

Autoreferat

do wniosku o nadanie stopnia naukowego doktora habilitowanego

Andrzej Bąk

**Środowisko analiz masowych danych molekularnych – Molekularna
i Strukturalna Bioinformatyczna Baza Danych (MoStBioDat)**



Katowice 2014

Spis treści

1.	Dane osobowe	3
2.	Wykaz publikacji stanowiących osiągnięcie naukowe	3
3.	Omówienie wyników prac stanowiących osiągnięcie naukowe	6
3.1	Wprowadzenie	6
3.2	Cel rozprawy habilitacyjnej	7
3.3	Ogólny zarys problematyki badań z uwzględnieniem prac własnych	9
3.3	Omówienie wyników	13
3.3.1	Modelowanie RI 4D-QSAR	13
3.3.2	Architektura platformy MoStBioDat	16
3.3.3	Analiza podobieństwa cząsteczkowego przy użyciu środowiska MoStBioDat	18
3.3.4	Fragmentaryczne mapowanie lekotypii w środowisku MoStBioDat: Analiza strukturalna motywu katecholu, β -ketoenu oraz diketokwasu	21
3.3.5	Amidy etylowe kwasu malonowego – nowa grupa inhibitorów integrazy wirusa HIV	31
3.3.6	Metody fragmentaryczne oraz koncepcja motywów uprzywilejowanych	34
3.3.7	Sztuczne receptory w chemoinformatyce	38
3.3.8	Modelowanie RD 4D-QSAR	42
3.2	Podsumowanie i wnioski	44
3.3	Omówienie pozostałych osiągnięć naukowo-dydaktycznych	48
3.4	Literatura	49
4.	Aktywność naukowa	53
4.1	Przed uzyskaniem stopnia doktora	53
4.2	Po uzyskaniu stopnia doktora	54
5.	Podsumowanie dorobku naukowo-dydaktycznego	62

**Wykaz opublikowanych prac naukowych lub twórczych prac zawodowych oraz
informacja o osiągnięciach dydaktycznych, współpracy naukowej i popularyzacji
nauki**

1. Imię i nazwisko: **Andrzej Bąk**

2. Posiadane dyplomy i stopnie naukowe:

1995–2000, studia magisterskie, Uniwersytet Śląski, Wydział Matematyki, Fizyki i Chemii, kierunek Chemia, tytuł magistra chemii, „Diagramatyczne poprawki do modelu CCSDT na poziomie V, VI i VII rzędu MBPT” wykonana pod kierunkiem prof. dr hab. Stanisława Kucharskiego

2000–2004, studia doktoranckie, Uniwersytet Śląski, Wydział Matematyki, Fizyki i Chemii, kierunek Chemia, stopień doktora nauk chemicznych, „Modelowanie zależności 4D-QSAR techniką samoorganizującej się sieci neuronowej; opracowanie nowego formalizmu, optymalizacja i wykorzystanie do badania grup związków biologicznie czynnych” wykonana pod kierunkiem prof. dr hab. Jarosława Polańskiego

2003–2006, studia zawodowe, Uniwersytet Śląski, Wydział Matematyki, Fizyki i Chemii, kierunek Informatyka, tytuł zawodowy licencjata, „Chemoinformatics Toolbox – nowy pakiet funkcji dla Matlab 7.0” wykonana pod kierunkiem dr Sławomira Michalika

3. Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych:

2005–2006, asystent, Uniwersytet Śląski w Katowicach, Instytut Chemii

2006–2014, adiunkt, Uniwersytet Śląski w Katowicach, Instytut Chemii

2005–2006, administrator sieci uczelnianej, Uniwersytet Ekonomiczny w Katowicach

2007–2009, pracownik naukowy w Austrian Research Institute w Seibersdorf k. Wiednia, Austria

4. Wskazanie osiągnięcia wynikającego z art. 16 ust. 2 Ustawy z dnia 14 marca 2003 roku o Stopniach Naukowych i Tytule Naukowym oraz o Stopniach i Tytule w zakresie Sztuki (Dz. U. nr 65, poz. 595 ze zm.)

4A. Tytuł osiągnięcia naukowego:

Środowisko analiz masowych danych molekularnych – Molekularna i Strukturalna Bioinformatyczna Baza Danych (MoStBioDat)

4B. Wykaz publikacji stanowiących osiągnięcie naukowe:

[H1]. **A. Bak**, J. Polanski, **2007**, 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.*, 47, 1017-1021. **IF₂₀₀₇=3,1 MSWiA₂₀₀₇=32** (IF₂₀₁₃=4,3 MSWiA₂₀₁₃=40 L. cyt.=13)

Mój wkład polegał na opracowaniu koncepcji badań, komputerowej implementacji oraz praktycznym przeprowadzeniu symulacji, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu, odpowiedzi na uwagi recenzentów, końcowym przygotowaniu manuskryptu do druku. Mój udział procentowy szacuję na 80%.

- [H2]. **A. Bak**, J. Polanski, A. Kurczyk, **2009**, The use of MoStBioDat for rapid screening of molecular diversity, *Molecules*, 14(9), 3436-3445. **IF₂₀₀₉=1,7 MSWiA₂₀₀₉=20** (IF₂₀₁₃=2,4 MSWiA₂₀₁₃=30 L. cyt.=5)

Mój wkład polegał na opracowaniu koncepcji badań, komputerowej implementacji oraz praktycznym przeprowadzeniu symulacji, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu, odpowiedzi na uwagi recenzentów, końcowym przygotowaniu manuskryptu do druku. Mój udział procentowy szacuję na 80%.

- [H3]. **A. Bak**, J. Polanski, T. Stockner, A. Kurczyk, **2010**, MoStBioDat - Molecular and Structural Bioinformatics Database, *Comb. Chem. High Throughput Screen.*, 13, 366-374. **IF₂₀₁₀=2,5 MSWiA₂₀₁₀=32** (IF₂₀₁₃=2,0 MSWiA₂₀₁₃=25 L. cyt.=4)

Mój wkład polegał na opracowaniu koncepcji badań, komputerowej implementacji oraz praktycznym przeprowadzeniu symulacji, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu, odpowiedzi na uwagi recenzentów, końcowym przygotowaniu manuskryptu do druku. Mój udział procentowy szacuję na 80%.

- [H4]. P. Mazur, T. Magdziarz, **A. Bak**, Z. Chilmonczyk, T. Kasprzycka-Guttman, I. Misiewicz, K. Skupinska, J. Polanski, **2010**, Does molecular docking reveal alternative chemopreventive mechanism of activation of oxidoreductase by sulforaphane isothiocyanates?, *J. Mol. Model.*, 16, 1205-1212. **IF₂₀₁₀=2,4 MSWiA₂₀₁₀=32** (IF₂₀₁₃=2,0 MSWiA₂₀₁₃=25 L. cyt.=1)

Mój wkład polegał na przygotowaniu oraz praktycznym przeprowadzeniu części obliczeń teoretycznych, wstępnym opracowaniu wyników, ostatecznym przygotowaniu tekstu manuskryptu, odpowiedzi na część uwag recenzentów, końcowym przygotowaniu manuskryptu do druku. Mój udział procentowy szacuję na 40%.

- [H5]. **A. Bak**, T. Magdziarz, A. Kurczyk, J. Polanski, **2011**, Mapping Drug Architecture by MoStBioDat: Rapid Screening of Intramolecular Hydrogen Bonded Motifs in Catechols, *Drug Develop. Res.*, 72, 209-218. **IF₂₀₁₁=1,3 MSWiA₂₀₁₁=20** (IF₂₀₁₃=0.9 MSWiA₂₀₁₃=15 L. cyt.=6)

Mój wkład polegał na opracowaniu koncepcji badań, komputerowej implementacji oraz praktycznym przeprowadzeniu symulacji, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu, odpowiedzi na uwagi recenzentów, końcowym przygotowaniu manuskryptu do druku. Mój udział procentowy szacuję na 90%.

- [H6]. **A. Bak**, T. Magdziarz, A. Kurczyk, J. Polanski, **2011**, Mapping Fragmental Druglikeness in the MoStBioDat Environment: Intramolecular Hydrogen Bonding in Beta-Ketoenols, *Comb. Chem. High Throughput Screen.*, 14(7), 560-569, **IF₂₀₁₁=2,6 MSWiA₂₀₁₁=32** (IF₂₀₁₃=2,0 MSWiA₂₀₁₃=25 L. cyt.=4)

Mój wkład polegał na opracowaniu koncepcji badań, komputerowej implementacji oraz praktycznym przeprowadzeniu symulacji, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu, odpowiedzi na uwagi recenzentów, końcowym przygotowaniu manuskryptu do druku. Mój udział procentowy szacuję na 80%.

- [H7]. K. Serafin, P. Mazur, **A. Bak**, E. Laine, L. Tchertanov, J-F Mouscadet, J. Polanski, **2011**, Ethyl malonate amides: A diketo acid offspring fragment for HIV integrase inhibition, *Bioorg. Med. Chem.*, 19, 5000-5005. **IF₂₀₁₁=2,2 MSWiA₂₀₁₁=32** (IF₂₀₁₃=2,9 MSWiA₂₀₁₃=30 L. cyt.=2)

Mój wkład polegał na przygotowaniu oraz praktycznym przeprowadzeniu części obliczeń teoretycznych, wstępnym opracowaniu wyników, ostatecznym przygotowaniu tekstu manuskryptu, odpowiedzi na część uwag recenzentów, końcowym przygotowaniu manuskryptu do druku. Mój udział procentowy szacuję na 40%.

- [H8]. **A. Bak**, T. Magdziarz, J. Polanski, **2012**, Pharmacophore-Based Database Mining for Probing Fragmental Drug-Likeness of Diketo Acid Analogues, *SAR&QSAR Env. Res.*, 23, 185-204. **IF₂₀₁₂=1,6 MSWiA₂₀₁₂=32** (IF₂₀₁₃=1,7 MSWiA₂₀₁₃=25 L. cyt.=4)

Mój wkład polegał na opracowaniu koncepcji badań, komputerowej implementacji oraz praktycznym przeprowadzeniu symulacji, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu, odpowiedzi na uwagi recenzentów, końcowym przygotowaniu manuskryptu do druku. Mój udział procentowy szacuję na 90%.

- [H9]. J. Polanski, A. Kurczyk, **A. Bak**, R. Musiol, **2012**, Privileged Structures – Dream or reality: Preferential Organization of Azanaphthalene Scaffold, *Curr. Med. Chem.*, 19, 1921-1945. **IF₂₀₁₂=4,6 MSWiA₂₀₁₂=32** (IF₂₀₁₃=4,1 MSWiA₂₀₁₂=40 L. cyt.=6)

Mój wkład polegał na przygotowaniu oraz praktycznym przeprowadzeniu części obliczeń teoretycznych, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu, odpowiedzi na część uwag recenzentów. Mój udział procentowy szacuję na 50%.

- [H10]. **A. Bak**, T. Magdziarz, A. Kurczyk, K. Serafin, J. Polanski, **2013**, 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.*, 16(4), 274-287, **IF₂₀₁₃=2,0 MSWiA₂₀₁₃=25**

Mój wkład polegał na opracowaniu koncepcji badań, komputerowej implementacji oraz praktycznym przeprowadzeniu symulacji, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu, odpowiedzi na uwagi recenzentów, końcowym przygotowaniu manuskryptu do druku. Mój udział procentowy szacuję na 80%.

- [H11]. **A. Bak**, M. Daszykowski, Z. Kaminski, K. Kiec-Kononowicz, K. Kuder, J. Fraczyk, B. Kolesinska, P. Ciosek, J. Polanski, **2014**, Probing an artificial polypeptide receptor library using a series of novel histamine H3 receptor ligands, *Comb. Chem. High Throughput Screen.*, 7(2), 141-156, **IF₂₀₁₃=2,0 MSWiA₂₀₁₃=25**

Mój wkład polegał na opracowaniu koncepcji badań, komputerowej implementacji oraz praktycznym przeprowadzeniu symulacji, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu, odpowiedzi na uwagi recenzentów, końcowym przygotowaniu manuskryptu do druku. Mój udział procentowy szacuję na 60%.

[H12]. **A. Bak**, M. Wyszomirski, T. Magdziarz, A. Smoliński, J. Polanski, **2014**, Structure-based modeling of dye-fiber affinity with SOM-4D-QSAR paradigm: Application to set of anthraquinone derivatives. *Comb. Chem. High Throughput Screen.*, 17(10), 1-18, **IF₂₀₁₃=2,0 MSWiA₂₀₁₃=25**

Mój wkład polegał na opracowaniu koncepcji badań, komputerowej implementacji oraz praktycznym przeprowadzeniu symulacji, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu, odpowiedzi na uwagi recenzentów, końcowym przygotowaniu manuskryptu do druku. Mój udział procentowy szacuję na 80%.

Łączny IF₂₀₀₇₋₂₀₁₃=**28,0** MSWiA₂₀₀₇₋₂₀₁₃=**339** (IF₂₀₁₃=**28,3** MSWiA₂₀₁₃=**330**)

Oświadczenia współautorów dotyczące ich wkładu w powstaniu wspólnych publikacji zostały zamieszczone w załączniku numer 4.

4C. Omówienie celu naukowego w/w prac oraz otrzymanych wyników

Wprowadzenie

Poszukiwanie nowych związków o określonych właściwościach fizykochemicznych i pożądanym profilu aktywności biologicznej jest jednym z podstawowych celów współczesnej chemii medycznej oraz farmacji [1]. Finansowa atrakcyjność rynku farmaceutyków, potocznie zwanych lekami, stwarza dodatkowy bodziec sprzyjający badaniom nad nowymi połączeniami tego typu. Procedurę związaną z otrzymywaniem takich substancji określa się często mianem *projektowania leków*. W miarę rozwoju technologicznego oraz coraz lepszego rozumienia skomplikowanych procesów zachodzących w organizmach żywych dąży się do racjonalizacji procesu poszukiwania nowych leków. Złożoność procesu opracowania nowego leku podkreśla stosowany czasem termin *racjonalne projektowanie leków* lub jego bardziej precyzyjny odpowiednik – *projektowanie molekularne*. Projektowanie molekularne jest obecnie nauką interdyscyplinarną wykorzystującą w coraz większym stopniu matematyczne techniki obliczeniowe oraz procedury informatyczne. Interdyscyplinarne połączenie tradycyjnych metod w chemii oraz technik i zasobów informatycznych doprowadziło do uformowania nowej gałęzi nauki, *chemoinformatyki* – przedmiotu, którego centrum zainteresowania stanowi informatyka cząsteczki chemicznej jako podstawowego obiektu podlegającego manipulacji w projektowaniu molekularnym. W ostatnich latach obserwowany jest szybki

rozwój różnorodnych technik projektowania molekularnego, a konstruowane leki coraz dokładniej odwzorowują substancje wybierane przez Naturę do kontrolowania procesów biologicznych. Brakuje jednak, postulowanej przez Hammonda, efektywnej technologii stanowiącej swoistego rodzaju 'mapę drogową' dla tzw. molekularnej produkcji właściwości [2], a więc cząsteczek o optymalnym profilu farmakodynamicznym i farmakokinetycznym przy jednoczesnej niskiej biotoksyczności. Z drugiej strony wydaje się, że znaczna liczba aktualnie stosowanych w terapii specyfików powinna zaspakajać podstawowe potrzeby w zakresie leczenia różnych schorzeń. Niestety, w wielu przypadkach jest to jedynie leczenie objawowe, które łagodzi dolegliwości, a nie zwalcza przyczyny. Ponadto, w niektórych chorobach stosowane leki są coraz mniej skuteczne lub wywołują niepożądane efekty uboczne. Projektowanie nowych celów syntetycznych oraz przewidywanie ich potencjalnej aktywności biologicznej jest ważne, szczególnie biorąc pod uwagę fakt, iż zaledwie jeden z kilku tysięcy nowosyntezowanych związków ma szansę pomyślnego przejścia przez cykl testów farmakologicznych oraz klinicznych. W chwili obecnej wyróżnia się dwie podstawowe strategie projektowania związków aktywnych. Pierwsza opiera się na poszukiwaniu niskocząsteczkowego bioefektora jako komplementarnego dopełnienia receptora (RD), druga zaś ogranicza się do badania analogii w zbiorze aktywnych ligandów z wytworzeniem przestrzennej mapy oddziaływań zwanej *farmakoforem* (RI) [3]. Informacje zebrane w trakcie symulacji komputerowych pozwalają ograniczyć kosztochłonność procesu poprzez ograniczenie liczby eksperymentów laboratoryjnych.

Cel rozprawy habilitacyjnej

Celem projektu zatytułowanego „Molekularna i Strukturalna Bioinformatyczna Baza Danych – MoStBioDat” było opracowanie, utworzenie oraz praktyczne wykorzystanie programowej platformy zarządzającej do łatwego przechowywania, przetwarzania oraz przesyłania (3xp) danych strukturalnych o substancjach wykazujących potencjalną aktywność biologiczną oraz o makromolekułach białkowych wraz z możliwymi adnotacjami o ich profilu aktywności farmakologicznej. Realizacja poszczególnych etapów projektu wymagała nie tylko fizycznego dostępu do danych strukturalnych, przechowywanych w zaprojektowanym systemie bazodanowym, lecz przede wszystkim zbioru narzędzi programistycznych do wydajnej analizy danych masowych. Koncepcyjnie, baza MoStBioDat składa się z dwóch głównych składników: warstwy przechowywania z relacyjnym systemem tabel do fizycznego przechowywania danych o geometrii ligandów i makromolekuł, oraz warstwy zarządzającej z zestawem procedur do efektywnego pobierania, składowania oraz korekcji danych. Warstwa użytkownika zaimplementowana została w jednym z języków wysokiego poziomu wspierającym programowanie obiektowo zorientowane, Pythonie. Jedną z zalet pakietu jest relatywnie prosta procedura instalacji. Modułarna architektura środowiska

MoStBioDat umożliwiła wygodne zwiększenie funkcjonalności pakietu poprzez implementację dodatkowych procedur.

Niniejsza rozprawa stanowi kompendium, opisujące etapy realizacji projektu, który podzielony został na 3 zasadnicze zadania obejmujące:

- zaprojektowanie oraz komputerową implementację bazy danych przechowującej informacje o geometrii i właściwościach fizykochemicznych, profilu biologicznym ligandów oraz celów makromolekularnych wraz ze zbiorem modułów do obsługi bazy i analizy przechowywanych danych;
- przeprowadzenie badań przesiewowych repozytoriów zawierających dane strukturalne przy pomocy filtrów wybranych właściwości fizykochemicznych i biologicznych oraz analiz podobieństwa cząsteczkowego dla wybranych układów molekularnych
- modelowanie mechanizmu rozpoznawania molekularnego cząsteczek do biblioteki sztucznych receptorów oraz zastosowano neuronowy formalizm analiz 4D-QSAR, który uwzględnia struktury receptorowe w odniesieniu do modelowania oddziaływań ligand-receptor.

Podkreślić należy, iż wykonanie badań i symulacji było możliwe jedynie dzięki opracowaniu, zaimplementowaniu oraz praktycznym wykorzystaniu odpowiedniego oprogramowania, w szczególności środowiska programistycznego MoStBioDat. **Zasadniczym etapem projektu było zaprojektowanie, komputerowa implementacja oraz testowanie zarówno bazy danych, jak i zintegrowanego w pakiecie MoStBioDat oprogramowania.** Cały pakiet, który składa się z zestawu modułów wykonujących określone zadania, dostarcza narzędzi do obsługi i analizy informacji zgromadzonej w bazie danych. Uwzględniając specyfikę niektórych symulacji, w części analiz rozszerzone zostały pierwotne możliwości pakietu bazowego poprzez implementację dodatkowych metod do m.in. filtrowania oraz analizy kątów torsyjnych i obliczania wybranych deskryptorów molekularnych. Metody odkrywania wiedzy oparte na statystycznych analizach danych masowych umożliwiają badanie trendów, preferencji i dostępności syntetycznej w chemii organicznej na przestrzeni wybranych okresów czasowych.

Otrzymane wyniki zostały przedstawione w formie 12 publikacji naukowych w czasopismach o zasięgu międzynarodowym oraz na wybranych konferencjach tematycznych.

