8.15, Ar-H); 8.14 (d, 1H, J=8.11, Ar-H)

REFERENCES

- [1] Oprea, T.I. Property distribution of drug-related chemical databases. *J. Comput. Aided Mol. Des.*, **2000**, *14*, 251-264.
- [2] Fukunishi, Y.; Nakamura, H. Definition of drug-likeness for compound affinity. *J. Chem. Inf. Model.*, **2011**, *51*, 1012-1016.
- [3] Veber, D.F.; Johnson, S.R.; Cheng, H.Y.; Smith, B.R.; Ward, K.W.; Kopple, K.D. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.*, **2002**, *45*, 2615-2623.
- [4] Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.*, **2001**, *46*, 3-26.
- [5] Lipinski, C.A.; Hopkins, A. Navigating chemical space for biology and medicine. *Nature*, **2004**, *432*, 855-861.
- [6] Ghose, A.K.; Herbertz, T.; Salvino, J.M.; Mallamo, J.P. Knowledge-based chemoinformatic approaches to drug discovery. *Drug Discov. Today*, **2006**, *11*, 1107-1114.
- [7] Stahl, M.; Mauser, H. Database clustering with a combination of fingerprint and maximum common substructure methods. *J. Chem. Inf. Model.*, **2005**, *45*, 542-548.
- [8] Heikamp, K.; Bajorath, J. How do 2D fingerprints detect structurally diverse active compounds? Revealing compound subset-specific fingerprint features through systematic selection. *J. Chem. Inf. Model.*, **2011**, *51*, 2254-2265.
- [9] Steffen, A.; Kogej, T.; Tyrchan, C.; Engkvist, O. Comparison of molecular fingerprint methods on the basis of biological profile data. *J. Chem. Inf. Model.*, **2009**, *49*, 338-347.
- [10] Johnson, M.A.; Maggiora, G.M. *Concepts and applications of molecular similarity*; Wiley & Sons: New York, **1990**.
- [11] Zarlter, E.R.; Shapiro, M.J. Fragonomics: fragment-based drug discovery. *Curr. Opin. Chem. Biol.*, **2005**, *9*, 366-370.
- [12] Erlanson, D.A.; McDowell, R.S.; O'Brien, T. Fragment-based drug discovery. *J. Med. Chem.*, **2004**, *47*, 3463-3482.
- [13] Hoffer, L.; Renaud, J.P.; Horvath, D. Fragment-based drug design: computational & experimental state of art. *Comb. Chem. High Throughput Screen.*, **2011**, *14*, 500-520.
- [14] Gianti, E.; Sartori, L. Identification and selection of "privileged fragments" suitable for primary screening. *J. Chem. Inf. Model.*, **2008**, *48*, 2129-2139.
- [15] Salum, L.B.; Andricopulo, A.D. Fragment-based QSAR strategies in drug design. *Expert Opin. Drug Discov.*, **2010**, *5*, 405-412.
- [16] Kolb, P.; Cafflisch, A. Automatic and efficient decomposition of two-dimensional structures of small molecules for fragment-based high-throughput docking. *J. Med. Chem.*, **2006**, *49*, 7384-7392.
- [17] Lameijer, E.W.; Kok, J.N.; Back, T.; Ijzerman, A.P. Mining a chemical database for fragment co-occurrence: discovery of "chemical clichés". *J. Chem. Inf. Model.*, **2006**, *46*, 553-562.
- [18] Baskin, I.; Varnek, A. Building a chemical space based on fragment descriptors. *Comb. Chem. High Throughput Screen.*, **2008**, *11*, 661-668.
- [19] Congreve, M.; Carr, R.; Murray, C.; Shoji, H. A "rule of three" for fragment-based lead discovery? *Drug Discov. Today*, **2003**, *8*, 876-877.
- [20] Hestekamp, T.; Whittaker, M. Fragment-based activity space: smaller is better. *Curr. Opin. Chem. Biol.*, **2008**, *12*, 260-268.
- [21] Ursu, O.; Oprea, T.I. Model-free drug-likeness from fragments. *J. Chem. Inf. Model.*, **2010**, *50*, 1387-1394.
- [22] Hann, M.M.; Oprea, T.I. Pursuing the leadlikeness concept in pharmaceutical research. *Curr. Opin. Chem. Biol.*, **2004**, *8*, 255-263.
- [23] Bemis, G.W.; Murcko, M.A. The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.*, **1996**, *39*, 2887-2893.