Ogólny zarys problematyki badań z uwzględnieniem prac własnych

Symulacje komputerowe odgrywają coraz większą rolę we współczesnych analizach chemicznych generując jednocześnie znaczną liczbę danych. Wymusza to konieczność ciągłego poszukiwania coraz to wydajniejszych metod ich przetwarzania, kompresji i wizualizacji. Jak każda nauka także i chemia jest nieprzebranym zbiorem informacji, którego analiza, a w szczególności poprawna interpretacja, stanowi obecnie poważne wyzwanie metodologiczne. Utrudniony przepływ informacji oraz danych jest jednym z głównych czynników hamujących rozwój różnych dyscyplin naukowych – nie tylko chemii. Dostęp do informacji jest w chemii problemem podstawowym, który warunkuje możliwość prowadzenia badań m.in. syntez organicznych. Gromadzenie, przetwarzanie oraz wymiana informacji o obiektach przestrzeni chemicznej wymaga wydajnych systemów bazodanowych. System taki pozwala na przeglądanie oraz weryfikację strukturalnych, fizykochemicznych i biologicznych właściwości związków chemicznych. Innowacyjne zastosowanie komputera wyposażonego w bazy danych chemicznych spowodowało swoistą rewolucję w efektywności przechowywania i przeszukiwania rekordów chemicznych. Od wielu lat prowadzone są prace nad ujednoliceniem standardów wymiany informacji o strukturze związków chemicznych pomiędzy różnymi systemami obliczeniowymi. Należy tak reprezentować cząsteczkę, aby w dalszych analizach tworzyła zintegrowaną strukturalnie całość. Intensywny rozwój technologii informatycznych, w tym systemów bazodanowych, daje nowe możliwości jednoznacznego kodowania obiektów chemicznych oraz zarządzania informacją masową. Stale tworzone są nowe, dostępne *on-line* repozytoria, w których przechowywane są dane strukturalne oraz właściwości fizykochemiczne opisujące małe cząsteczki (ligandy) oraz makromolekuły (białka, kwasy nukleinowe). Niestety istnieje zaledwie kilka ogólnie dostępnych systemów, które łącząc ze sobą dane strukturalne ligandów z ich celami biologicznymi (receptorami) dostarczają równocześnie narzędzi do badania różnych aspektów podobieństwa cząsteczek.

Eksploracja danych sprzężona z technikami odkrywania wiedzy zawartej w bazach danych jest obecnie ważnym kierunkiem badań, który umożliwia pozyskanie wiedzy przy wykorzystaniu szczególnych możliwości informatyki do szybkiego wyszukiwania i analizy informacji. Pozwala to na ujawnienie relacji, które są szczególnie trudne do intuicyjnego odkrycia przez człowieka. Należy jednak pamiętać, iż baza danych jest narzędziem bezwartościowym, gdy brakuje efektywnych i szybkich metod jej przeszukiwania. Konieczność wydajnego zarządzania informacją chemiczną była podstawą projektu, którego celem nadrzędnym było dostarczenie narzędzi do przechowywania oraz eksploracji znacznych zasobów bazodanowych.

Utworzony system MoStBioDat przechowuje informacje w ujednolicony sposób, co redukuje tzw. dane nadmiarowe. Relatywnie prosty schemat hierarchicznego ułożenia danych na poziomie bazy danych stwarza przyjazne dla użytkownika środowisko pracy. Programowe metody korekcji błędów w plikach formatu SDF oraz PDB zapewniają integralność i spójność danych. Warstwa programowa zaimplementowana w języku Python dostarcza metod do składowania i pobierania danych oraz umożliwia przeszukiwanie baz danych (skrining) z równoczesną analizą podobieństwa międzymolekularnego. Szczegółowy opis architektury bazy danych wraz z modułami programowymi udostępniony został na stronie domowej projektu MoStBioDat oraz omówiony w cyklu publikacji [H2, H3].

Przeszukiwanie baz danych, oparte o podstruktury molekularne, to proces identyfikacji cząsteczek, które zawierają określone motywy strukturalne definiowane w zapytaniu. Filtrowanie może odbywać się na podstawie określonych deskryptorów strukturalnych lub właściwości molekuł, które są przechowywane w bazie lub obliczane na podstawie struktury cząsteczki. Połączenie informacji strukturalnej, która przechowywana jest w bazie danych z wydajnymi systemami przeszukiwań umożliwia przeprowadzenie wysokowydajnych badań przesiewowych. Dane tego typu mogą być źródłem informacji niezbędnej do przeprowadzenia analiz między strukturą związków a ich aktywnością farmakologiczną [H1]. Należy podkreślić, iż obecne analizy QSAR polegają bardziej na wizualizacji pewnych trendów zmienności aktywności niż na precyzyjnym prognozowaniu aktywności pojedynczo zaprojektowanej cząsteczki. Technika ta umożliwia, co prawda ograniczone do pewnej podprzestrzeni, przewidywanie struktur nowych związków o danym kierunku działania. Historycznie uwarunkowanym schematem poszukiwania aktywnego biologicznie związku jest modyfikowanie struktury wiodącej, czyli znanego związku wykazującego interesującą aktywność biologiczną. Optymalizację struktury wiodącej można przeprowadzić syntezując podobne do niej związki i analizując wpływ budowy na aktywność. Badanie wpływu zmian struktury na aktywność jest więc w dużej mierze uzależnione od możliwości syntetycznych modyfikacji struktury wiodącej. Modyfikacje takie w pierwszej fazie optymalizacji dokonywane są często w oparciu o bardziej lub mniej intuicyjne reguły wyboru analogów struktury wiodącej - nie bez znaczenia jest w tym przypadku ich syntetyczna dostępność. Z drugiej strony istotne jest uzyskanie możliwie pełnej charakterystyki zmienności (różnorodności) strukturalnej takich analogów. Koncepcja ilościowego opisu stopnia zróżnicowania cząsteczek chemicznych jest stosunkowo nowa. Poszukiwanie związków biologicznie aktywnych sprowadza się bowiem do poszukiwania podobieństwa do cząsteczek działających w naturalnych systemach biologicznych. Ostatnio często dyskutowane jest występowanie tzw. motywów uprzywilejowanych, które zawarte są w wielu lekach oraz produktach pochodzenia naturalnego [H9]. Problemem jest jednak ustalenie ścisłej koncepcji lekotypii lub uprzywilejowanych wzorców leków, czyli zdefiniowanie reguł

pozwalających odróżnić cząsteczkę leku od innych cząsteczek chemicznych. Lekotypia zakłada, iż kategoria leków spełnia pewne reguły empiryczne związane z deskryptorami lub właściwościami fizykochemicznymi. Zasady te mogą wyznaczać wstępne kierunki poszukiwania leków. Implikuje to poszukiwanie podobieństw, na przykład podobnych ugrupowań uprzywilejowanych wzorców leków lub podobieństwa między cząsteczkami leku a kieszenią receptorową. Ponadto znacznie łatwiej zdefiniować kategorię podobieństwa niż różnorodności. Jednakże odkrywanie nowych leków wymaga także eksploracji różnych obszarów przestrzeni chemicznej. Konieczna więc staje się konstrukcja molekuł wypełniających możliwie równomiernie takie obszary. Celem działania leku w układzie biologicznym jest makromolekuła określana mianem receptora. W projektowaniu leków receptorem mogą być białka receptorowe, enzymy oraz inne układy molekularne (kwasy nukleinowe). Pełny opis mechanizmu działania leku wymaga więc wyjaśnienia oddziaływań w układzie lek-receptor. Metody projektowania w oparciu o dane strukturalne receptora stanowią cel, do którego dąży się w dzisiejszych technologiach komputerowego wspomagania projektowania molekularnego (CAMD). Bezpośrednie włączenie do opisu oddziaływań ligand-receptor geometrii miejsca wiążącego powinno pozwolić na zrozumienie procesów rozpoznawania molekularnego. Przestrzenne rozmieszczenia poszczególnych atomów kompleksu ligand-receptor, które uwzględnia efekty steryczne oraz elektrostatyczne definiuje technika nazywana dokowaniem [H4]. Jedną z technik, która umożliwia wzajemne geometryczne dopasowanie konformacyjnie labilnego układu ligand-receptor jest procedura dynamiki molekularnej (MD). Często spotykanym problemem w projektowaniu leków jest określenie kształtu i konformacji cząsteczki w chwili jej dopasowania do miejsca wiążącego tzw. konformacji bioaktywnej. Jest to szczególnie ważne dla cząsteczek o znacznej labilności konformacyjnej wynikającej ze znacznej liczby wiązań rotacyjnych. Jednym ze sposobów poszukiwania konformacji aktywnych jest analizowanie wszystkich sensownych konformerów szeregu związków aktywnych a następnie wyznaczenie wspólnej przestrzeni grup oddziałujących z celem makromolekularnym. W ostatnim czasie klasyczne metody RI wielowymiarowej techniki QSAR, bazujące na analizie deskryptorów molekularnych, uzupełniane są o dane opisujące układ receptora rozpoznający szereg ligandów. Wciąż brakuje jednak dostatecznie wiarygodnych procedur RD, które umożliwiłyby praktyczne projektowanie nowych celów syntetycznych. Badania stosunkowo prostych bibliotek receptorów peptydowych umożliwiają przeprowadzenie prac modelowych zmierzających do opracowania efektywniejszych metod RD-QSAR [H11].

Oddziaływania cząsteczek z receptorem biologicznym ograniczają się z reguły do kilku wybranych obszarów – do opisu tego typu oddziaływań potrzebna jest niewielka liczba zmiennych opisujących położenia poszczególnych atomów. Zastosowanie jednej z technik wyboru zmiennych eliminuje dane, które niejednokrotnie stanowią szum informacyjny [4].

W toku realizacji badań zaadaptowano opracowane wcześniej procedury RI QSAR przez włączenie do opisu geometrii receptora [H12].

Pomimo stale rosnącej liczby metod obliczeniowych, wciąż brakuje skutecznych metod terapii przeciwko wielu schorzeniom. Poważne wyzwanie nadal stanowi AIDS, mimo iż udowodniono wirusową etiologię tej choroby. W chwili obecnej wieloskładnikowa farmakoterapia przeciwko AIDS (HAART) opiera się głównie na lekach hamujących odwrotną transkryptazę (RT) i proteazę (P), choć duże nadzieje wiąże się z zastosowaniem nowej klasy połączeń, inhibitorów integrazy (IN) wirusa HIV - związków, które blokują proces wprowadzenia materiału genetycznego wirusa do genomu zainfekowanej komórki gospodarza [5]. Dotychczas wprowadzono na rynek dwa specyfiki, których działanie polega na hamowaniu integrazy wirusa HIV. Opracowanie oraz synteza efektywnych i bezpiecznych preparatów tego typu umożliwia prowadzenie tzw. politerapii przeciwko AIDS. Pomimo intensywnych prac badawczych nadal nie rozstrzygnięty jednoznacznie pozostaje problem mechanizmu hamowania replikacji wirusa HIV. Dodatkowe utrudnienie stanowi duża zmienność genetyczna, która prowadzi do licznych mutacji wirusa HIV. Postuluje się, iż znaczną rolę odgrywać mogą konkurencyjne miejsca wiązania inhibitora z DNA wirusa. W literaturze opisano wiele grup związków organicznych hamujących działanie integrazy wirusa HIV m.in. styrylochinoliny lub diketokwasy [6]. Przeprowadzone badania koncentrowały się wokół zagadnienia wiązań wodorowych występujących w związkach zawierających tego typu motywy strukturalne. Zbadano częstość występowania konformacji sprzyjających tworzeniu wewnątrzcząsteczkowych wiązań wodorowych dla dwóch grup związków: pochodnych katecholiny oraz β -ketoenoli [H5, H6]. Potencjalne tworzenie wewnątrzcząsteczkowych wiązań wodorowych jest istotne z punktu widzenia możliwości oddziaływania analizowanych cząsteczek z makromolekułami. Odpowiednia liczba donorów i akceptorów wiązań wodorowych znajdujących się w cząsteczce jest powiązana z jej aktywnością biologiczną. Obliczenia przeprowadzono dla zasobów kilku baz molekularnych, w tym bazy PubChem, ZINC, DrugBank i Ligand.Info. Uzyskane wyniki zostały porównane z danymi otrzymanymi przez innych badaczy dla zasobów bazy Cambridge Structural Database. Przeprowadzone analizy umożliwiły wskazanie baz, które są źródłem wiarygodnych danych strukturalnych. Opracowano również kryteria analizy i wyszukiwania konformacji sprzyjających tworzeniu wiązań wodorowych w powiązaniu z potencjałem bioaktywnym analizowanych cząsteczek.

Wykorzystano procedury filtrowania związków oparte o kryteria geometryczne lub analizę wybranych deskryptorów molekularnych przy użyciu techniki sieci neuronowych SOM, analizy czynników głównych PCA oraz dokowania molekularnego [H7]. Miarą podobieństwa analizowanych podprzestrzeni chemicznych był współczynnik Tanimoto, zaś szeroko rozumiana analiza podobieństwa dostarczyła cennych informacji o znaczeniu

poszczególnych farmakoforów. Zastosowanie metod projekcyjnych pozwoliło na wizualizację oraz analizę rozkładu wybranych właściwości molekularnych w badanych podzbiorach molekuł, uwidocznienie różnic i podobieństw pomiędzy szeregami związków [H10]. Przeprowadzono również badania przesiewowe bazujące na koncepcji przestrzennej farmakofora, w szczególności analizę (współ)występowania motywu diketokwasu [H8]. Wybrano zbiór molekuł docelowych (szereg bioizosterycznych pochodnych DKA), w szczególności monohalogenowanych pochodnych arylowych z podstawnikiem w pozycji *para*, które spełniają określone kryteria geometrii oraz podobieństwa do związków odniesienia raltegraviru, elvitegraviru – inhibitorów integrazy wirusa HIV. Identyfikacja tych klas oraz możliwość ich wyodrębnienia z bazy danych daje możliwość dalszych badań zmierzających do identyfikacji nowych potencjalnych inhibitorów integrazy HIV.

Omówienie wyników

1. Modelowanie RI 4D-QSAR

Modelowanie efektów chemicznych i biologicznych stanowi poważne wyzwanie stawiane współczesnej chemii i farmakologii. Od dłuższego czasu obserwuje się stały rozwój technik badania zależności między aktywnością biologiczną a strukturą chemiczną (QSAR). Technika QSAR pozwala badać ilościowe zależności pomiędzy aktywnością biologiczną a właściwościami fizykochemicznymi lub parametrami opisującymi strukturę związków chemicznych. Klasyczne modele QSAR (tzw. 2D-QSAR), ograniczając się jedynie do dwuwymiarowej reprezentacji cząsteczek, nie są w stanie wyjaśnić różnic wynikających z ich przestrzennej budowy. Tylko uwzględnienie trójwymiarowej struktury związków chemicznych daje możliwość pełnego opisu wpływu efektów strukturalnych na ich aktywność biologiczną. W końcu lat 80-tych ubiegłego wieku, opierając się na nowych technikach statystycznych (PCA, PLS), opracowano teoretyczne podstawy porównawczej analizy pola cząsteczkowego (CoMFA). Jedną z alternatywnych metod modelowania zależności 3D-QSAR polega na zastosowaniu samoorganizującej się sieci neuronowej Kohonena (SOM). Stwierdzono, że sieci tego typu można m.in. wykorzystać do tworzenia dwuwymiarowych map potencjału elektrostatycznego cząsteczek chemicznych [7]. Analiza szeregu takich map z zastosowaniem metody PLS umożliwia modelowanie 3D-QSAR, zaś technikę taką nazwano porównawczą analizą powierzchni cząsteczkowej (CoMSA) [8,9].

Ważnym problemem związanym z modelowaniem 3D-QSAR jest analiza konformacyjna. Analiza taka jest jedną z technik badania farmakofora w przypadku, gdy nie jest znana struktura receptora (RI). Ciekawą techniką tego typu może być, zaproponowana w 1997 roku przez Hopfingera metoda 4D-QSAR, która w porównaniu do 3D-QSAR, jako dodatkowy wymiar wprowadza zmiany konformacyjne cząsteczek [10]. Metody mapowania ugrupowań

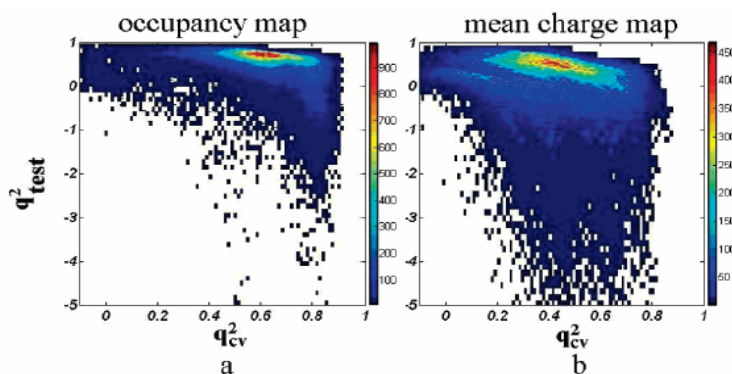
farmakoforowych postulują istnienie przestrzennych ugrupowań strukturalnych, które musi posiadać związek chemiczny, aby był rozpoznawalny przez receptor. Konceptyjnie 4D-QSAR zakłada, że cząsteczki o podobnych cechach strukturalnych i właściwościach fizykochemicznych charakteryzują się podobnym mechanizmem tworzenia kompleksu lek-receptor, a w konsekwencji porównywalnym profilem aktywności biologicznej.

Metodologia 4D-QSAR pozwala konstruować przestrzenne modele QSAR opisujące podobieństwo strukturalne szeregu związków chemicznych w funkcji zmian konformacyjnych, sposobu nałożenia i rodzaju grup farmakoforowych. Algorytm kodowania cząsteczek w metodzie 4D-QSAR opiera się na dyskretyzacji przestrzeni przez układ jednostkowych komórek przestrzennej sieci sześciątów. Podobną dyskretyzację można uzyskać transformując przestrzeń z zastosowaniem sieci neuronowych Kohonena (SOM-4D-QSAR). Zrozumienie wszystkich zależności pomiędzy strukturą przestrzenną i elektronową różnorodnych grup związków chemicznych a ich aktywnością biologiczną pozwoli na skuteczne projektowanie leków oraz pełną interpretację ogółu ich oddziaływań z receptorami w układach biologicznych. Układy te są skomplikowane, co utrudnia ich matematyczne modelowanie. Z reguły jednak oddziaływanie cząsteczek z receptorami ogranicza się do kilku wybranych obszarów charakteryzowanych przez położenia poszczególnych atomów. W większości analiz QSAR występuje duża liczba zmiennych (deskryptorów) opisujących strukturę badanych związków. Umiejętna eliminacja nieistotnych zmiennych pozwala nie tylko ograniczyć analizy do najważniejszych obszarów oddziaływań molekularnych, ale również poprawić efektywność modelowania. Jedną z metod eliminacji zmiennych jest technika UVE-PLS. Klasyczny algorytm UVE-PLS charakteryzuje znaczna szybkość obliczeniowa, nie daje jednak kontroli nad liczbą eliminowanych zmiennych [11]. Rozwiązaniem problemu może być modyfikacja powyższej procedury tzw. mUVE-PLS lub metoda IVE-PLS [12].

Wzajemne oddziaływanie pomiędzy cząsteczką barwnika a celulozą jest złożonym zjawiskiem opisywanym izotermą Langmuira [13]. Znaczny stopień skomplikowania oddziaływań powoduje, że opis ten nie wystarcza do molekularnego modelowania omawianych procesów. Różnorodność procesów zachodzących podczas wiązania się molekuł barwnika z włóknem celulozowym (m.in. oddziaływania elektrostatyczne, van der Waalsa, wodorowe itd.) sugeruje, iż mechanizm ten może być podobny do sposobu oddziaływania w układzie ligand-receptor. Dodatkowo, rozmieszczenie barwnika na powierzchni celulozy wskazuje, iż istnieją określone miejsca wiązania barwnika z włóknem [14]. Pomimo, iż brakuje jednoznacznych dowodów potwierdzających receptorowy charakter oddziaływań, w literaturze opisano próby modelowania powinowactwa barwników do włókna celulozowego techniką 2D i 3D-QSAR [15]. Aby porównać wyniki modelowania 4D-QSAR tego typu związków z danymi literaturowymi przeprowadzono symulacje dla szeregów

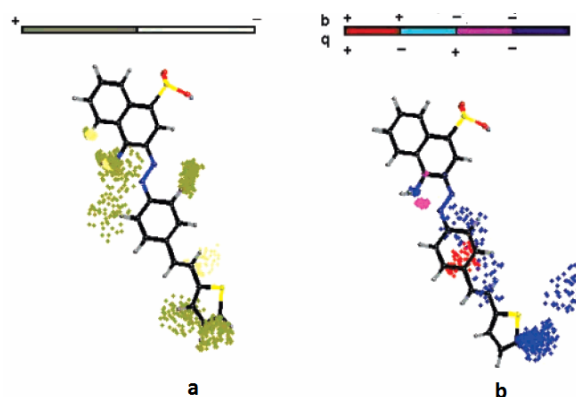
barwników azowych i antrachinonowych [16]. W celu wyjaśnienia skomplikowanych oddziaływań grupy barwników z włóknem celulozowym wprowadzono pojęcie *tinktofora*, swoistego odpowiednika *farmakofora* w modelowaniu QSAR [17]. Wyniki przestrzennego mapowania oddziaływań barwnik-włókno sugerują, iż na powierzchni celulozy istnieją struktury o charakterze receptorowym [18].

W wielu analizach QSAR walidacja modelu ogranicza się jedynie do obliczenia pojedynczych wartości parametrów dopasowania dla jednego, wybranego zbioru treningowego i testowego. Zamiana jednego obiektu w tych zbiorach prowadzić może do zasadniczych różnic w charakterystyce modelowania. Opierając się na metodzie krosvalidacji LOO do modelowania oddziaływań barwnik-celuloza zastosowano technikę sprzężonej walidacji LSO [19]. Metoda ta polega na równoczesnym testowaniu wartości estymatora dopasowania dla obiektów zbioru treningowego oraz wartości estymatora przewidywania obliczonego dla obiektów zbioru testowego. Iteracyjna procedura próbkowania różnych kombinacji molekuł z wybranej podprzestrzeni chemicznej może być traktowana jako metoda o charakterze probabilistycznym [20]. Stochastyczny algorytm oceny modeli (SMV) umożliwia wskazanie kombinacji cząsteczek w zbiorach treningowych i testowych, dla których utworzony model charakteryzuje się odpowiednio wysokimi parametrami dopasowania oraz przewidywania.



Rys. 1. Mapa częstości występowania parametrów oceny dopasowania i predykcji dla podziału 20/10 zbiór treningowy/testowy szeregu barwników azowych modelowanych procedurą SOM-4D-QSAR (a) deskryptor występowania (b) deskryptor uśrednionego ładunku.

Przypadkowo wybrane 100 modeli z obszarów o największej częstości występowania poddane zostały procedurze eliminacji zmiennych nieistotnych metodą IVE-PLS, wskazując obszary o największym znaczeniu dla powinowactwa do włókna celulozowego tej grupy związków. Technika ta pozwala na wizualizację niektórych aspektów podobieństwa cząsteczkowego, szczególnie w przypadku, gdy oddziaływania limitowane są kształtem molekuł. Po eliminacji identyfikowane są fragmenty przestrzeni, które potencjalnie determinują aktywność szeregu związków.



Rys. 2. Obszary istotne dla modelowania SOM-4D-QSAR szeregu barwników azowych wybrane metodą IVE-PLS dla (a) deskryptorów występowania (b) deskryptorów uśrednionego ładunku. Kolory kodują znak przy danej zmiennej oraz wartość ładunku.

Wydaje się, iż większa część molekuly zaangażowana jest w interakcje z celulozą, co w pewnym sensie odróżnia *tinktofor* od *farmakofora* wskazując jednocześnie na mniejszą specyficzność oddziaływań w układzie barwnik-włókno.

Szczegółowa analiza problemów związanych z modelowaniem QSAR/QSPR, przedstawiona w serii prac po tytule 'Modeling Robust QSAR' [21,22] wykazała, iż metody stochastyczne mogą odgrywać znaczącą rolę w procesie deterministycznego modelowania ilościowej zależności struktura-aktywność, wzbogacając narzędzia badawcze oraz prowadząc do lepszego zrozumienia tego typu relacji [H1]. Mnogość danych generowanych w tego typu analizach wskazała na potrzebę utworzenia spójnego środowiska programistycznego, złożonego z systemu przechowywania, zarządzania oraz pakietu narzędzi do eksploracji zebranych informacji. W ramach współpracy z Austrian Institute of Technology w Seibersdorfie podjęta została próba opracowania swoistego rodzaju platformy do analiz o charakterze deterministycznym oraz probabilistycznym – Molekularnej i Strukturalnej Bioinformatycznej Bazy Danych (MoStBioDat).

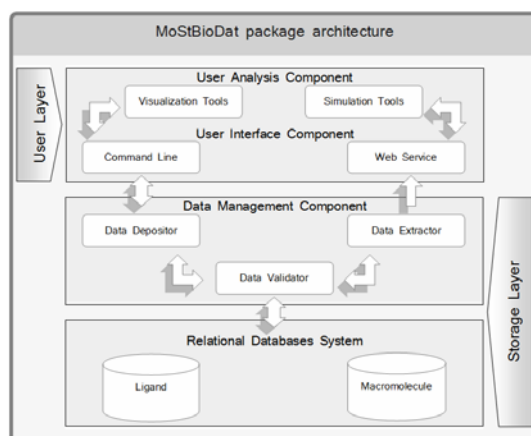
2. Architektura platformy MoStBioDat

Procedury przeszukiwań bibliotek związków chemicznych oparte jedynie na podobieństwie strukturalnym do związku referencyjnego są obecnie zastępowane technikami eksploracji danych i pozyskiwania wiedzy z chemoinformatycznych baz danych [23]. Łatwy dostęp do repozytoriów danych chemicznych wymusił rozwój metod zarządzania informacją masową. Baza danych to połączenie wiedzy i technologii, które zintegrowane jest w wyspecjalizowanym oprogramowaniu nazywanym systemem zarządzania bazami danych (DBMS) lub potocznie systemem baz danych. DBMS stanowi narzędzie przeznaczone do tworzenia oraz wydajnego zarządzania dużymi zbiorami danych oraz bezpiecznego ich przechowywania [24]. Tak więc, system bazy danych składa się z bazy danych, systemu

zarządzania bazą danych oraz opcjonalnego zestawu programów wspomagających użytkowników.

Koncepcyjnie baza MoStBioDat składa się z dwóch głównych składników: warstwy przechowywania oraz warstwy użytkownika. Warstwa przechowywania odpowiada za fizyczne składowanie danych na dyskach oraz dostarcza metod do pobierania danych. Przy projektowaniu warstwy przechowywania zastosowano model relacyjny, co w znacznej mierze zapobiega replikacji danych. Dane zorganizowano w tabele z kluczami podstawowymi (PR), które stanowią jednoznaczne identyfikatory poszczególnych rekordów. Tabele powiązane są ze sobą kluczami obcymi (FK), co zapewnia spójność oraz integralność przechowywanych danych. Baza danych MoStBioDat oparta jest na serwerze MySQL z silnikiem InnoDB, wspierającym model relacyjny. Relacyjny system baz danych składa się z dwóch składników: Ligand i Macromolecule. Baza danych Ligand zawiera zbiór tablic podstawowych zawierających różnorodne reprezentacje molekuł włączając reprezentacje jednowymiarowe cząsteczek z kodem SMILES. Do każdej tablicy podstawowej zdefiniowano zbiór tablic wspierających, które przechowują numeryczne i alfanumeryczne wartości właściwości fizykochemicznych i biologicznych. Główna koncepcja biomolekuł zakłada że składają się one z odpowiednio zdefiniowanych bloków. Topologia takiego systemu przechowywana jest w tablicach, które odzwierciedlają budowę plików w formacie PDB. System baz został zintegrowany i tworzy spójne środowisko do zarządzania danymi o małych cząsteczkach i ich celach makromolekularnych.

Składnik zarządzania danymi składa się z zestawu programów do obsługi danych – pobieranie/składowanie/zmiana. Warstwa programowa zaimplementowana została w Pythonie, który jest darmowym, niezależnym od systemu operacyjnego językiem wysokiego poziomu, wspierającym programowanie obiektowo zorientowane.



Rys. 3. Architektura bazy danych MoStBioDat.

Cały pakiet zawierający ponad 20 tys. linii kodu składa się z zestawu modułów wykonujących określone zadania, dostarczających narzędzi do importu, logowania i semi-automatycznych

operacji na bazie danych. Jednym z głównych założeń początkowych było utworzenie oprogramowania z łatwą i szybką metodą instalacji, dlatego też pakiet oferuje prostą procedurę instalacji. W chwili obecnej lokalna baza danych MoStBioDat zawiera ponad 12 milionów rekordów związków chemicznych wraz z danymi o ich właściwościach fizycznych, chemicznych i biologicznych. Szczegółowy opis bazy danych MoStBioDat wraz z dokumentacją oraz oprogramowaniem dostarczanym na warunkach licencji GPL 3 można znaleźć na stronie domowej projektu <http://chemoinformatyka.us.edu.pl/~zchorg/mostbiodat/>. Omówienie poszczególnych modułów pakietu oraz ich przykładowe zastosowanie zostało dokładnie przedstawione w publikacji [H3].

3. Analiza podobieństwa cząsteczkowego przy użyciu środowiska MoStBioDat

Komputer wyposażony w informację chemiczną wraz z oprogramowaniem zdolnym do przetwarzania i zarządzania danymi chemicznymi tworzy platformę wspomagającą chemika w badaniach. Porównywanie struktur cząsteczek jest jedną z podstawowych technik wykorzystywanych w różnych metodach modelowania molekularnego (MM). Zakłada się, iż molekuly wykazujące podobieństwo strukturalne powinny charakteryzować się podobnym profilem farmakologicznym [25]. Definiowanie oraz pomiar podobieństwa strukturalnego całych cząsteczek lub ich fragmentów (podstruktur) zależy od sposobu reprezentacji molekuł. Badania przesiewowe baz danych są interesującą strategią będącą swoistego rodzaju połączeniem projektowania racjonalnego z masowym podejściem do poszukiwania nowych bioefektorów. Przeszukiwanie znacznych zasobów baz danych z uwzględnieniem różnych aspektów podobieństwa molekularnego zależy od przyjętego sposobu kanonicznej reprezentacji cząsteczki. Chemia kombinatoryczna w połączeniu z wysokowydajnymi metodami badania przesiewowego (HTS) może być skutecznym narzędziem służącym wydajnej syntezie i efektywnej identyfikacji cząsteczek o potencjalnym zastosowaniu farmakologicznym. Analizy podobieństwa oparte mogą być na badaniu analogii w zbiorze aktywnych ligandów z utworzeniem przestrzennej mapy oddziaływań (farmakofora). Farmakofor traktowany może być jako swoistego rodzaju negatyw geometrii receptora, który można wykorzystać do poszukiwań nowych związków aktywnych względem danego receptora. Zastosowanie metryk podobieństwa umożliwia numeryczne określenie podobieństwa w zbiorze ligandów. Jedną z często stosowanych miar podobieństwa jest współczynnik Tanimoto.

Przeszukiwanie baz danych oparte o podstruktury molekularne to proces identyfikacji cząsteczek, które zawierają określone motywy strukturalne definiowane w zapytaniu. Przeszukiwanie baz danych, tzw. wirtualny skrining oparty o (pod)strukturę lub właściwości fizykochemiczne związków, stał się standardową procedurą we wstępnych etapach racjonalnego projektowania substancji o potencjalnej aktywności biologicznej. Najprostsza

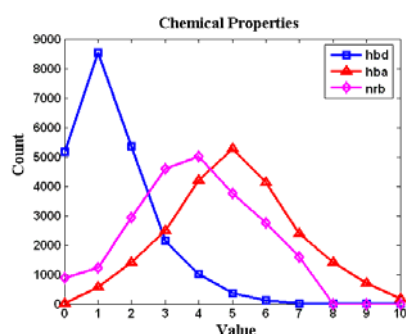
odmiana skringingu bazy danych polega na porównaniu wzorów sumarycznych cząsteczek. Molekuły, które nie zawierają określonej sekwencji atomów są automatycznie odrzucane. Komputer, operując na danych zapisanych w systemie binarnym, nie rozumie tak abstrakcyjnych typów danych jak molekuła czy wiązanie chemiczne. Z punktu widzenia systemu komputerowego dane te muszą zostać zakodowane w postaci binarnej tzn. ciągu kolejnych zer i jedynek. W notacji liniowej struktura związku chemicznego reprezentowana jest w postaci ciągu znaków alfanumerycznych, czyli liniowej sekwencji liter i cyfr. Istnieją algorytmy, które generują tzw. kanoniczną reprezentację cząsteczek w notacji liniowej (np. algorytm Morgana). Oznacza to, iż w sposób jednoznaczny i unikatowy można utworzyć sekwencję znaków kodujących strukturę cząsteczki. Taki format danych jest szczególnie przydatny w systemach bazodanowych, gdzie znaczna liczba molekuł może być przechowywana, przesyłana oraz wydajnie zarządzana. SMILES stał się jednym z najbardziej rozpowszechnionych systemów przetwarzania i wymiany informacji strukturalnej w systemach komputerowych. Jest to kompletny język kodowania topologii układu w notacji liniowej oparty na teorii wiązań walencyjnych. Stanowi konstrukcję językową z prostym słownictwem i kilkoma regułami gramatycznymi, szczególnie przydatną przy przeszukiwaniu baz danych. Filtrowanie może odbywać się na podstawie określonych deskryptorów lub właściwości molekuł, które są przechowywane w bazie lub obliczane na podstawie struktury cząsteczki. Wstępnym etapem procedury przesiewowej jest odrzucenie związków o niepożądanych parametrach farmakologicznych lub posiadających niekorzystne dla aktywności elementy strukturalne. W koncepcji lekotypii zakłada się, iż istnieje zespół cech strukturalnych i fizykochemicznych charakterystycznych dla danej populacji leków. Jednym z często stosowanych kryteriów filtrowania jest reguła 5 Lipińskiego (Ro5). Zasada ta podaje wartości właściwości molekularnych, które są istotne dla farmakokinetyki potencjalnego leku w żywym organizmie, włączając absorpcję, dystrybucję, metabolizm, wydalenie oraz toksyczność leku (ADMET) [26]. Przestrzeń właściwości definiowana jest przez współczynnik podziału pomiędzy dwie niemieszające się fazy (woda/n-oktanol) Clog P, masę molową MW, liczbę donorów wiązania wodorowego (liczba atomów N lub O połączonych z jednym lub większą liczbą atomów H) HDO oraz liczbę akceptorów wiązania wodorowego (liczba atomów N lub O) HAC. W przypadku, gdy cząsteczka nie znajduje się w podprzestrzeni właściwości ograniczonej przez

$$MW \leq 500, \text{Clog P} \leq 5, \text{HDO} \leq 5, \text{HAC} \leq 10$$

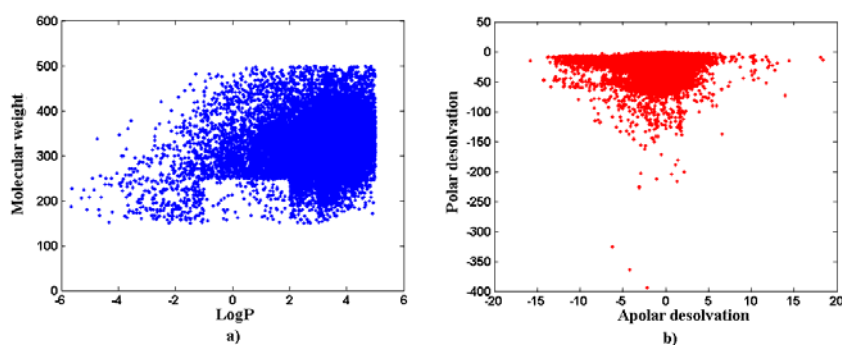
wówczas z dużym prawdopodobieństwem można stwierdzić, iż substancja nie ma cech potencjalnego leku. Zakłada się, iż naruszenie dwóch z podanych wartości granicznych właściwości molekularnych znacznie zmniejsza szanse zastosowania cząsteczki jako leku.

Pomimo stale zwiększającej się liczby dostępnych *on-line* baz danych istnieje zaledwie kilka repozytoriów łączących dane o małych cząsteczkach - ligandach oraz celach molekularnych,

głównie makromolekułach białkowych. Konkretnie konformacje ligandów pobierane są z repozytoriów zawierających dane strukturalne przeznaczone do programów dokujących z predefiniowanymi współrzędnymi przestrzennymi atomów np. baza danych ZINC. Baza ZINC stanowi repozytorium w większości komercyjnych danych przeznaczonych do badań przesiewowych HTS. W przeprowadzonych symulacjach baza ZINC w wersji 8 zaimportowana została przy użyciu środowiska MoStBioDat na lokalny serwer baz danych.

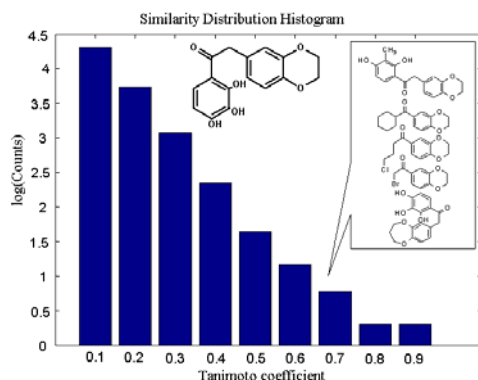


Rys. 4. Histogram deskryptorów molekularnych, uzyskany w trakcie skringu bazy danych MoStBioDat/Ligand (hbd/hba – donor/akceptor wiązania wodorowego, nrb – liczba wiązań zdolnych do rotacji).



Rys. 5. Rozkład obliczonej wartości logP względem masy cząsteczkowej (a) oraz desolvatacji polarnej względem desolvatacji polarnej (b).

Bardziej zaawansowane techniki skringu opierają się na tekstowych ciągach znakowych SMARTS. W swoich założeniach notacja SMARTS stanowi uogólnienie kodowania liniowego struktury cząsteczki w postaci ciągu znaków języka SMILES. Poszukiwanie wspólnych motywów strukturalnych polega na zakodowaniu fragmentu referencyjnego w postaci ciągu znaków SMARTS, a następnie identyfikacji poszukiwanej podstruktury w określonej cząsteczce pobranej z bazy danych. Algorytm porównania opiera się na zasadach podobnych do reguł wyrażeń regularnych (RE) znanych z systemu operacyjnego Linux.



Rys. 6. Histogram dystrybucji cząsteczek w bazie ZINC dla wybranych wartości współczynnika podobieństwa Tanimoto obliczony przy pomocy środowiska MoStBioDat dla przykładowego związku referencyjnego.

Przeprowadzone analizy wskazują jednoznacznie na przydatność pakietu MoStBioDat do badania lekotypii oraz (bio)różnorodności znacznych zasobów cząsteczek chemicznych. Dokładne omówienie modułów pakietu MoStBioDat oraz ich przykładowe zastosowanie do analizy podobieństwa cząsteczkowego zostało przedstawione w publikacji [H2].

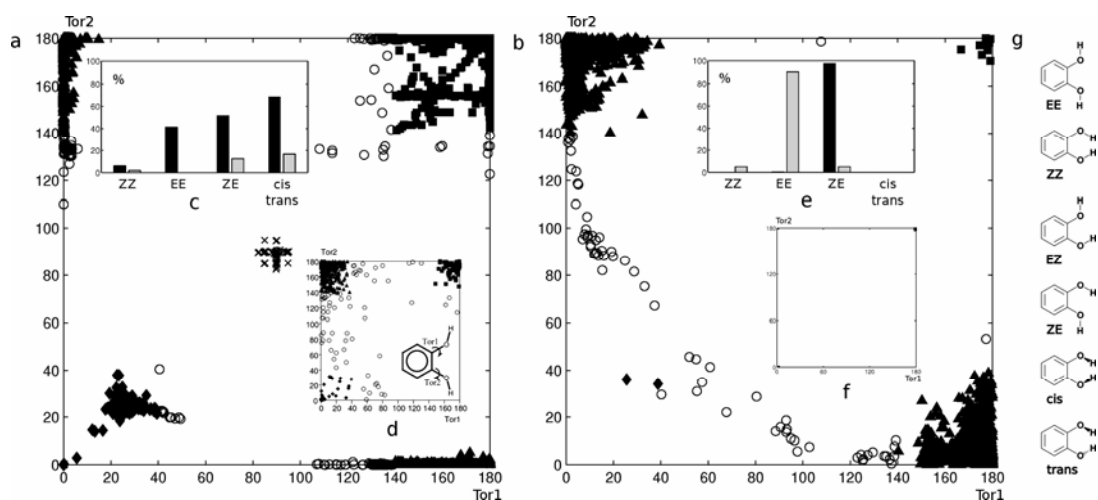
4. Fragmentaryczne mapowanie lekotypii w środowisku MoStBioDat: Analiza strukturalna motywu katecholu, β -ketoenu oraz diketokwasu.

Pomimo wzrostu liczby dostępnych danych strukturalnych, zarówno doświadczalnych jak i generowanych *in silico*, stanowią one jedynie niewielki fragment ‘nieskończonej’ przestrzeni chemicznej szacowanej na 10^{100} - 10^{200} molekuł [27]. Wydaje się, iż statystyczna analiza zbiorów danych, charakteryzujących się znaczną różnorodnością strukturalną powinna ułatwić identyfikację struktury wiodącej, która podlegając procedurze optymalizacji może doprowadzić do leku. Niestety, relatywnie mała liczba wprowadzanych na rynek leków wskazują na niską skuteczność metody HTS. Powstaje więc pytanie o jakość i różnorodność ‘wirtualnych’ danych strukturalnych oraz granice ich stosowalności. W przeprowadzonych analizach porównano geometrie molekuł zawierających układy o znanych właściwościach inhibicyjnych enzymu integrazy wirusa HIV z dostępnymi danymi doświadczalnymi.