- [24] Bemis, G.W.; Murcko, M.A. Properties of known drugs. 2. Side chains. *J. Med. Chem.*, **1999**, *42*, 5095-5099.
- [25] Lounkine, E.; Bajorath, J. Topological fragment index for the analysis of molecular substructures and their topological environment in active compounds. *J. Chem. Inf. Model.*, **2009**, *49*, 162-168.
- [26] Polanski, J.; Kurczyk, A.; Bak, A.; Musiol, R. Privileged structures - dream or reality: preferential organization of azanaphthalene scaffold. *Curr. Med. Chem.*, **2012**, *19*, 1921-1945.
- [27] Muegge, I.; Heald, S.L.; Britelli, D. Simple selection criteria for drug-like chemical matter. *J. Med. Chem.*, **2001**, *44*, 1841-1846.
- [28] Evans, B.; Rittle, K.; Bock, M.; DiPardo, R.; Freidinger, R.; Whitter, W.; Lundell, G.; Veber, D.; Anderson, P. Methods for drug discovery: development of potent, selective, orally effective cholecystokinin antagonists. *J. Med. Chem.*, **1988**, *31*, 2235-2246.
- [29] Kubinyi, H. Privileged structures and analogue-based drug discovery. In: *Analogue-based drug discovery*; Fischer, J.; Ganellin, C.R., Eds.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, **2006**; pp. 53-68.
- [30] Welch, M.E.; Snyder, S.A.; Stockwell, B.R. Privileged scaffolds for library design and drug discovery. *Curr. Opin. Chem. Biol.*, **2010**, *14*, 347-361.
- [31] Hajduk, P.J.; Bures, M.; Praestgaard, J.; Fesik, S.W. Privileged molecules for protein binding identified from NMR-based screening. *J. Med. Chem.*, **2000**, *43*, 3443-3447.
- [32] DeSimone, R.W.; Currie, K.S.; Mitchell, S.A.; Darrow, J.W.; Plippin, D.A. Privileged structures: applications in drug discovery. *Comb. Chem. High Throughput Screen.*, **2004**, *7*, 473-494.
- [33] Hestekamp, T.; Barker, J.; Davenport, A.; Whittaker, A. Fragment based drug discovery using fluorescence correlation spectroscopy techniques: challenges and solutions. *Curr. Top. Med. Chem.*, **2007**, *7*, 1582-1591.
- [34] Oprea, T.I.; Allu, T.K.; Fara, D.C.; Rad, R.F.; Ostropovici, L.; Bologa, C.G. Lead-like, drug-like or "Pub-like": how different are they? *J. Comput. Aided Mol. Des.*, **2010**, *21*, 113-119.
- [35] Costantino, L.; Barlocco, D. Privileged structures as leads in medicinal chemistry. *Curr. Med. Chem.*, **2006**, *13*, 65-85.
- [36] Horton, D.A.; Bourne, G.T.; Smythe, M.L. The combinatorial synthesis of bicyclic privileged structures or privileged substructures. *Chem. Rev.*, **2003**, *103*, 893-930.
- [37] Schomburg, K.; Ehrlich, H.C.; Stierand, K.; Rarey, M. From structure diagrams to visual chemical patterns. *J. Chem. Inf. Model.*, **2010**, *50*, 1529-1535.
- [38] Kazius, J.; Nijssen, S.; Kok, J.; Bäck, T.; Ijzerman, A.P. Substructure mining using elaborate chemical representation. *J. Chem. Inf. Model.*, **2006**, *46*, 597-605.
- [39] Weininger, D.; Weininger, A.; Weininger, J.L. SMILES. 2. Algorithm for generation of unique SMILES notation. *J. Chem. Inf. Comput. Sci.*, **1989**, *29*, 97-101.
- [40] Lisurek, M.; Rupp, B.; Wichard, J.; Neuenschwander, M.; von Kries, J.P.; Frank, R.R.; Rademann, J.; Kuhne, R. Design of chemical libraries with potentially bioactive molecules applying a maximum common substructure concept. *Mol. Divers.*, **2010**, *14*, 401-408.
- [41] Chen, H.; Yang, Y.; Engkvist, O. Molecular topology analysis of the differences between drugs, clinical candidate compounds, and bioactive molecules. *J. Chem. Inf. Model.*, **2010**, *50*, 2141-2150.
- [42] Bak, A.; Polanski, J.; Stockner, T.; Kurczyk, A. MoSiBioDat - molecular and structural bioinformatics database. *Comb. Chem. High Throughput Screen.*, **2010**, *13*, 366-374.