Obecne metody projektowania leków w coraz większym stopniu wykorzystują w pełni zdefiniowane trójwymiarowe struktury receptorów i enzymów białkowych. Przestrzenna budowa wielu z nich pozostaje jednak wciąż nieznana, a dostępna wiedza ogranicza się jedynie do znajomości struktur wiązanych przez nie ligandów. Zbiór taki pośrednio opisuje miejsce oddziaływania, a poszukiwanie wspólnych elementów strukturalnych szeregu aktywnych ligandów zastępuje bezpośrednie modelowanie oddziaływań ligand-receptor. Tworzone międzycząsteczkowe lub wewnątrzcząsteczkowe wiązania wodorowe należą do istotnych czynników warunkujących wzajemne rozpoznawanie oraz sposób niekonwalencyjnego wiązania ligandu do jego celu makromolekularnego.

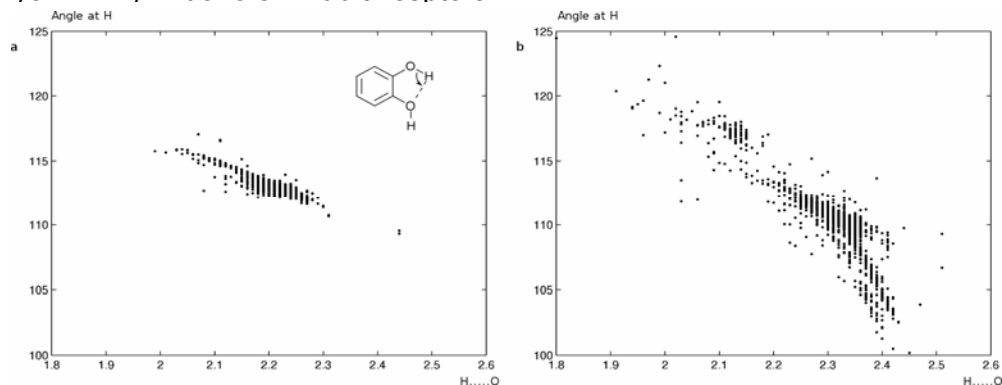
W przeprowadzonych badaniach przeanalizowano geometryczną strukturę fragmentów strukturalnych katecholu, β -ketoenu oraz diketokwasu oraz wpływ przestrzennego rozmieszczenia poszczególnych podstawników na możliwość i siłę tworzonych wewnątrz- oraz zewnątrzcząsteczkowych wiązań wodorowych. W symulacjach komputerowych wykorzystano oprogramowanie MoStBioDat do porównania zasobów bazy danych ZINC, PubChem, ChemBank oraz DrugBank.

Motyw strukturalny katecholu jest często spotykany w substancjach biologicznie czynnych m.in. 17 dostępnych na rynku specyfików, które zawierają ten fragment, zarejestrowanych jest przez FDA [28]. Analiza danych eksperymentalnych bazy danych CSD wykazała, iż 30% wszystkich rekordów zawierających fragment katecholu wykazuje aktywność biologiczną. Pomimo potencjalnych efektów ubocznych (znaczna toksyczność) około 1% leków posiada w swej budowie fragment katecholu. Wzajemne ułożenie dwóch hydroksylowych grup funkcyjnych powoduje, iż całość wykazuje zdolność do chelatowania dwuwartościowych jonów metali, co pośrednio wyjaśnia inhibicyjne właściwości w stosunku do integrazy wirusa HIV. Wydaje się, iż geometryczna orientacja grup hydroksylowych w pochodnych katecholu determinuje możliwość tworzenia wewnątrz- i zewnątrzcząsteczkowych wiązań wodorowych wpływając bezpośrednio na zdolność rozpoznawania molekularnego oraz oddziaływania z enzymem tej grupy związków. Poszczególne konformacje katecholu, wartości kątów torsyjnych wraz z ich dystrybucją dla wybranych baz danych przeanalizowane zostały przy użyciu środowiska MoStBioDat oraz porównane z dostępnymi danymi eksperymentalnymi [29]. Różne rozkłady absolutnych wartości kątów torsyjnych katecholu obserwowano dla populacji związków pochodzących z różnych zasobów baz danych. Największe podobieństwo geometryczne do danych eksperymentalnych wykazały rekordy bazy PubChem/ChemBank, podczas gdy związki pobrane z bazy ZINC posiadały głównie konformację Z,Z-cis/trans. Prawdopodobieństwo występowania poszczególnych konformacji wskazuje dominującą orientację Z,E dla bazy PubChem i ChemBank oraz E,E dla bazy DrugBank [30].



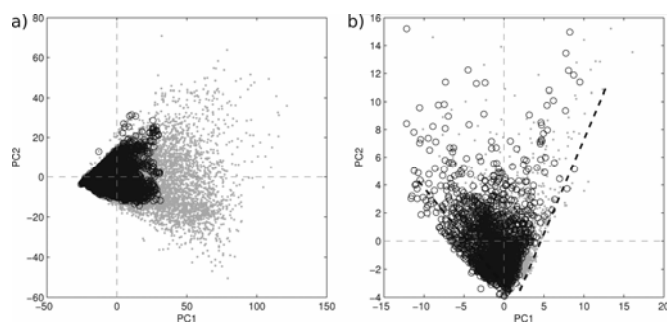
Rys. 7. Wykres zależności kątów torsyjnych określających konformacje katecholu dla baz danych ZINC vs. PubChem (a) oraz ChemBank (b) vs. DrugBank (f) wraz z porównaniem do bazy danych eksperymentalnych CSD (d). Histogram (c) ilustruje populacje poszczególnych konformacji dla baz PubChem (kolor czarny) oraz ZINC (kolor szary), zaś histogram (e) odpowiednio dla baz ChemBank vs. DrugBank.

Możliwość tworzenia wewnątrzcząsteczkowych wiązań wodorowych dla konfiguracji Z,E przebadano na podstawie analizy kątów i odległości pomiędzy poszczególnymi atomami grup hydroksylowych. Wewnątrzcząsteczkowe wiązanie wodorowe w katecholu, stabilizujące konformację Z,E jest na tyle słabe, że nie wpływa znacząco na możliwość tworzenia wiązań wodorowych z innymi donorami lub akceptorami.



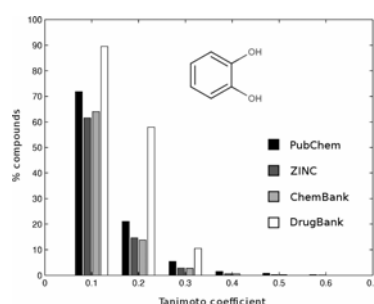
Rys. 8. Wykres zależności odległości H...O i kątów konformacji Z,E dla bazy Puchem (a) i ChemBank (b).

Populacje związków baz PubChem oraz ZINC, zawierające fragment katecholu, zostały opisane przez zbiór 1271 deskryptorów, obliczonych przy pomocy programu ADRIANA tworząc zbiór poddany następnie analizie PCA. Projekcja obiektów na płaszczyzny zdefiniowane przez czynniki główne PC1 i PC3 wskazuje na mniejszą różnorodność strukturalną związków bazy ZINC. Zdolność do tworzenia wiązań wodorowych przebadana została przy pomocy wybranych 15 deskryptorów dla zbioru 2578 struktur występujących zarówno w bazie ZINC jak i PubChem.



Rys. 9. Obiekty baz ZINC i PubChem rzutowane na płaszczyznę PC1 vs PC3 dla całego zbioru deskryptorów (a) oraz zbioru 15 deskryptorów określających możliwość tworzenia wiązań wodorowych (b).

Podobieństwo strukturalne do wybranego związku referencyjnego 1,2-dihydroksybenzenu w poszczególnych bazach danych przebadana została przy użyciu współczynnika Tanimoto.



Rys. 10. Analiza podobieństwa strukturalnego obiektów baz ZINC, PubChem, ChemBank i DrugBank.

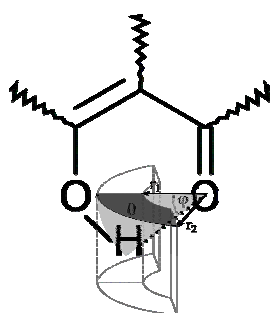
Mapowanie właściwości 1D (MW, logP), 2D (HBA, HBD) oraz liczby naruszeń reguły Ro5 Lipińskiego wykonane zostały przy pomocy sieci neuronowych Kohonena. Metoda okazała się przydatna do tworzenia dwuwymiarowych map wybranych deskryptorów molekularnych i analizy rozkładu określonych właściwości cząsteczkowych. Przeprowadzone analizy wykazały przydatność pakietu MoStBioDat do badania właściwości i geometrii znacznych zasobów baz danych. Dokładne omówienie poszczególnych modułów pakietu MoStBioDat oraz ich zastosowanie do analizy strukturalnej fragmentu katecholu zostało zamieszczone w publikacji [H5].

Związki zawierające motyw β -ketoenu, który jest tautomeryczną formą β -diketonu, stały się interesującym celem badań, ponieważ wykazały zdolności inhibycyjne w stosunku do enzymu integrazy wirusa HIV. Chelatujące właściwości tego układu atomów mogą znacząco wpływać na aktywność przeciwwirusową tej grupy związków. Eksploracja bazy DrugBank wykazała, iż 31 substancji z podstrukturą β -ketoenu zostało zarejestrowanych jako leki przez FDA.

Przestrzenne rozmieszczenie poszczególnych podstawników fragmentu β -ketoenu determinuje sposób oddziaływania z enzymem m.in. przez tworzenie zewnątrz- i wewnątrzcząsteczkowych wiązań wodorowych. W przeprowadzonych symulacjach komputerowych zastosowano środowisko MoStBioDat do analizy prawdopodobieństwa

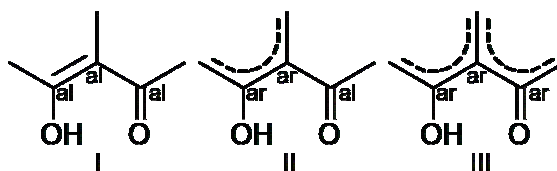
występowania wewnątrzcząsteczkowego wiązania wodorowego w związkach z fragmentem β -ketoenu. Dane dotyczące geometrii cząsteczek zostały pobrane z bazy danych ZINC (>8 milionów rekordów) oraz PubChem (>39 milionów rekordów) i porównane z wynikami krystalograficznymi dostępnymi w bazie danych CSD (>0,5 miliona rekordów) [31]. Liczba danych eksperymentalnych stanowi jedynie małą część ogólnej liczby dostępnych danych o topologii i topografii cząsteczek chemicznych. W tym kontekście pojawia się pytanie o jakość danych generowanych komputerowo, zwłaszcza w analizie niekowalencyjnych oddziaływań ligand-receptor w tym wiązań wodorowych.

Przeszukiwanie baz danych oparte o podstrukturę β -ketoenu pozwoliło zidentyfikować wszystkie związki zawierające ten motyw farmakoforowy istotny dla hamowania procesu integracji wirusa HIV z genomem gospodarza. Analiza danych eksperymentalnych bazy CSD ujawniła regularności kątów i odległości atomowych we fragmencie β -ketoenu.



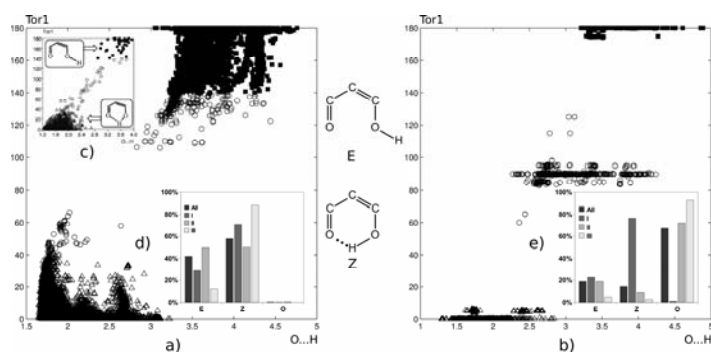
Rys. 11. Uprzywilejowane położenia grup donorowych i akceptorowych obliczone na podstawie danych eksperymentalnych $r_2=2.2\text{\AA}$, $r_1=1.6\text{\AA}$ and $|\phi|=120^\circ$, $|\theta|<30^\circ$.

Utworzenie wewnątrzcząsteczkowego wiązania wodorowego (IHB) jest wynikiem delokalizacji elektronów π , które obserwowane jest, gdy donorowy i akceptorowy atom tlenu znajdują w jednej płaszczyźnie. Postulowane zjawisko rezonansowo wspieranego wiązania wodorowego (RAHB) w tym fragmencie może wyjaśniać mniejsze odległości pomiędzy atomami tlenu oraz relatywną równowagę wszystkich wiązań C—C (długość pośrednia pomiędzy wiązaniem pojedynczym a podwójnym) [32]. W przeprowadzonych analizach wirtualnego skringu znacznych zasobów strukturalnych baz danych (ZINC, PubChem, DrugBank) określono uprzywilejowane położenie poszczególnych podstawników motywu β -ketoenu oraz bezpośredni wpływ przestrzennego rozmieszczenia tych podstawników na prawdopodobieństwo utworzenia IHB oraz RAHB. Znaleziono związki podzielono na 3 zasadnicze grupy uwzględniając charakter atomów węgla fragmentu β -ketoenu (alifatyczny lub aromatyczny) na β -ketoenie, pochodne α -acylofenylowe oraz pochodne w układzie bicyklicznym.



Rys. 12. Grupy związków β -ketoenu uwzględniające charakter atomów węgla (al – alifatyczny, ar – aromatyczny).

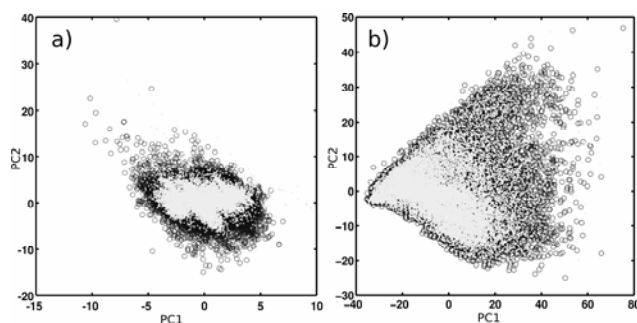
Szczegółowa analiza budowy przestrzennej fragmentu β -ketoenu wykazała, iż 58% rekordów bazy PubChem i 14% bazy ZINC przyjmuje konformację Z według przyjętych kryteriów danych eksperymentalnych bazy CSD. Dane PubChem są pogrupowane w 2 klasterzy związków o konformacjach Z i E, zaś dane ZINC tworzą 3 klasterzy zdominowane głównie przez konformacje przejściowe. W tym przypadku konformacje związków bazy danych PubChem wykazują lepszą zgodność z danymi krystalograficznymi bazy CSD, co wskazuje, iż generatory struktur 3D zaimplementowane w bazie PubChem uwzględniają informacje eksperymentalne.



Rys. 13. Wykres zależności odległości O...H i kąta torsyjnego Tor, dla baz danych PubChem (a) i ZINC (b) oraz CSD (c). Histogramy obrazują dystrybucję poszczególnych konformacji dla baz danych PubChem (d) i ZINC (e).

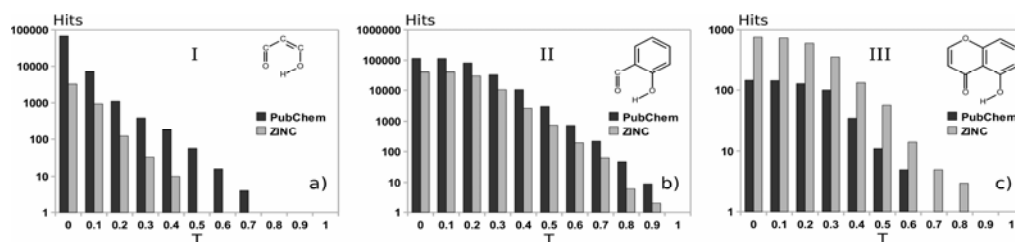
Obliczone wartości odległości i kątów istotnych dla tworzenia wiązań IHB oraz RAHB dla baz danych PubChem oraz ZINC pozostają w zgodzie z danymi doświadczalnymi [30]. Stabilizacja fragmentu β -ketonu przez tworzenie wiązania wewnątrzcząsteczkowego, zależna od charakteru atomów węgla, powoduje słabsze oddziaływania z innymi grupami donorowymi lub akceptorowymi.

Użycie sieci neuronowej Kohonena SOM umożliwiło analizę rozkładu wybranych właściwości fizykochemicznych, włączając w to dane strukturalne, przestrzenne oraz przynależność do określonych grup związków zawierających motyw β -ketoenu. Zastosowanie metody PCA dla zbioru związków PubChem i ZINC opisanych zbiorem deskryptorów obliczonych programem ADRIANA wykazało, iż rekordy bazy ZINC charakteryzują się mniejszą różnorodnością strukturalną związków z podstrukturą β -ketoenu. Przeprowadzono również analizę podobieństwa do wybranych związków referencyjnych w poszczególnych grupach molekuł z fragmentem β -ketoenu używając współczynnika Tanimoto jako miary podobieństwa cząsteczkowego w zakresie wartości od 0,1 do 0,9.



Rys. 14. Wykres zależności odległości O...H i kąta torsyjnego Tor_1 dla baz danych PubChem (a) i ZINC (b) oraz CSD (c). Histogramy obrazują dystrybucję poszczególnych konformacji dla baz danych PubChem (d) i ZINC (e).

W grupie, gdzie wszystkie atomy węgla motywu β -ketoenu są alifatyczne (I), wyraźne podobieństwo do 3-hydroksy-propenal wykazują cząsteczki bazy PubChem. Dokładnie odwrotna zależność obserwowana jest dla grupy ze wszystkimi atomami C o charakterze aromatycznym, gdzie dominują rekordy bazy ZINC. W przypadku grupy II obserwowana jest podobny rozkład związków w danym zakresie wartości współczynnika Tanimoto dla bazy PubChem i ZINC.

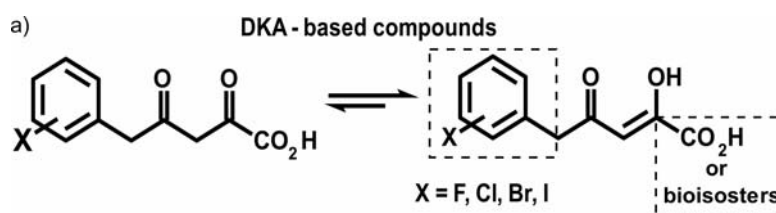


Rys. 15. Histogram różnorodności cząsteczkowej mierzonej współczynnikiem Tanimoto dla poszczególnych grup związków zawierających fragment β -ketoenu.

Szczegółowa analiza parametrów wiązań IHB oraz RAHB przy użyciu pakietu MoStBioDat została opisana w pracy [H6].

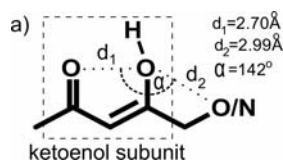
Integraza jest enzymem niezbędnym do replikacji wirusa HIV. Katalizuje on integrację (włączenie) wirusowego DNA (donora) do DNA gospodarza (akceptora) [33]. Brak funkcjonalnych homologów integrazy w ludzkim genomie spowodował znaczne zainteresowanie blokowaniem działania tego enzymu, szczególnie na etapie przeniesienia nici (ST) wirusowego DNA do chromosomowego DNA gospodarza. Enzym integrazy funkcjonalnie zbudowany jest z 3 zasadniczych części: domeny N-końcowej (NTD), centrum katalicznego (CCD) oraz domeny C-końcowej (CTD) [34]. Wolne grupy karboksylowe trzech reszt aminokwasowych (Asp, Asp, Glu) w CCD zdolne są do tworzenia wiązania koordynacyjnego z jonami dwuwartościowymi (Mn^{2+} , Mg^{2+}), co jest istotne w procesie katalicznym. Kompetencyjne wiązanie inhibitora z integrazą w CCD może prowadzić do dezaktywacji enzymu lub zmniejszenia jego aktywności katalicznej. W znacznym uproszczeniu mechanizm działania IN wirusa HIV można przedstawić w trzech zasadniczych

etapach. Pierwszy, 3'-modyfikacja (3'P), który przebiega w cytoplazmie, polega na odszczepieniu dwóch lub trzech nukleotydów z każdego końca 3' nici wirusowego DNA z utworzeniem kompleksu przedintegracyjnego (PIC). Uwolnione w ten sposób grupy hydroksylowe stają się czynnikami nukleofilowymi, które w reakcji transestryfikacji (S_N2) oddziałują z DNA gospodarza w jądrze zainfekowanej komórki. W drugim etapie przeniesienia nici (ST) następuje wbudowanie pary zmodyfikowanych końców wirusowego DNA do genomu ludzkiego – miejsca integracji w obu niciach oddalone są o pięć par zasad purynowych lub pirymidynowych. W etapie trzecim następuje swoista naprawa przerwanych nici DNA gospodarza, co prowadzi do przywrócenia ciągłości poszczególnych nici. Dwuwartościowe jony metali (kofaktory) pośredniczą w połączeniu integrazy ze szkieletem fosfodiesterowym DNA na etapie 3'-terminacji oraz przeniesienia nici. Odkrycie chelatujących właściwości fragmentu diketokwasu (DKA) okazało się punktem zwrotnym w projektowaniu potencjalnych inhibitorów integrazy etapu przeniesienia nici (INSTI). Inhibitor, chelatując jony metali, blokuje centrum aktywne enzymu, co zapobiega integracji wirusowego DNA z ludzkim genomem. Analiza przestrzenna grupy inhibitorów IN wirusa HIV umożliwiła opracowanie różnych modeli farmakoforowych. Współpłaszczyznowe ułożenie fragmentu β -ketoenolu oraz donorowych atomów tlenu lub azotu jest często spotykanym motywem strukturalnym w molekułach INSTI, gdzie odpowiednio trzy heteroatomy tworzą sferę koordynującą dwuwartościowe jony magnezu lub manganu. Wątpliwość budzi jednak liczba jonów metali dwuwartościowych niezbędnych w centrum aktywnym integrazy podczas procesu katalitycznego, gdyż istnieją inhibitory, których działanie wyjaśnia mechanizm jednokofaktorowy.



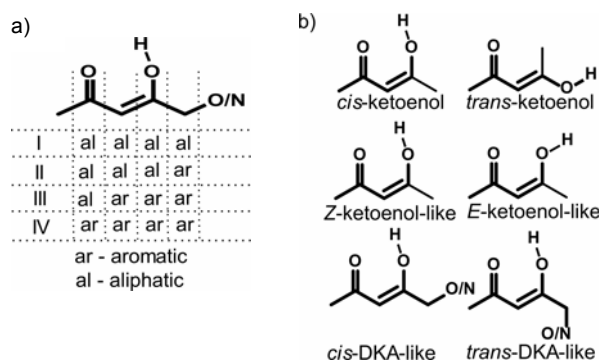
Rys. 16. Chemotypia układu DKA.

Zastąpienie grupy karboksylowej grupą bioizosteryczną (np. układy triazolu lub tetrazolu) wpływa na zmniejszenie toksyczności projektowanych inhibitorów integrazy wirusa HIV. Przestrzenne rozmieszczenie poszczególnych grup fragmentu DKA jest kluczowe dla procesu potencjalnego tworzenia i stabilizacji wewnątrzcząsteczkowych wiązań wodorowych z wytworzeniem sześciocząłowego pseudopierścienia $=C-O-H \cdots O=C-C$. Opierając się na danych eksperymentalnych opracowano model farmakofora układu DKA.



Rys. 17. Model farmakofora układu DKA .

W przeprowadzonych analizach przestrzennego rozmieszczenia poszczególnych grup atomów motywu DKA zastosowano literaturowy model farmakofora do wirtualnego skryningu znacznych zasobów baz danych ZINC, PubChem, Ligand.Info oraz DrugBank używając pakietu MoStBioDat. Związki zawierające podstrukturę DKA podzielone zostały na 4 grupy w zależności od charakteru atomów węgla wchodzących w skład tego motywu strukturalnego.

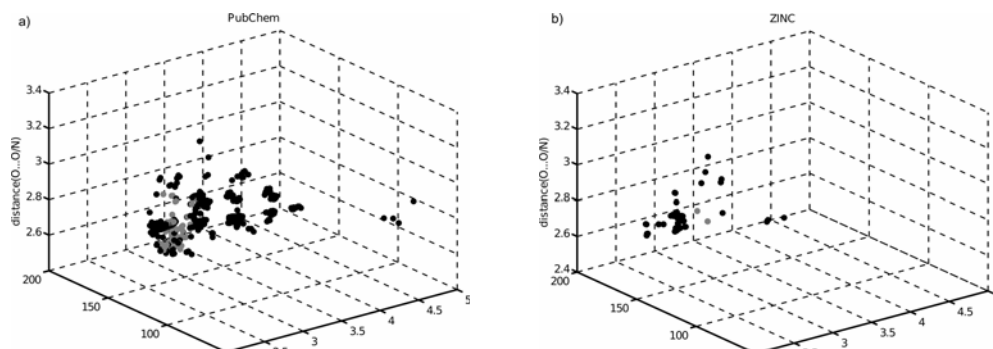


Rys. 18. Główne grupy związków zawierających fragment DKA (a) oraz możliwe konformacje (b).

Opierając się na dostępnych danych literaturowych oraz doświadczalnych danych krystalograficznych bazy danych PDB dla znanych inhibitorów integrazy (kody: 3YOA, 3L2W, 3L2V, 1QS4) wprowadzono szereg ograniczeń geometrycznych dla motywu DKA (tzw. model-DKA) gdzie:

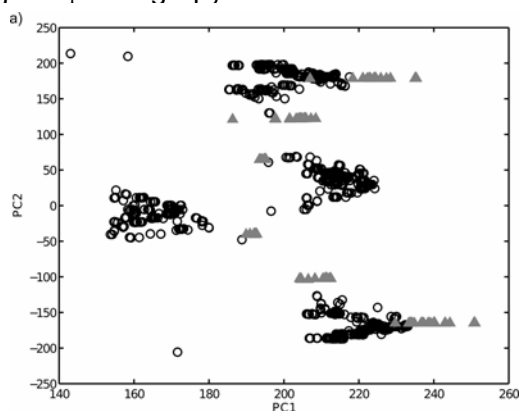
- odległość $2.65 \leq d_1(\text{O} \cdots \text{O}) \leq 2.90 \text{ \AA}$
- odległość $2.60 \leq d_2(\text{O} \cdots \text{O/N}) \leq 2.99 \text{ \AA}$
- kąt $130 \leq \alpha(\text{O} \cdots \text{O} \cdots \text{O/N}) \leq 150^\circ$

Przeprowadzono szczegółową analizę odległości i kątów pomiędzy poszczególnymi atomami fragmentu DKA dla komputerowo generowanych rekordów baz danych, zaś otrzymane wyniki porównano z danymi doświadczalnymi bazy CSD.



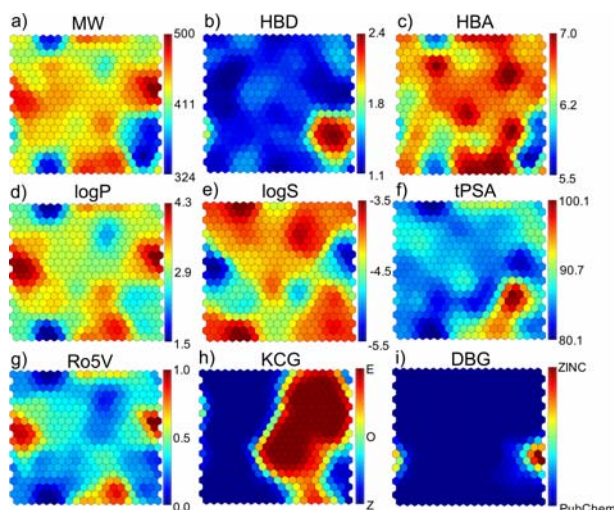
Rys. 19. Wykres zależności odległości d1, d2 i kąta α dla bazy PubChem (a) i ZINC (b) .

Podjęto próbę interpretacji wpływu geometrycznego rozmieszczenia atomów motywu DKA na możliwe tworzenie wewnątrzcząsteczkowych wiązań wodorowych. W przypadku, gdy wszystkie atomy węgla są alifatyczne potencjalnie tworzone wiązanie wewnątrzcząsteczkowe jest najsilniejsze, zaś oddziaływania te słabną wraz ze wzrostem liczby aromatycznych atomów węgla motywu DKA. W szeregu związków o najniższej energii wiązań wewnątrzcząsteczkowych obserwowany jest największy efekt delokalizacji elektronów π . Zbiór molekuł opisanych parametrami geometrycznymi obliczonymi dla fragmentu DKA poddano analizie PCA. Rzutowanie obiektów na płaszczyznę zdefiniowaną przez parę czynników głównych PC1 vs. PC2 (75% opisanej wariancji) ujawnia 4 klaster, co jest zgodne z podziałem tej związków grupy.



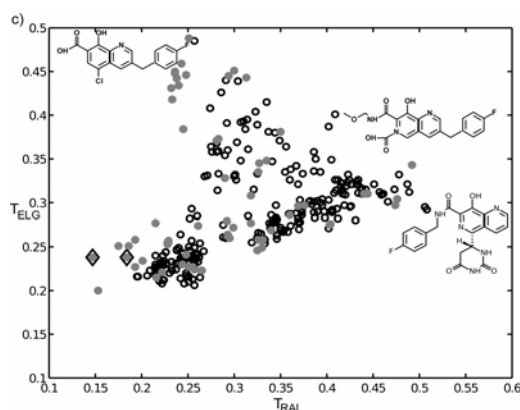
Rys. 20. Rzutowanie DKA na PC1 vs. PC2 dla bazy PubChem (kolor czarny) i ZINC (kolor szary).

Wykonano projekcję wybranych właściwości fizykochemicznych przy pomocy samoorganizującej się sieci Kohonena wytrenowanej zbiorem deskryptorów geometrycznych motywu DKA.



Rys. 21. Projekcja wybranych właściwości fizykochemicznych związków DKA przy użyciu sieci SOM.

Dokonano analizy różnorodności strukturalnej w poszczególnych grupach związków zawierających podstrukturę DKA w oparciu o współczynnik podobieństwa Tanimoto względem raltegraviru i elvitegraviru jako molekuł odniesienia. Wydaje się, iż *mono*-halogenopochodne podstawione w pozycji *para* wykazują najwyższe podobieństwo strukturalne do cząsteczek referencyjnych.



Rys. 22. Podobieństwo strukturalne wyrażone współczynnikiem Tanimoto względem raltegraviru (RAL) i elvitegraviru (ELG) dla rekordów bazy danych PubChem (kółka) i ZINC (romby).

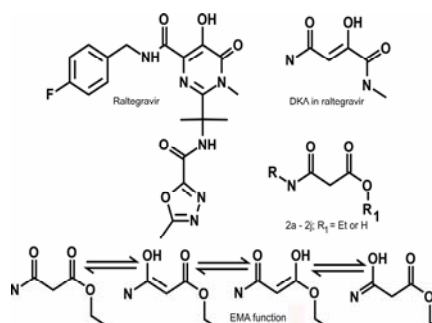
Szczegółowa analiza parametrów opisujących geometrię związków z motywem DKA przy użyciu pakietu MoStBioDat opisana została w pracy [H8].

4. Amidy etylowe kwasu malonowego – nowa grupa inhibitorów integrazy wirusa HIV.

Racjonalne projektowanie nowych farmaceutyków sprowadza się często do poszukiwania analogii w zbiorze znanych cząsteczek działających w naturalnych systemach biologicznych. Porównanie struktur cząsteczek, definiowanie oraz pomiar podobieństwa strukturalnego jest jedną z podstawowych technik wykorzystywanych w różnych metodach modelowania molekularnego m.in. w analizie podobnych ugrupowań uprzywilejowanych wzorców leków

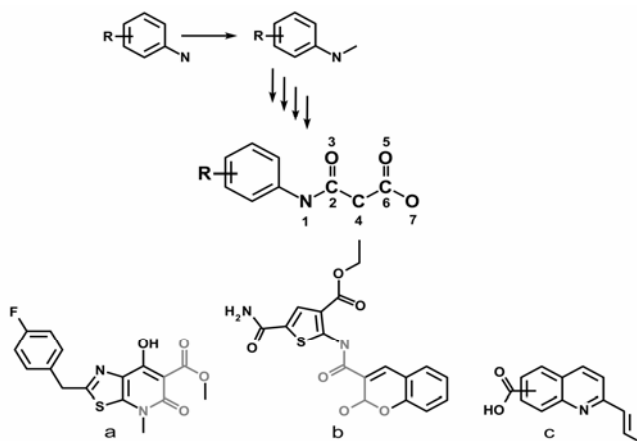
lub podobieństwie między lekiem a jego celem makromolekularnym tzw. receptorem. W projektowaniu molekularnym mianem receptora określa się zarówno białka receptorowe, enzymy a także inne układy molekularne, z którymi oddziałują leki. Pełny opis mechanizmu działania leku wymaga wyjaśnienia oddziaływania typu lek-receptor. Metody projektowania w oparciu o dane strukturalne receptora stanowią cel, do którego dąży się w dzisiejszych technologiach komputerowego wspomaganie projektowania molekularnego (CAMD). Wydaje się, że bezpośrednie włączenie do opisu oddziaływań ligand-receptor geometrii miejsca wiążącego daje możliwość lepszego zrozumienia procesów rozpoznawania molekularnego [35]. Przestrzenne rozmieszczenie poszczególnych atomów kompleksu ligand-receptor, które uwzględnia efekty steryczne oraz elektrostatyczne definiuje technika zwana dokowaniem. Poszukiwana jest struktura niskocząsteczkowego bioefektora jako komplementarnego dopełnienia receptora. Jedną z technik, która umożliwia wzajemne dopasowanie geometryczne konformacyjnie labilnego układu ligand-receptor jest procedura dynamiki molekularnej (MD). Analiza różnych aspektów podobieństwa molekularnego w zbiorze aktywnych ligandów umożliwia generowanie swobodnego rodzaju 'negatywu' struktury receptora. Proces tworzenia tego odwrotnego obrazu receptora na bazie analogii do aktywnych ligandów nazywany jest mapowaniem receptora.

Opracowanie nowych inhibitorów IN wirusa HIV-1 wydaje się kluczowe z punktu widzenia polifarmakologicznej terapii anty-HIV. Odkrycie motywu diketokwasu (DKA), który może być odpowiedzialny za blokowanie działania integrazy wydaje się być przełomem w badaniach nad inhibitorami integrazy. Podkreślić należy, iż pomimo znacznych wysiłków komercyjnie dostępne są jedynie dwa inhibitory IN – raltegravir (isentress) oraz mieszanina zawierająca elivitegravir (stribild). Co więcej, nie został w pełni poznany mechanizm dezaktywacji IN, ponieważ wciąż brakuje pełnej struktury krystalograficznej IN. Pojawiły się jednak modele teoretyczne kompleksu DNA-IN reprezentujące tzw. układ przedintegracyjny (PIC) oraz krystalograficzne prototypy PFV kompleksowane raltegravirem i elivitegravirem co umożliwia prowadzenie symulacji komputerowych technikami modelowania molekularnego [36]. Badania teoretyczne umożliwią opracowanie nowych modeli inhibitorów enzymu integrazy wirusa HIV, zaś praktyczne badania aktywności potencjalnych celów syntetycznych dostarczą nowych danych eksperymentalnych. W przeprowadzonych badaniach bazujących na technice fragmentacji zastosowano wysokowydajne badania przesiewowe (HTS) sprzężone z technikami dokowania molekularnego do projektowania związków o potencjalnej aktywności biologicznej (głównie inhibitorów integrazy wirusa HIV). Pomimo, iż amidy etylowe kwasu malonowego (EMA) wykazują strukturalne podobieństwo do fragmentu DKA w raltegravirze rzadko badano ich potencjalną aktywność anty-HIV.



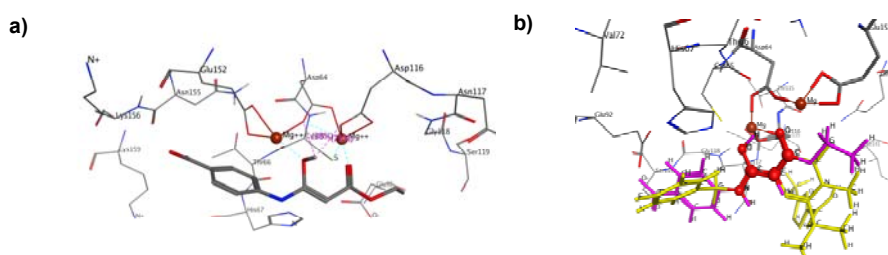
Rys. 23. Podobieństwo strukturalne EMA i raltegraviru.

W przeprowadzonych badaniach opisano metody syntezy oraz wyznaczono aktywność biologiczną serii związków EMA, co pozwoliło na modelowanie zależności aktywność-struktura tej grupy cząsteczek. Eksploracja bazy danych ChemDB pozwoliła na wybranie molekuł o potencjalnej aktywności anty-HIV. Utworzona biblioteka poddana została analizom różnorodności strukturalnej fragmentu DKA przy użyciu metody fragmentacji inkrementacyjnej w środowisku MoStBioDat.



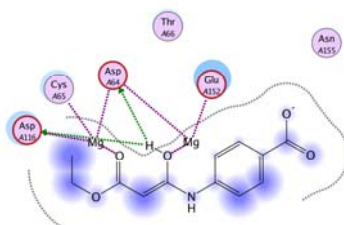
Rys. 24. Konstrukcja kolejnych zapytań do bazy danych NIAID ChemDB przy pomocy technik fragmentacji inkrementacyjnej (dodawanie atomów 1-7).

Nie znaleziono rekordów zawierających pełny fragment EMA, lecz zidentyfikowano substancje zawierające ten motyw maskowany układem heterocyklicznym (rys. 24a,b). Fragmentaryczna procedura badań zależności QSAR sprzężona z technikami dokowania molekularnego ułatwia identyfikację konformacji i orientacji (pozy) wybranych pochodnych EMA w miejscu aktywnym enzymu IN. W modelowaniu wykorzystano geometrię układu DNA-IN pobranego z bazy danych PDB (kod: 1WKN). Ułożenie przestrzenne zadokowanej serii pochodnych EMA było zgodne z danymi literaturowymi oraz układem referencyjnym – raltegravirem.



Rys. 25. Najaktywniejszy związek serii EMA zadokowany w miejscu aktywnym IN (a) oraz nałożony na strukturę raltegraviru (b).

Pomimo przeciwnej orientacji przestrzennej zadokowanej (rys. 25a) oraz nałożonej (rys. 25b) najaktywniejszej cząsteczki serii EMA, sposób oddziaływania wskazuje na istotne chelatujące znaczenie fragmentu EMA w modelowaniu inhibicyjnych właściwości tej grupy związków.



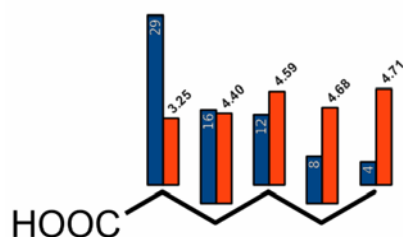
Rys. 26. Dwuwymiarowa mapa oddziaływań najaktywniejszej pochodnej EMA z wybranymi resztami aminokwasowymi miejsca aktywnego enzymu IN.

Szczegółowa analiza parametrów opisujących geometrię związków EMA, metody przeszukiwania bazy ChemDB przy użyciu pakietu MoStBioDat oraz ustawień dokowania molekularnego opisana została w pracy [H7].

5. Metody fragmentaryczne oraz koncepcja motywów uprzywilejowanych

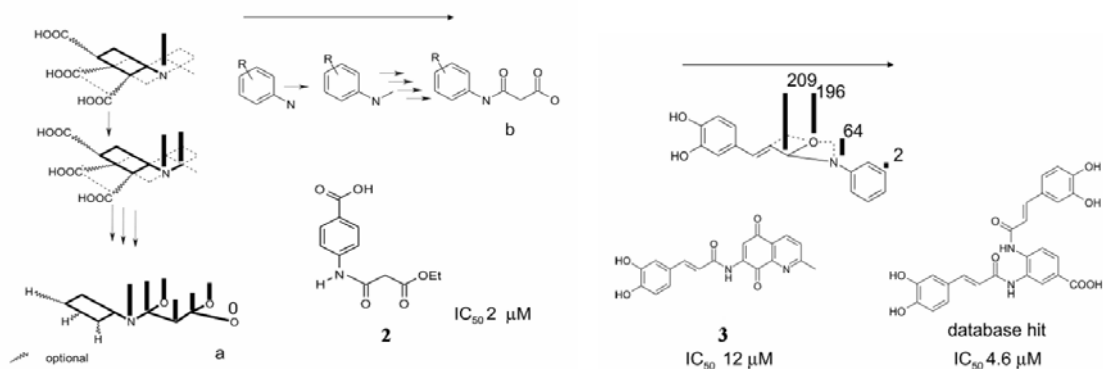
Procedura fragmentaryczna lub używane niekiedy zamiennie określenie *fragonomika* stała się alternatywnym sposobem identyfikacji fragmentów molekularnych o potencjalnie uprzywilejowanej konstytucji w stosunku do wybranych celów makromolekularnych. Samo pojęcie uprzywilejowania odnosi się głównie do najczęściej występujących lub współwystępujących ugrupowań cząsteczek oddziałujących z określonymi receptorami [37]. W tej metodologii zakłada się, iż istnieją preferowane 'trzony' stanowiące szkielet molekularny oraz łączące je motywy strukturalne. Strukturalnie różnorodna biblioteka fragmentów uprzywilejowanych może zostać wykorzystana jako podstawowy zestaw bloków do budowy cząsteczek o określonym profilu farmakologicznym względem różnych receptorów, pomimo relatywnie niższej złożoności (Ro3) oraz aktywności struktur wyjściowych [38]. W przeprowadzonych badaniach wykorzystano środowisko MoStBioDat do analizy zależności pomiędzy topologią a aktywnością wybranych fragmentów molekularnych

(FRAGTAL) otrzymanych przy pomocy procedury fragmentacji inkrementacyjnej lub dekrementacyjnej molekuł pobranych z bazy danych ChemDB oraz Ligand.Info.