- [43] Bak, A.; Magdziarz, T.; Kurczyk, A.; Polanski, J. Mapping fragmental drug-likeness in the MoSiBioDat environment: intramolecular hydrogen bonding motifs in α -ketonols. *Comb. Chem. High Throughput Screen.*, **2011**, *14*, 560-569.
- [44] Bak, A.; Magdziarz, T.; Polanski, J. Pharmacophore-based database mining for probing fragmental drug-likeness of diketone analogues. *SAR & QSAR Environ. Res.*, **2012**, *23*, 185-204.
- [45] Bak, A.; Magdziarz, T.; Kurczyk, A.; Polanski, J. Mapping drug architecture by MoSiBioDat - Rapid screening of intramolecular hydrogen bonded motifs in catechols. *Drug Dev. Res.*, **2011**, *72*, 209-218.
- [46] Fialkowski, M.; Bishop, K.J.M.; Chubukov, V.A.; Campbell, C.J.; Grzybowski, B.A. Architecture and evolution of organic chemistry. *Angew. Chem. Int. Ed.*, **2005**, *44*, 7263-7269.
- [47] Reutlinger, M.; Guba, W.; Martin, R.E.; Alanine, A.I.; Hoffmann, T.; Klemmer, A.; Hias, J.A.; Schneider, P.; Schneider, G. Neighborhood-preserving visualization of adaptive structure-activity landscapes: application to drug discovery. *Angew. Chem. Int. Ed.*, **2011**, *50*, 11633-11636.
- [48] *OEChem - Python theory manual*, OpenEye Scientific Software Inc., Santa Fe, NM 87508, **2008**.
- [49] O'Boyle, N.M.; Morley, C.; Hutchison, G.R. Pybel: a Python wrapper for the OpenBabel cheminformatics toolkit. *Chem. Cent. J.*, **2008**, *9*, 2-5.
- [50] Berman, H.; Henrick, K.; Nakamura, H.; Markley, J.L. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.*, **2007**, *35*, 301-303.
- [51] Available from: ACD/Labs. <http://www.acdlabs.com>

- [52] Zupan, J.; Gasteiger, J. *Neural networks and drug design for chemists*. 2nd ed.; Wiley-VCH: Weinheim, 1999.
- [53] Bak, A.; Polanski, J. Modeling robust QSAR 3: SOM-4D-QSAR with iterative variable elimination IVE-PLS: application to steroid, azo dye and benzoic acid series. *J. Chem. Inf. Model.*, **2007**, *47*, 1469-1480.
- [54] Bak, A.; Polanski, J. The 4D-QSAR study on anti-HIV HEPT analogues. *Bioorg. Med. Chem.*, **2006**, *14*, 273-279.
- [55] Stahl, M.T. Open-source software: not quite endsville. *Drug Discov. Today*, **2005**, *10*, 219-122.
- [56] National Institute of Allergy and Infectious Diseases. <http://chemdb.niaid.nih.gov/>
- [57] von Grothuis, M.; Pas, J.; Rychlewski, L. Ligand-Info: searching for similar small compounds using index profiles. *Bioinformatics*, **2003**, *19*, 2041-2042.
- [58] Hajduk, P.J.; Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug Discov.*, **2007**, *6*, 211-219.
- [59] Dayam, R.; Sanchez, T.; Clement, O.; Shoemaker, R.; Sei, S.; Neamati, N. -Diketo acid pharmacophore hypothesis. 1. Discovery of a novel class of HIV-1 integrase inhibitors. *J. Med. Chem.*, **2005**, *48*, 111-120.
- [60] Dayam, R.; Sanchez, R.; Neamati, N. Diketo acid pharmacophore. 2. Discovery of structurally diverse inhibitors of HIV-1 integrase. *J. Med. Chem.*, **2005**, *48*, 8009-8015.
- [61] Solomon, V.R.; Lee, H. Quinoline as a privileged scaffold in cancer drug discovery. *Curr. Med. Chem.*, **2011**, *18*, 1488-1508.
- [62] Musiol, R.; Serda, M.; Hensel-Bielowka, S.; Polanski, J. Quinoline-based antifungals. *Curr. Med. Chem.*, **2010**, *17*, 1960-1973.
- [63] Serafin, K.; Mazur, P.; Bak, A.; Laine, E.; Tchertanov, L.; Mouscadet, J.F.; Polanski, J. Ethyl malonate amides: a diketo acid offspring fragment for HIV integrase inhibition. *Bioorg. Med. Chem.*, **2011**, *16*, 5000-5005.