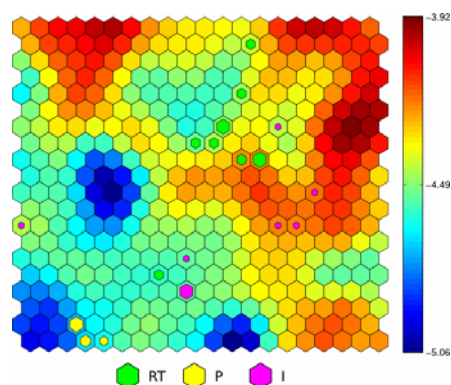
Rys. 27. Histogram częstotliwości (kolor niebieski) oraz wartości pKa (kolor pomarańczowy) dla kwasów karboksylowych o określonej liczbie atomów węgla.

Procedura FAGTAL pozwala określić stopień rozpowszechnienia danego fragmentu molekularnego w odpowiednio zdefiniowanej podprzestrzeni chemicznej. Algorytm fragmentacji inkrementacyjnej zastosowany został dla molekuł bazy ChemDB Anti-HIV/OI/TB oraz Ligand.Info zawierających motyw strukturalny diketokwasu oraz katecholu spotykany w lekach stosowanych w terapii przeciwko HIV. Odpowiednia modyfikacja aktywnej podstruktury powinna wskazać nowy cel syntetyczny o pożądanym profilu farmakologicznym. Nowosyntezowany związek **2** (rys. 28) wykazał aktywność biologiczną IC_{50} na poziomie 2 μM , zaś dla pochodnych kwasu kofeinowego uzyskano odpowiednio $IC_{50}=12 \mu M$ i $IC_{50}=4,6 \mu M$.



Rys. 28. Inkrementacyjna konstrukcja kolejnych deskryptorów FRAGTAL (a) dla układu określonego motywu strukturalnego (b). Wysokość słupków ilustruje częstotliwość występowania poszczególnych fragmentów.

Projekcja deskryptorów typu FRAGTAL dla pochodnych DKA, które opisane zostały zbiorem dodatkowych zmiennych na dwuwymiarowe mapy Kohonena, wskazała obszary o różnej aktywności względem odwrotnej transkryptazy, proteazy oraz integrazy. Pomimo, iż od leku oczekuje się selektywnego działania, w praktyce selektywność receptorowa jest to trudna do osiągnięcia. Analiza polifarmakologiczna, rozumiana jako innowacyjny model oddziaływania w układzie wieloreceptorowym, umożliwiła identyfikację cząsteczek charakteryzujących się znacznymi zdolnościami inhibicyjnymi poszczególnych enzymów wirusa HIV.

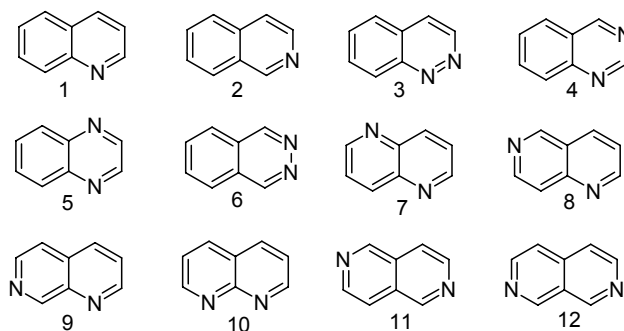


Rys. 29. Polifarmakologiczna mapa aktywności inhibitorów wirusa HIV względem odwrotnej transkryptazy (RT), proteazy (P) oraz integrazy (I). Wielkość sześciokąta wskazuje liczbę cząsteczek w neuronie.

Wizualna inspekcja utworzonej mapy polifarmakologicznej pokazuje, iż jedynie inhibitory proteazy lokują się w obszarach o dużej aktywności. Wydaje się, iż powyższa zależność wynika z relatywnie większej reprezentacji inhibitorów proteazy w stosunku do wszystkich analizowanych związków w przeszukiwanych bazach danych. Wskazane zostały też topologie układów z motywem DKA o potencjalnych właściwościach inhibicyjnych w stosunku do enzymów wirusa HIV. Szczegółowa analiza parametrów analizy FRAGTAL dla pochodnych DKA przy użyciu pakietu MoStBioDat opisana została w pracy [H10].

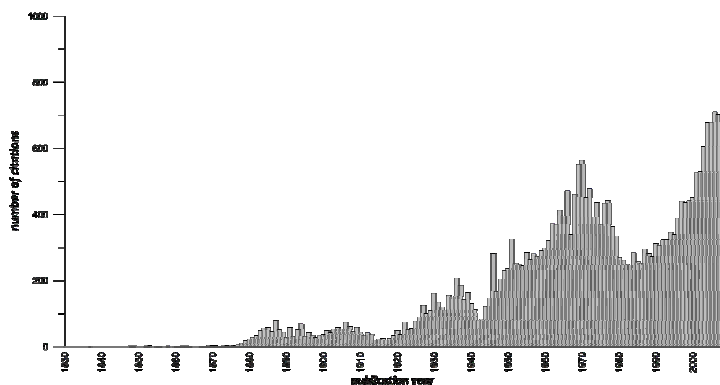
Wielokierunkowe działanie leków może powodować spektrum efektów ubocznych lub ewentualną aktywację grupy receptorów w celu uzyskania pożądanej odpowiedzi farmakologicznej [39]. Efektywność terapii przy użyciu tego typu leków zależy w głównej mierze od patogenezы schorzenia, a więc liczby czynników etiologicznych wywołujących daną chorobę. Selektywność receptorowa determinowana jest strukturą przestrzenną (topografią) bioefektora. Eksploracja repozytoriów chemicznych wykazała, iż uprzywilejowane fragmenty molekularne (PS) zawarte są w strukturach leków o wieloreceptorowym profilu działania. Biblioteka motywów o preferowanej konstytucji może stanowić interesujące źródło danych wyjściowych dla leków o potencjalnym działaniu wielokierunkowym. Praktycznie, przeprowadza się głównie analizę częstości występowania danych podstruktur w określonych populacjach ligandów np. leków wiążących się specyficznie z receptorami danej rodziny, choć można wskazać podjednostki strukturalne wykazujące szeroki zakres aktywności farmakologicznej [40]. Wydaje się, iż topologia takich fragmentów odpowiada za wielokierunkowe działanie (niespecyficzność) i powinowactwo względem wielu struktur białkowych, zaś wzajemne ułożenie podstawników determinuje selektywną zdolność rozpoznawania molekularnego. Okazało się również, iż wiele preferowanych fragmentów obecnych jest w strukturze związków o charakterze naturalnym, gdzie zapewne czynnik ewolucyjny wpłynął na skuteczność pewnych połączeń. Analiza danych literaturowych wykazała, iż pierścień naftalenowy jest ważną częścią strukturalną

wielu połączeń chemicznych. Znane są również przykłady leków oraz produktów pochodzenia naturalnego, które zawierających układ azanaftalenowy np. chinolinę. Pochodne chinoliny stanowią grupę związków o aktywności inhibicyjnej w stosunku do enzymu integrazy wirusa HIV [41]. W badaniach przeprowadzono analizę statystyczną wydzielonej uprzednio podprzestrzeni azanaftalenów zawierających układ mono- i diazanaftalenowy w oparciu o bazy danych Beilstein, DrugBank i PubChem.



Rys. 30. Fragment mono- (1-2) i diazanaftalenowy (3-12).

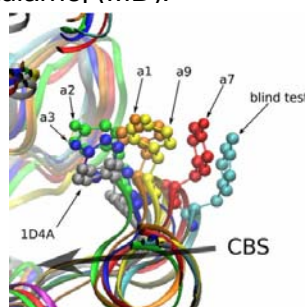
Otrzymane wyniki nie wskazały istotnej korelacji pomiędzy częstotliwością występowania oraz bioaktywnością danego fragmentu w populacji leków a jego stopniem rozpowszechnienia w badanej podprzestrzeni chemicznej. Czasowa propagacja liczby powyższych fragmentów, obserwowana dla bazy PubChem, potwierdziła powyższą zależność. Eksploracja danych bibliograficznych ujawniła okresy wzmożonego zainteresowania wykorzystaniem syntetycznym danego fragmentu azanaftalenowego wskazując na pewne trendy w syntezie organicznej. Ta swoista periodyczność obserwowana jest we fluktuacjach czasowych liczby publikacji poświęconych pochodnym chinoliny.



Rys. 31. Tendencje syntezy organicznej motywu chinoliny jako produktu reakcji.

Szczegółowe studium wiedzy o strukturach uprzywilejowanych wraz z praktycznymi badaniami lekotypii fragmentów mono- i diazanaftaleinowego, które obejmuje eksplorację dostępnych systemów bazodanowych pod kątem potencjalnego wykorzystania do konstrukcji nowych inhibitorów wirusa HIV zostało omówione w pracy przeglądowej [H9].

Izotiocyaniany (ITC) to grupa związków o profilaktycznym działaniu przeciwnowotworowym, których aktywność związana jest z aktywacją enzymu oksyreduktazy [42]. Badania wykazały, iż konsumpcja pokarmów bogatych w ITC znacznie zmniejsza ryzyko zachorowania na choroby nowotworowe [43]. W przeprowadzonych symulacjach komputerowych wskazano na możliwy niekompetencyjny mechanizm aktywacji enzymu, który posiada właściwości przeciwutleniające katalizując redukcję chinonów do hydrochinonów. Z drugiej strony enzym odpowiada za bioaktywację grupy proleków do form wykazujących znaczną cytotoksyczność [44]. Obserwacje doświadczalne wskazały grupę pochodnych tioizocyjanianów wiążących miejsce aktywne enzymu jako potencjalnych inhibitorów dysmutazy nadadtlenkowej. Zależności 3D-QSAR, otrzymane przy uwzględnieniu geometrii enzymu, charakteryzują się podobnymi parametrami dopasowania i predykcji jak typowe modele farmakoforowe np. CoMSA [45], dlatego podjęto próbę poszukiwań alternatywnych miejsc niekompetencyjnego wiązania (UBS) tej grupy inhibitorów. Przeanalizowano wpływ efektu allosterycznego na jakość modelowania oddziaływań enzymu z szeregiem inhibitorów indochinolinowych przy użyciu technik dynamiki molekularnej (MD).



Rys. 32. Zmiany w położeniu tyrozyny 128 w centrum katalitycznym powodowane wiązaniem aktywatora w UBS.

Aktywatory o dużej aktywności (a2,a3) wiązane w UBS powodują przesunięcie reszty aminokwasowej Tyr128, która jest istotna dla wiązania inhibitorów w centrum katalitycznym (CBS) w kierunku miejsca aktywnego. Przeciwny kierunek przesunięcia Tyr128 obserwowany jest dla aktywatorów o średniej i niskiej aktywności (a1,a7,a9). Wydaje się, że w układzie enzym/aktywator/inhibitor aktywator ułatwia penetrację miejsca aktywnego przez inhibitory. Szczegółowa analiza parametrów przygotowania kompleksu enzym/inhibitor/aktywator, dokowania molekularnego oraz dynamiki molekularnej dla szeregu izotiocyanianów i indochinolin przy częściowym wykorzystaniu programów zaimplementowanych w środowisku MoStBioDat opisana została w pracy [H4].

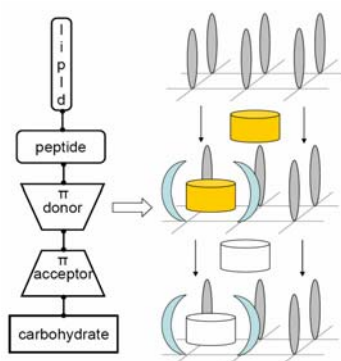
6. Sztuczne receptory w chemoinformatyce

Funkcjonowanie leków związane jest z makromolekułami białkowymi, pełniącymi funkcję receptorów lub enzymów, które odpowiadają za wywołanie określonych sygnałów biologicznych [46]. Poszukiwanie niskocząsteczkowego bioefektora, stanowiącego

komplementarne dopełnienie receptora, wymaga zdefiniowania sekwencji oraz geometrii makromolekuły. Czasem brakuje danych doświadczalnych niezbędnych do bezpośredniego badania oddziaływań w układzie ligand-receptor, a wiedza w tym zakresie ogranicza się do znajomości struktur wiązanych przez nie ligandów. Tworzony jest wtedy zastępczy obraz receptora - farmakofora. Pośrednie modelowanie złożonych interakcji tego typu może odbiegać od rzeczywistych procesów rozpoznawania molekularnego.

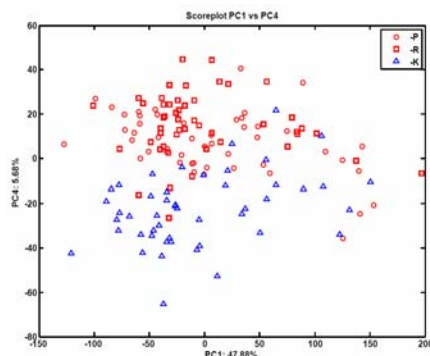
Koncepcyjnie sztuczne receptory to uproszczone modele rzeczywistych układów makromolekularnych, które zdolne są do selektywnego wiązania określonych szeregów ligandów [47]. Stanowią swoistego rodzaju sensory biomedyczne lub całe platformy diagnostyczne do systematycznego badania oddziaływań ligand-receptor. Kombinatorycznie tworzone biblioteki sztucznych receptorów mogą być użyte jako narzędzia do projektowania nowych związków chemicznych o potencjalnym zastosowaniu jako leki. Tego typu technikę traktować można jako fragmentaryczną procedurę badań zależności QSAR sprzężoną z technikami dokowania molekularnego. Fragmentami budulcowymi jest zbiór aminokwasów białkowych, kwasów tłuszczowych i cukrów. Istotnym problemem jest strukturalna aranżacja bloków budulcowych rozumiana jako ich sekwencja i trójwymiarowa geometria, która najlepiej odzwierciedli naturalne systemy biologiczne przy jednoczesnym zachowaniu względnej prostoty strukturalnej i możliwości syntetycznych.

W przeprowadzonych symulacjach podjęto próbę modelowania rozpoznawania molekularnego zbioru sztucznych receptorów (APR) otrzymanych w procesie samoorganizacji *N*-lipidowanych peptydów przyłączonych do celulozy przez *m*-aminofenyloamino-1,3,5-triazynę [48]. Zdolność selektywnego wiązania cząsteczek przeanalizowano na grupie inhibitorów receptora histaminowego H₃. Dostępne dane literaturowe oraz praktyka laboratoryjna wskazały na zbiór 7 aminokwasów (Glu, Lys, His, Trp, Ser, Ala, Pro) o różnym charakterze chemicznym i strukturalnym, które mogą być istotne dla rozpoznawania tych ligandów przez receptor. W znacznym uproszczeniu sztuczny receptor zbudowany jest z osnowy cukrowej (celuloza), części o charakterze π -akceptora oraz π -donora, dipeptydu oraz labilnej części lipidowej (reszty wyższych kwasów karboksylowych nasyconych i nienasyconych). Znaczna swoboda rotacji reszt kwasowych powoduje zmiany konformacyjne tych części sztucznego receptora w czasie wiązania ligandu, co odpowiada procesowi indukcyjnego dopasowania lub efektowi allosterycznemu w rzeczywistych systemach biologicznych. Siłę i selektywność powinowactwa szeregu inhibitorów histaminowych mierzono metodą pośrednią, poprzez wymywanie zadokowanych uprzednio molekuł barwnika i obserwację zmian w zabarwieniu matrycy sztucznych receptorów.



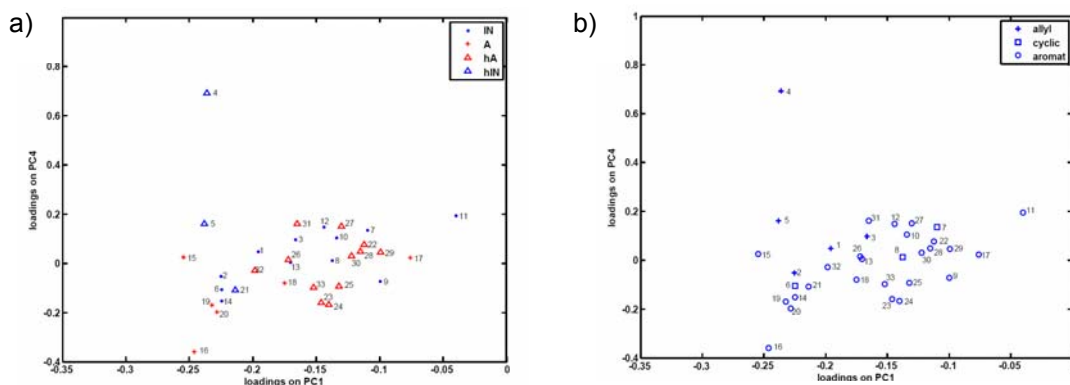
Rys. 33. Schemat budowy sztucznego receptora oraz procedury pomiaru powinowactwa inhibitorów.

Badany szereg związków podzielono na grupy uwzględniając ich chemotypię, charakter lipofilowy oraz profil aktywności do rzeczywistej makromolekuły. Do wstępnej analizy danych zastosowano metodę czynników głównych (PCA), która jest narzędziem szczególnie przydatnym w przypadku analiz wzajemnie skorelowanych danych wielowymiarowych. Umożliwia nie tylko ich kompresję i wizualizację, ale również eliminację części błędu eksperymentalnego [49]. W efekcie przyjętego sposobu doboru kolejnych czynników głównych (np. według kryterium maksymalnej wariancji, które opisują), znaczna część wariancji danych określana jest kilkoma początkowymi czynnikami, pozostałe zaś związane są z błędem pomiarowym i opisują niewielką część wariancji.



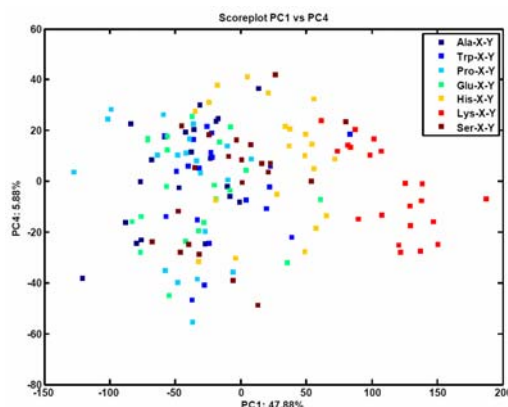
Rys. 34. Projekcja powinowactwa do sztucznych receptorów szeregu inhibitorów histaminowych na płaszczyzny zdefiniowane przez czynniki główne PC1 vs. PC4 z uwzględnieniem charakteru chemicznego (nasyconego i nienasyconego) oraz lipofilowości ($\log P > 5$ vs $\log P < 5$) części lipidowej.

Projekcja powinowactwa do sztucznych receptorów szeregu inhibitorów receptora H_3 na pary istotnych czynników głównych wskazuje na heterogeniczną strukturę danych, która ujawnia się poprzez zróżnicowanie względem czynnika PC4. Niejednorodność obserwowana jest w charakterze chemicznym oraz lipofilowości wyższych kwasów tłuszczowych użytych do konstrukcji sztucznych receptorów. Ocena wpływu parametrów, które różnicują poszczególne receptory, dokonywana jest przy projekcji wag na płaszczyzny zdefiniowane przez pary czynników głównych.



Rys. 35. Projekcja wag na płaszczyzny czynników głównych PC1 vs PC4 z (a) profilem aktywności oraz (b) chemotypią inhibitorów.

Obserwowana jest tendencja do grupowania się związków o podobnym profilu aktywności oraz budowie strukturalnej. Przebadano wpływ charakteru chemicznego poszczególnych reszt aminokwasowych w strukturze dipeptydu, tworzącego sztuczny receptor, na siłę oddziaływania ligand – sztuczny receptor.



Rys. 36. Projekcja powinowactwa do sztucznych receptorów szeregu inhibitorów histaminowych na płaszczyzny zdefiniowane przez czynniki główne PC1 vs. PC4 z uwzględnieniem charakteru chemicznego reszt aminokwasowych.

Obserwowana jest klasteryzacja układów zawierających reszty aminokwasowe o charakterze zasadowym (Lys i His) z dodatnio naładowanym atomem azotu w warunkach płynu fizjologicznego ($\text{pH}=7,2\div 7,4$) względem czynnika PC1. Wydaje się, że wiązanie inhibitora w tym obszarze determinowane jest czynnikami elektrostatycznymi. Zwiększona zdolność wiązania ligandów dla układów z Lys oraz His przekłada się na zmniejszenie bioselektywności sztucznych receptorów. Przeprowadzono analizę regresyjną dla eksperymentalnie wyznaczonej wartości aktywności oraz powinowactwa szeregu inhibitorów przy pomocy metody PLS sprzężonej z procedurą eliminacji zmiennych nieistotnych IVE. W podzbiorach o różnej lipofilowości części lipidowej wybrane zostały układy, które głównie zawierały reszty aminokwasowe Glu, Ser oraz His. Fakt ten potwierdza obserwowany eksperymentalnie wpływ Glu206 na zdolność rozpoznawania molekularnego naturalnego

receptora histaminowego. Szczegółowa analiza parametrów syntezy kompleksu enzym-inhibitor, analizy PCA i PLS przy częściowym użyciu programów zaimplementowanych w środowisku MoStBioDat opisana została w pracy [H11].

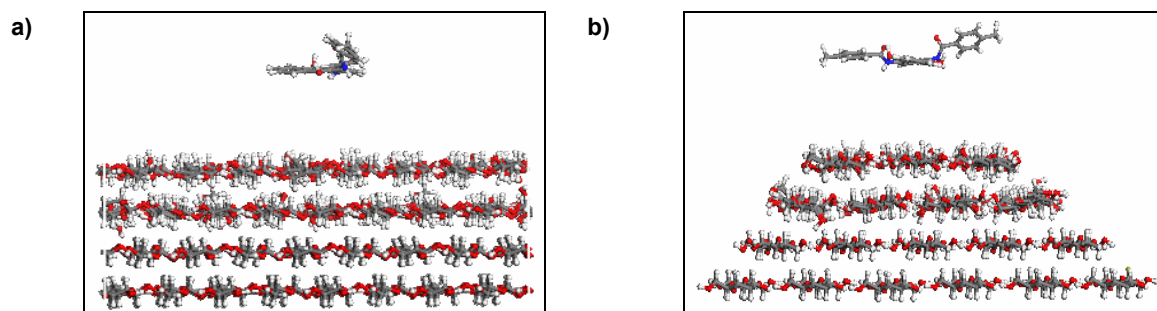
7. Modelowanie RD 4D-QSAR

Istotnym czynnikiem związanym z oddziaływaniem ligandów z receptorem są efekty określone mianem rozpoznawania molekularnego. Tego typu efekty nie są brane pod uwagę w technikach RI, a jedynym sposobem ich uwzględnienia jest dołączenie do badań struktury receptora. W przypadku, gdy znana jest struktura miejsca receptorowego projektowanie leku sprowadza się do poszukiwania przestrzennego rozmieszczenia atomów ligandu, które uwzględnia efekty steryczne i elektrostatyczne tworzonego kompleksu [50]. W takim podejściu prognozowanie stałych wiązania uwzględniać musi pełną termodynamikę wiązania, którą opisuje czynnik entalpowy i entropowy. Rozwój technologii informatycznych oraz eksperymentalnych technik analiz chemicznych (m.in. NMR) doprowadził do zwiększenia liczby dostępnych danych, które określają geometrię 3D makromolekuł białkowych pełniących funkcję receptorów. Jedną z technik, która wykorzystuje poszukiwania niskocząsteczkowego bioefektora jako komplementarne dopełnienie receptora nazywana jest dokowaniem molekularnym. Interesujące wydają się również uogólnienie wielowymiarowych technik QSAR poprzez uwzględnienie struktur receptorowych (RD), w szczególności wykorzystanie neuronowego formalizmu analiz 3D i 4D-QSAR sprzężonego z procedurami eliminacji zmiennych nieistotnych.

Różnorodność zjawisk, które należy uwzględnić przy wiązaniu się barwnika na powierzchni włókna celulozowego sugeruje na podobny mechanizm rozpoznawania molekularnego jak w układzie ligand-receptor. Wykazano, iż nierównomierne rozmieszczenie molekuł barwnika na powierzchni polisacharydu sugerować może istnienie miejsc o charakterze receptorowym, które charakteryzują się zwiększonym powinowactwem do molekuł barwnika [51].

W modelowaniu RI-QSPR oddziaływań barwnik-celuloza zastosowano techniki generowania *tinktofora*, który jest osobliwym odpowiednikiem *farmakofora* w klasycznym badaniu zależności aktywność-struktura. Pomimo, iż nie potwierdzono jednoznacznie receptorowego charakteru oddziaływań podjęte zostały próby modelowania powinowactwa barwników do celulozy technikami RI-QSAR. Znaczny stopień skomplikowania oddziaływań sugeruje, iż taki opis jest niewystarczający do pełnego modelowania opisanych procesów. Wydaje się, iż bezpośrednie włączenie do symulacji struktury przestrzennej celulozy umożliwi lepsze zrozumienie procesów zachodzących na jej powierzchni podczas wiązania barwnika. W celu porównania wyników modelowania RD 4D-QSAR oraz RD SOM-4D-QSAR tego typu związków z danymi literaturowymi RI przeprowadzono symulacje dla szeregu barwników

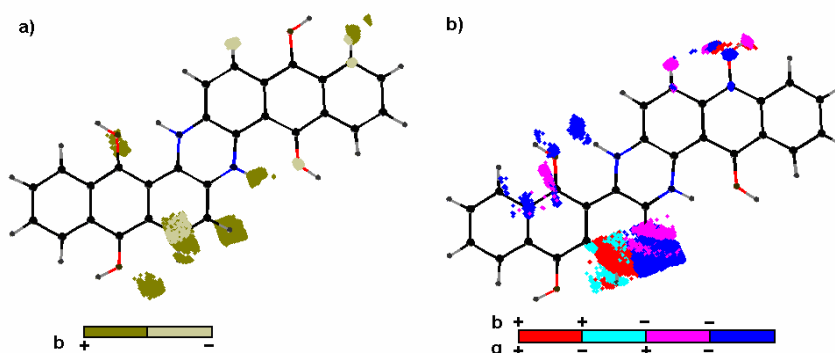
antrachinonowych. Walidacja modeli została wykonana przy użyciu stochastycznej procedury oceny modeli SMV, która iteracyjnie dzieli próbkowaną przestrzeń na dwa podzbiory: treningowy i testowy. Zastosowana procedura eliminacji nieistotnych zmiennych IVE-PLS wskazała obszary istotne dla modelowania powinowactwa barwników do celulozy tworząc przestrzenny obraz *tinktofora*. Zbiór molekuł barwnika do analizy konformacyjnej wygenerowany został przy pomocy procedur dynamiki molekularnej w układzie barwnik-celuloza.



Rys. 37. Początkowa geometria układu barwnik-celuloza. Rzuty odpowiednio równoległy (a) oraz prostopadły (b) do osi włókna celulozowego.

Otrzymane wyniki modelowania wskazują, że czynnik steryczny związany z przestrzennym rozmieszczeniem i kształtem barwnika na powierzchni celulozy ma pośredni wpływ na ich wzajemne oddziaływania elektrostatyczne. Konformacja barwnika decyduje o rozkładzie ładunków cząstkowych na poszczególnych jego atomach. Niewielkie zmiany w konfiguracji molekuly indukować mogą zmiany geometrii włókna celulozowego. Nadal trwają spekulacje czy kształt przestrzenny barwnika ma bezpośredni wpływ na wartości powinowactwa do włókna. W celu oceny statystycznej poprawności wybranych modeli zastosowana została iteracyjna procedura podziału zbioru barwników na dwa podzbiory: treningowy (2/3) i testowy (1/3). Ze względu na złożoność obliczeniową systematyczny podział zastąpiono procedurą o charakterze probabilistycznym wybierając 10^6 przypadkowo wygenerowanych dystrybucji związków na dwa podzbiory. Obserwacja wartości parametrów dopasowania modelu oraz jego zdolności predykcyjnych potwierdziła brak bezpośredniej zależności pomiędzy tymi zmiennymi. Wzrost wartości dopasowania modelu nie musi oznaczać jego lepszych zdolności predykcyjnych. Z podprzestrzeni chemicznej, charakteryzującej się największą częstotliwością występowania modeli, w przypadkowy sposób wybrano 1000 próbek do monitorowania liczby zmiennych istotnych dla oddziaływań barwnik-włókno przy użyciu metody IVE-PLS. Takie podejście pozwala na identyfikację obszarów specyficznych oddziaływań międzymolekularnych determinujących powinowactwo szeregu barwników do włókna przy zachowaniu jakości otrzymywanych modeli. Wskazane zostały obszary molekuł, które wnoszą największy wkład do modelowania powinowactwa do celulozy tej grupy

związków. Graficzna reprezentacja wybranych obszarów, wartość i znak współczynników regresji, umożliwia bezpośrednią ocenę cech przestrzennych barwnika oraz pośrednio włókna, które mają istotne znaczenie dla ich wzajemnych interakcji – tworzony jest więc model *tinktofora* uwzględniający geometrię celulozy. Szczegółowa analiza parametrów dynamiki molekularnej kompleksu barwnik-celuloza, analizy RD 4D-QSAR oraz RD SOM-4D-QSAR przy częściowym użyciu programów zaimplementowanych w środowisku MoStBioDat opisana została w pracy [H12].



Rys. 38. Obszary istotne dla modelowania RD 4D-QSAR szeregu barwników antrachinonowych wybrane metodą IVE-PLS dla (a) deskryptorów występowania (b) deskryptorów uśrednionego ładunku. Kolory kodują znak przy danej zmiennej oraz wartość ładunku.

Podsumowanie i wnioski

Znaczny postęp technologiczny prowadzi do wzrostu liczby dostępnych danych. Symulacje komputerowe oraz praktyczne analizy chemiczne generują informacje, które wymagają opracowania procedur efektywnego ich przetworzenia, analizy, kompresji oraz składowania. Z tego względu, w badaniach związanych z konstruowaniem nowych obiektów chemicznych nie należy pomijać wydajnego systemu gromadzenia dokumentacji, przeszukiwania oraz wymiany danych. Komputer stał się nieodzownym atrybutem spotykanym we współczesnych laboratoriach chemicznych, zaś specjalistyczne oprogramowanie zdolne do przechowywania oraz zarządzania informacją masową tworzy platformę wspomagającą chemika w codziennej pracy. Zarządzanie danymi o obiektach przestrzeni chemicznej wymaga jednorodnej reprezentacji cząsteczki w systemie binarnym. Jednoznaczne kodowanie obiektów złożonych z atomów i łączących je wiązań chemicznych w języku maszynowym nie jest zadaniem łatwym. Intensywny rozwój technologii informatycznych, w tym systemów bazodanowych, dostarcza narzędzi do przechowywania oraz eksploracji repozytoriów chemicznych.

W ramach prowadzonych badań zaprojektowano, zaimplementowano oraz praktycznie wykorzystano w analizach platformę obliczeniową MoStBioDat (Molekularną i Strukturalną Bioinformatyczną Bazę Danych). Środowisko MoStBioDat umożliwia poprawę oraz

przechowywanie danych w ujednoliconym formacie co zapobiega ich replikacji, tworzy także przyjazną platformę programistyczną z łatwym dostępem do analizowanych danych. Opracowane narzędzie wykorzystane zostało w szeregu analiz znacznych zasobów strukturalnych baz danych, m.in. ZINC, PubChem, DrugBank i Ligand.Info. Odkrywanie wiedzy zawartej w systemach bazodanowych, rozumiane jako pozyskiwanie potencjalnie użytecznej informacji na podstawie danych masowych, jest obecnie istotnym kierunkiem badań, który ma coraz szersze zastosowanie w racjonalnym projektowaniu nowych celów syntetycznych m.in. nowych substancji o potencjalnym zastosowaniu jako leki np. inhibitorów wirusa HIV. Przeszukiwanie baz danych oparte o podstruktury molekularne oraz przesiewowa analiza cząsteczek wraz z zastosowanym filtrem, który określa wymagany motyw strukturalny w molekuale, nazywa się skринingiem (HTS). W przeprowadzonych symulacjach wykorzystano procedury eksploracji baz danych w celu pozyskania wiedzy przydatnej do poszukiwania nowych celów syntetycznych oraz uprzywilejowanych fragmentów molekularnych w odniesieniu do koncepcji lekotypii.

Coraz szersze zastosowanie w racjonalnym projektowaniu nowych celów syntetycznych znalazły również techniki dekompozycji związków chemicznych (szczególnie znanych leków) do mniejszych fragmentów strukturalnych. Generowane w ten sposób biblioteki 'uprzywilejowanych' motywów strukturalnych o określonych właściwościach fizykochemicznych (np. reguła Ro3) tworzą podprzestrzeń przestrzeni leków. Różnorodne kombinacje motywów wyjściowych znacznie lepiej mogą opisywać skomplikowane oddziaływania w układzie ligand-receptor, co bezpośrednio wpływa na aktywność biologiczną takich połączeń. Porównawcze analizy HTS wskazują, iż struktury wiodące są również znacznie mniej złożone w porównaniu do wywodzących się z nich leków, na co wskazują niższe wartości mas molowych oraz wartości hydrofobowości. Tworzone zbiory charakteryzują się zazwyczaj znacznie większą różnorodnością strukturalną, która dodatkowo uzyskiwana jest mniejszym kosztem obliczeniowym. Z drugiej strony, badania praktyczne wykazały, iż większa złożoność strukturalna tworzonych bibliotek molekularnych nie zwiększa prawdopodobieństwa uzyskania pozytywnego wyniku w teście biologicznym. Niniejsza rozprawa stanowi kompendium, które obejmuje cykl 12 publikacji reprezentujących badania, gdzie:

- wykorzystano neuronowy formalizm analiz 4D-QSAR do masowego generowania oraz porównania podobieństwa bibliotek ligandów;
- zaprojektowano oraz praktycznie zaimplementowano bazy danych przechowujących informację o geometrii i właściwościach fizykochemicznych, profilu biologicznym ligandów oraz celów makromolekularnych wraz ze zbiorem modułów do obsługi bazy i analizy przechowywanych danych;

- przeprowadzono badania przesiewowe repozytoriów ZINC, PubChem, DrugBank itd. przy pomocy filtrów strukturalnych oraz analiz podobieństwa cząsteczkowego dla pochodnych zawierających motyw katecholu, β -ketoenu, diketokwasu;
- mapowano lekotypę fragmentów katecholu, β -ketoenu, diketokwasu przy użyciu metod projekcyjnych;
- wskazano nowe cele syntetycznych o potencjalnym zastosowaniu jako leki m.in. inhibitory enzymu integrazy wirusa HIV - analiza amidów etylowych kwasu malonowego;
- opracowano oraz wykorzystano procedurę fragmentaryczną FRAGTAL w koncepcji polifarmakologii inhibitorów integrazy wirusa HIV;
- modelowano mechanizm rozpoznawania molekularnego cząsteczek do biblioteki sztucznych receptorów;
- zastosowano neuronowy formalizm analiz 4D-QSAR, który uwzględnia struktury receptorowe w odniesieniu do modelowania oddziaływań ligand-receptor.

Reasumując, celem projektu było zastosowanie metod przesiewowych wobec dużych molekularnych baz danych w poszukiwaniu nowych bioefektorów, w szczególności potencjalnych inhibitorów integrazy wirusa HIV. Zasadniczym etapem projektu było zaprojektowanie, komputerowa implementacja oraz testowanie zarówno bazy danych jak i zintegrowanego w pakiecie MoStBioDat oprogramowania. Cały pakiet, zawierający ponad 20 tys. linii kodu, składa się z zestawu modułów wykonujących określone zadania, dostarczając narzędzi do obsługi i analizy informacji zgromadzonej w bazie danych. W toku prac wykonano lokalne kopie całości lub części ogólnodostępnych baz internetowych: ZINC, PubChem, DrugBank, ChemBank. W celu wskazania nowych inhibitorów integrazy wyodrębniono grupy związków zawierające określone motywy strukturalne: katecholu, β -ketoenu, diketokwasu, chinoliny.

W toku dalszych prac planowane jest udoskonalenie bioinformatycznej bazy danych ligandów i struktur receptorowych przygotowanej we współpracy z Austrian Institute of Technology.

Omówienie pozostałych osiągnięć naukowo-dydaktycznych

Moje zainteresowania naukowe skupiają się w głównej mierze wokół problemów modelowania molekularnego w projektowaniu leków. Swoją pracę naukową rozpocząłem w Zakładzie Chemii Teoretycznej UŚ pod kierunkiem prof. dr hab. Stanisława Kucharskiego, gdzie zajmowałem się zastosowaniem metody sprzężonych klasterów w obliczeniach kwantowych. Stosunkowo szybko zainteresowałem się jednak praktycznymi aplikacjami obliczeń teoretycznych w chemii organicznej oraz projektowaniu leków. Po przeniesieniu do Zakładu Chemii Organicznej UŚ, pod opieką prof. dr hab. Jarosława Polańskiego przygotowałem pracę doktorską poświęconą zastosowaniu metod dynamiki molekularnej w chemii organicznej, w szczególności modelowaniu QSAR.

Grupa pod kierownictwem prof. dr hab. Jarosława Polańskiego posiada bogate doświadczenie w rozwoju oraz praktycznym wykorzystaniu metod RI m-QSAR, tzn. w modelowaniu wielowymiarowych zależności aktywność-struktura na podstawie szeregu aktywnych molekuł z wytworzeniem przestrzennej mapy oddziaływań [52]. Zbiór biologicznie czynnych ligandów opisuje szereg związków chemicznych stanowiąc swoisty odwrotny obraz receptora (negatyw). W dotychczasowych pracach przeanalizowałem problemy modelowania układów biologicznie czynnych w aspekcie oddziaływania leku z receptorem, przy czym metodologia tych badań opierała się na założeniach RI 3D-QSAR.

Cząsteczka o najwyższej aktywności użyta jako zastępczy obraz receptora tworzy wzorcową sieć Kohonena, która jest zdolna do symulacji rozpoznawania molekularnego oraz wizualizacji różnych aspektów podobieństwa międzycząsteczkowego. Opracowana metoda porównawczej analizy powierzchni cząsteczkowej (CoMSA) daje niejednokrotnie lepsze wyniki niż klasyczna metoda porównawczej analizy pola cząsteczkowego (CoMFA) [53]. W toku badań zmodyfikowana została standardowa procedura eliminacji zmiennych nieistotnych UVE-PLS. Opisałem zastosowanie metodologii 3D/4D-QSAR do oceny modelowania efektów elektronowych i sterycznych pKa pochodnych kwasu benzoowego, kwasów alkanowych, steroidów kompleksowanych przez globuliny CBT/TBG [52]. Opracowany iteracyjny algorytm eliminacji zmiennych IVE-PLS nie powoduje statystycznej degradacji jakości modeli, umożliwiając jednocześnie graficzną reprezentację obszarów wnoszących istotny wkład do modelowania [54]. W ramach prowadzonych badań opracowałem również neuronowy formalizm analiz 4D-QSAR, który jest swoistym połączeniem klasycznej techniki 4D-QSAR z samoorganizującą się siecią Kohonena (SOM-4D-QSAR) [55]. Zaproponowałem modyfikację klasycznych deskryptorów 4D-QSAR przez uwzględnienie dodatkowych deskryptorów opisujących zmiany ładunków cząstkowych na poszczególnych atomach w czasie dynamiki molekularnej. Procedura SOM-4D-QSAR umożliwia transformację trójwymiarowej przestrzeni konformacyjnej w jej dwuwymiarową

mapę z zachowaniem relacji topologicznych w zbiorze wejściowym. Sprzężenie metody SOM-4D-QSAR z technikami eliminacji zmiennych umożliwia rzutowanie wybranych neuronów dwuwymiarowej mapy na trójwymiarową trajektorię poszczególnych atomów cząsteczki, a tym samym atrakcyjną wizualizację fragmentów wpływających na aktywność biologiczną badanej klasy związków. Opracowane techniki znalazły szerokie zastosowanie w modelowaniu różnych szeregów związków chemicznych, w tym potencjalnych inhibitorów integrazy wirusa HIV. Efektywność metod opartych o sieci neuronowe wzbogacone o procedury selekcji zmiennych istotnych zademonstrowałem modelując aktywność pochodnych steroidowych inhibitorów aromatazy, pochodnych 2,4 – diamino – 5 – benzylopirymidyny, pochodnych styrylochinolinowych inhibitorów integrazy wirusa HIV, a także powinowactwo barwników azowych oraz antrachinonowych do celulozy. Omówiłem również sposoby walidacji modeli 3D i 4D-QSAR metodą sprzężoną walidacji LSO/LOO. Współczesne metody projektowania substancji biologicznie czynnych w coraz większym stopniu opierają się na w pełni zdefiniowanych trójwymiarowych strukturach białkowych. Interesujące wydaje się uogólnienie wielowymiarowych technik QSAR poprzez uwzględnienie struktury receptora, w szczególności wykorzystanie neuronowego formalizmu 3D/4D QSAR sprzężonego z technikami dokowania molekularnego.

W swojej pracy zajmowałem się również opracowaniem multimedialnych materiałów dydaktycznych, zaimplementowanych na portalu *e-Chem* oraz *iCSE*, przeznaczonych do wspomagania nauczania chemii metodą zdalną [56,57]. Zaangażowanie w utworzenie platformy *e-learningowej* zaowocowało również opracowaniem skryptu pt. „Podstawy chemoinformatyki leków” objaśniającego pojęcia i metody intensywnie rozwijającej się gałęzi chemii – chemoinformatyki [58]. Jest to podręcznik akademicki wprowadzający studentów w tajniki zastosowań informatyki w chemii ukierunkowany na problemy związane z projektowaniem i poszukiwaniem związków biologicznie czynnych.

W ostatnim czasie nauka, która częściowo kojarzyła się z pasją i zaangażowaniem, stała się domeną rankingów i klasyfikacji opartych na analizie publikacji ukazujących się w czasopiśmie branżowych co decyduje niejednokrotnie o karierach indywidualnych osób. Zasada *‘publikuj albo giń’* stała się podstawą działalności naukowej całych wydziałów a niejednokrotnie i uczelni. Opracowaniem statystycznych indeksów i parametrów dla monitorowania i oceny jakości nauki zajmuje się relatywnie nowa dziedzina zwana naukometrią lub *scientometrią*. W przeprowadzonych ostatnio analizach zwróciłem uwagę na zależność pomiędzy wkładem finansowym, a liczbą publikacji w najbardziej prestiżowych czasopiśmie o zasięgu międzynarodowym [59]. Badania wysokości finansowania nauki w poszczególnych krajach i na 50 najlepszych uniwersytetach wskazały na swoistego rodzaju *efekt świętego Mateusza*, gdzie *‘każdemu bowiem, kto ma, będzie dodane, tak że nadmiar mieć będzie. Temu zaś, kto nie ma, zabiorą nawet to, co ma’* [60]. Czysta

statystyka, która opiera się jedynie na analizie rozkładu publikacji i ich cytowań, pomijając czynnik ludzki, jest w mojej ocenie jedną z pułapek naukometrii.

W chwili obecnej zajmuję się rozwijaniem procedur i programów, które umożliwiają włączenie różnych technik dokowania do wielowymiarowych analiz QSAR tzw. dokowań konsensusowych.

Literatura

- [H1]. **A. Bak**, J. Polanski, 2007, 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.*, 47, 1017-1021.
- [H2]. **A. Bak**, J. Polanski, A. Kurczyk, 2009, The use of MoStBioDat for rapid screening of molecular diversity, *Molecules*, 14(9), 3436-3445.
- [H3]. **A. Bak**, J. Polanski, T. Stockner, A. Kurczyk, 2010, MoStBioDat - Molecular and Structural Bioinformatics Database, *Comb. Chem. High Throughput Screen.*, 13, 366-374.
- [H4]. P. Mazur, T. Magdziarz, **A. Bak**, Z. Chilmonczyk, T. Kasprzycka-Guttman, I. Misiewicz, K. Skupinska, J. Polanski, 2010, Does molecular docking reveal alternative chemopreventive mechanism of activation of oxidoreductase by sulforaphane isothiocyanates?, *J. Mol. Model.*, 16, 1205-1212.
- [H5]. **A. Bak**, T. Magdziarz, A. Kurczyk, J. Polanski, 2011, Mapping Drug Architecture by MoStBioDat: Rapid Screening of Intramolecular Hydrogen Bonded Motifs in Catechols, *Drug Develop. Res.*, 72, 209-218.
- [H6]. **A. Bak**, T. Magdziarz, A. Kurczyk, J. Polanski, 2011, Mapping Fragmental Druglikeness in the MoStBioDat Environment: Intramolecular Hydrogen Bonding in Beta-Ketoenols, *Comb. Chem. High Throughput Screen.*, 14(7), 560-569.
- [H7]. K. Serafin, P. Mazur, **A. Bak**, E. Laine, L. Tchertanov, J-F Mouscadet, J. Polanski, 2011, Ethyl malonate amides: A diketo acid offspring fragment for HIV integrase inhibition, *Bioorg. Med. Chem.*, 19, 5000-5005.
- [H8]. **A. Bak**, T. Magdziarz, J. Polanski, 2012, Pharmacophore-Based Database Mining for Probing Fragmental Drug-Likeness of Diketo Acid Analogues, *SAR&QSAR Env. Res.*, 23, 185-204.
- [H9]. J. Polanski, A. Kurczyk, **A. Bak**, R. Musiol, 2012, Privileged Structures – Dream or reality: Preferential Organization of Azanaphthalene Scaffold, *Curr. Med. Chem.*, 19, 1921-1945.
- [H10]. **A. Bak**, T. Magdziarz, A. Kurczyk, K. Serafin, J. Polanski, 2013, 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.*, 16(4), 274-287.
- [H11]. **A. Bak**, M. Daszykowski, Z. Kaminski, K. Kiec-Kononowicz, K. Kuder, J. Fraczyk, B. Kolesinska, P. Ciosek, J. Polanski, 2013, Probing an artificial polypeptide receptor library using a series of novel histamine H3 receptor ligands, *Comb. Chem. High Throughput Screen.*, 7(2), 141-156.
- [H12]. **A. Bak**, M. Wyszomirski, T. Magdziarz, A. Smoliński, J. Polanski, 2014, Structure-based modeling of dye-fiber affinity with SOM-4D-QSAR paradigm: Application to set of anthraquinone derivatives. *Comb. Chem. High Throughput Screen.*, 17(10), 1-18.
- [1]. J. Devillers, 2013, Methods for building QSARs. *Methods Mol. Biol.*, 930, 3-27.
- [2]. H.C. Kolb, M.G. Finn, K.B. Sharpless, 2001, Click chemistry: Diverse chemical function from a few good reactions. *Angew. Chem. Int. Ed. Engl.*, 40, 2004-2021.
- [3]. C.H. Andrade, K.F.M. Pasqualoto, E.I. Ferreira, A.J. Hopfinger, 2010, 3D-pharmacophore mapping of thymidine-based inhibitors of TMPK as potential antituberculosis agents. *J. Comput. Aided Mol. Des.*, 24, 157-172.

- [4]. M.P. Gonzalez, C. Teran, L. Saiz-Urra, M. Teijeira, 2008, Variable selection methods in QSAR: an overview. *Curr. Top Med. Chem.*, 18, 1606-1627.
- [5]. N. Nunthaboot, S. Pianwanit, V. Parasuk, S. Kokpol, J. Briggs, 2007, Computational studies of HIV-1 integrase and its inhibitors. *Curr. Comput. –Aided Drug Des.*, 3, 160-190.
- [6]. H. Niedbala, J. Polanski, R. Gieleciak, R. Musiol, D. Tabak, B. Podeszwa, **A. Bak**, A. Palka, J.-F. Mouscadet, J. Gasteiger, M. Le Bret, 2006, Comparative molecular surface analysis (CoMSA) for virtual combinatorial library screening of styrylquinoline HIV-1 blocking agents, *Comb. Chem. High Throughput Screen.*, 9, 753-770.
- [7]. J. Polanski, 2003, Self-organizing neural networks for pharmacophore mapping. *Adv. Drug Deliv. Rev.*, 55, 1149-1162.
- [8]. J. Polanski, R. Gieleciak, **A. Bak**, 2002, The Comparative Molecular Surface Analysis (CoMSA) - A Nongrid 3D QSAR Method by a Coupled Neural Network and PLS System: Predicting pKa Values of Benzoic and Alkanoic Acids, *J. Chem. Inf. Comput. Sci.*, 42(2), 184-191.
- [9]. J. Polanski, B. Walczak, 2000, The Comparative Molecular Surface Analysis (CoMSA):A Novel Tool for Molecular Design, *Comput. Chem.*, 24, 615-625.
- [10]. A. Hopfinger, S. Wang, J. Tokarski, B. Jin, M. Albuquerque, P. Madhav, C. Duraiswami, 1997, Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. *J. Am. Chem. Soc.*, 119, 10509-10524
- [11]. V. Centner, D.L. Massart, O.E. de Noord, S. de Jong, B.M.V. Vandeginste, C. Sterna, 1996, Elimination of Uninformative Variables for Multivariate Calibration. *Anal. Chem.*, 68, 3851-3858.
- [12]. J. Polanski, R. Gieleciak, 2003, The Comparative Molecular Surface Analysis (CoMSA) with Modified Uninformative Variable Elimination-PLS (UVE-PLS) Method: Application to The Steroids Binding the Aromatase Enzym. *J. Chem. Inf. Comput. Sci.*, 43, 656-666.
- [13]. G. Schuurmann, S. Funar-Timofei, 2003, Multilinear regression and comparative molecular field analysis (CoMFA) of azo dye-fiber affinities. 2. Inclusion of solution-phase molecular orbital descriptors. *J. Chem. Inf. Comput. Sci.*, 43, 1502-1512.
- [14]. S. Timofei, L. Kurunczi, W. Schmidt, Z. Simon, 1996, Lipophilicity in dye-cellulose fiber binding. *Dyes Pigm.*, 32, 25-42.
- [15]. J. Polanski, R. Gieleciak, M. Wyszomirski, 2003, Comparative molecular surface analysis (CoMSA) for modeling dye-fiber affinities of the azo and anthraquinone dyes. *J. Chem. Inf. Comput. Sci.*, 43, 1754-1762.
- [16]. J. Polanski, R. Gieleciak, **A. Bak**, K. Jarzembek, M. Wyszomirski, 2002, The Comparative Molecular Surface Analysis (CoMSA). A New Efficient Technique for Drug Design, *Acta Pol. Pharm.*, 59(6), 459-461.
- [17]. J. Polanski, R. Gieleciak, T. Magdziarz, **A. Bak**, 2004, The GRID Formalism for the Comparative Molecular Surface Analysis: Application to the CoMFA Benchmark Steroids, Azo dyes and HEPT derivatives, *J. Chem. Inf. Comput. Sci.*, 44(4), 1423-1435.
- [18]. K. Mazeau, M. Wyszomirski, M., 2012, Modelling of Congo red adsorption on the hydrophobic surface of cellulose using molecular dynamics. *Cellulose*, 9, 1495-1506.
- [19]. J. Polanski, R. Gieleciak, **A. Bak**, 2004, Probability Issues in Molecular Design. Predictive and modeling ability in 3D-QSAR schemes, *Comb. Chem. High Throughput Screen.*, 7(8), 793-807.
- [20]. **A. Bak**, J. Polanski, 2006, A 4D-QSAR study on anti-HIV HEPT analogues, *Bioorg. Med. Chem.*, 14, 273-279.

- [21]. R. Gieleciak, T. Magdziarz, **A. Bak**, J. Polanski, 2005, Modeling Robust QSAR. 1. Coding Molecules in 3D-QSAR - from a Point to Surface Sectors and Molecular Volumes, *J. Chem. Inf. Model.*, 45(5), 1447-1455.
- [22]. R. Gieleciak, J. Polanski, 2007, Modeling robust QSAR. 2. Iterative variable elimination schemes for CoMSA: application for modeling benzoic acid pKa values. *J. Chem. Inf. Model.*, 47, 547-556.
- [23]. S. Jónsdóttir, F. Jørgensen, B. Søren, 2005, Prediction methods and databases within chemoinformatics: emphasis on drugs and drug candidates. *Bioinformatics*, 21, 2145-2160.
- [24]. J. Burbaum, N. Sigal, 1997, New technologies for high-throughput screening. *Curr. Opin. Chem. Biol.*, 1, 72-78.
- [25]. D. Flower, 1998, On the properties of bit-string based measures of chemical similarity. *J. Chem. Inf. Comput. Sci.*, 38, 379-386.
- [26]. M. Hann, T. Oprea, 2004, Pursuing the leadlikeness concept in pharmaceutical research. *Curr. Opin. Chem. Biol.*, 8, 255-263.
- [27]. J.L. Medina-Franco, K. Martinez-Mayorga, M.A. Giulianotti, R.A. Houghten, C. Pinilla, 2008, Visualization of the Chemical Space in Drug Discovery. *Curr. Comput-Aided Drug Design*, 4, 322-333.
- [28]. D.P. Yang, H.F. Ji, G.Y. Tang, W. Ren, H.Y. Zhang, 2007, How Many Drugs Are Catecholics? *Molecules*, 12, 878-884.
- [29]. S. Sarkhel, G.R. Desiraju, 2004, N-H...O, O-H...O, C-H...O hydrogen bonds in protein-ligand complexes: strong and weak interactions in molecular recognition. *Proteins*, 54, 247-259.
- [30]. L. Tchertanov, J.F. Mouscadet, 2007, Target Recognition by Catechols and β -Ketoenols: Potential Contribution of Hydrogen Bonding and Mn/Mg Chelation to HIV-1 Integrase Inhibition. *J. Med. Chem.*, 50, 1133-1146.
- [31]. T. Scior, P. Bernard, J.L. Medina-Franco, G.M. Maggiora, 2007, Large compound databases for structure-activity relationships studies in drug discovery. *Mini Rev. Med. Chem.*, 7, 851-860.
- [32]. L. Infantes, W.D. Motherwell, 2004, The probable number of hydrogen-bonded contacts for chemical groups in organic crystal structures. *Chem. Commun.*, 1166-1167.
- [33]. R. Dayam, R. Sanchez, N. Neamati, 2005, Diketo acid pharmacophore. 2. Discovery of structurally diverse inhibitors of HIV-1 Integrase, *J. Med. Chem.*, 48, 8009-8015.
- [34]. S. Hare, S.S. Gupta, E. Volkov, A. Engelman, P. Cherepanov, 2010, Retroviral intasome assembly and inhibition of DNA strand transfer, *Nature*, 464, 232-236.
- [35]. J. Polanski, 2009, Receptor dependent multidimensional QSAR for modeling drug-receptor interactions, *Curr. Med. Chem.*, 16, 3243-3257.
- [36]. S. Hare, A.M. Vos, R.F. Clayton, J.W. Thuring, M.D. Cummings, P. Cherepanov, 2010, Molecular mechanisms of retroviral integrase inhibition and the evolution of viral resistance, *Proc. Natl. Acad. Sci.*, 107, 20057-20062.
- [37]. D.C. Rees, M. Congreve, C.W. Murray, R. Carr, 2004, Fragment-based lead discovery, *Nat. Rev. Drug Discov.*, 3, 660-672.
- [38]. P.J. Hajduk, J. Greer, 2007, A decade of fragment-based drug design: strategic advances and lessons learned, *Nat. Rev. Drug. Discov.* 6, 211-219.
- [39]. M. Fialkowski, K.J.M. Bishop, V.A. Chubukov, C.J. Campbell, B.A. Grzybowski, 2005, Architecture and evolution of organic chemistry. *Angew. Chem. Int. Ed. Engl.*, 44, 7263-7269.
- [40]. K. Grabowski, G. Schneider, 2007, Properties and architecture of drugs and natural products revisited. *Curr. Chem. Biol.*, 1, 115-127.
- [41]. V.R. Solomon, H. Lee, 2011, Quinoline as a privileged scaffold in cancer drug discovery. *Curr. Med. Chem.*, 18, 1488-1508.

- [42]. Y. Lingxiang, Y. Zang, 2001, Total intracellular accumulation levels of dietary isothiocyanates determines their activity in elevation of cellular glutathione and induction of Phase 2 detoxification enzymes. *Carcinogenesis*, 22, 1987–1992.
- [43]. J.W. Chiao, F.L. Chung, R. Kancherla, T. Ahmed, A. Mittelman, C.C Conaway, 2002, Sulforaphane and its metabolite mediate growth arrest and apoptosis in human prostate cancer cells. *Int. J. Oncol.*, 20, 631–636.
- [44]. P. Reigan, M.A. Colucci, D. Siegel, A. Chilloux, C.J. Moody, D. Ross, 2007, Development of indolequinone mechanism-based inhibitors of NAD(P)H:quinone oxidoreductase 1 (NQO1): NQO1 inhibition and growth inhibitory activity in human pancreatic MIA PaCa-2 cancer cells. *Biochemistry*, 46, 5941–5950.
- [45]. T. Magdziarz, P. Mazur, J. Polanski, 2009, Receptor independent and receptor dependent CoMSA modeling with IVE-PLS: application to CBG benchmark steroids and reductase activators. *J. Mol. Model.*, 15, 41–51.
- [46]. J. Fraczyk, B. Malawska, Z.J. Kaminski, 2009, Application of a library of artificial receptors formed by the self-organization of N-lipidated peptides immobilized on cellulose in studying the effects of the incorporation of a fluorine atom. *J. Comb. Chem.*, 11, 446–451.
- [47]. J. Fraczyk, B. Kolesinska, Z.J. Kaminski, 2012, The new approach to organocatalysts. Synthesis of a library of N-lipidated oligopeptides immobilized on cellulose and screening of their catalytic activity. *ARKIVOC*, 4, 186–195.
- [48]. J. Fraczyk, Z.J. Kaminski, 2008, Design, synthesis and application of a library of supramolecular structures formed by N-lipidated peptides immobilized on cellulose. *J. Comb. Chem.*, 10, 934–940.
- [49]. M. Daszykowski, B. Walczak, 2011, Methods for the exploratory analysis of two-dimensional chromatographic signals. *Talanta*, 83, 1088–1097.
- [50]. O.A. Santos-Filho, A. J. Hopfinger, A. Cherkasov, R.B. de Alencastro, 2009, The receptor-dependent QSAR paradigm: an overview of the current state of art. *Med. Chem.*, 5, 359–566.
- [51]. S. Funar-Timofei, W.M.F. Fabian, L. Kurunczi, M. Goodarzi, S.T. Ali, Y. Vander Heyden, 2012, Modelling heterocyclic azo dye affinities for cellulose fibres by computational approaches. *Dyes Pigm.*, 94, 278–289.
- [52]. J. Polanski, **A. Bak**, 2003, Modeling Steric and Electronic Effects in 3D- and 4D-QSAR Schemes: Predicting Benzoic pKa Values and Steroid CBG Binding Affinities, *J. Chem. Inf. Comput. Sci.*, 43(6), 2081–2092.
- [53]. T. Magdziarz, B. Lozowicka, R. Gieleciak, **A. Bak**, J. Polanski, Z. Chilmonczyk, 2006, 3D QSAR study of hypolipidemic asarones by comparative molecular surface analysis, *Bioorg. Med. Chem.*, 14, 273–279.
- [54]. J. Polanski, **A. Bak**, R. Gieleciak, T. Magdziarz, 2004, Self-organizing Neural Networks for Modeling Robust 3D and 4D QSAR: Application to Dihydrofolate Reductase Inhibitors, *Molecules*, 9(12), 1148–1159.
- [55]. **A. Bak**, J. Polanski, 2006, A 4D-QSAR study on anti-HIV HEPT analogues, *Bioorg. Med. Chem.*, 14, 273–279.
- [56]. <http://www.chemoinformatyka.us.edu.pl/e-chem/portal.html>
- [57]. <http://icse.us.edu.pl/>
- [58]. J. Polanski, **A. Bak**, 2010, Podstawy chemoinformatyki leków, Oficyna Wydawnicza, Katowice, ISBN 978-83-60743-30-0
- [59]. J. Bogocz, **A. Bak**, J. Polanski, 2014, No free lunches in Nature? An analysis of the regional distribution of the affiliations of Nature publications, *Scientometrics*, (w druku).
- [60]. Ewangelia Św. Mateusza.

5. Wykaz innych (nie wchodzą w skład osiągnięcia naukowego wymienionego w punkcie 4B) opublikowanych prac naukowych oraz wskaźniki dokonań naukowych

5A. Publikacje naukowe w czasopismach znajdujących się w bazie Journal Citation Reports (JRC)

Publikacje poza cyklem habilitacyjnym przed doktoratem

- [1]. J. Polanski, R. Gieleciak, **A. Bak**, **2002**, The Comparative Molecular Surface Analysis (CoMSA) - A Nongrid 3D QSAR Method by a Coupled Neural Network and PLS System: Predicting pK_a Values of Benzoic and Alkanoic Acids, *J. Chem. Inf. Comput. Sci.*, 42(2), 184-191. IF₂₀₀₂=2,8 MSWiA₂₀₀₂=32 (IF₂₀₁₃=4,3 MSWiA₂₀₁₃=40 L. cyt.=27)
Mój wkład polegał na komputerowej implementacji oraz praktycznym przeprowadzeniu części symulacji, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu. Mój udział procentowy szacuję na 30%.
- [2]. J. Polanski, R. Gieleciak, **A. Bak**, K. Jarzembek, M. Wyszomirski, **2002**, The Comparative Molecular Surface Analysis (CoMSA). A New Efficient Technique for Drug Design, *Acta Pol. Pharm.*, 59(6), 459-461. (IF₂₀₁₃=0,7 MSWiA₂₀₁₃=15 L. cyt.=5)
Mój wkład polegał na komputerowej implementacji oraz praktycznym przeprowadzeniu części symulacji, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu. Mój udział procentowy szacuję na 30%.
- [3]. J. Polanski, **A. Bak**, **2003**, Modeling Steric and Electronic Effects in 3D- and 4D-QSAR Schemes: Predicting Benzoic pK_a Values and Steroid CBG Binding Affinities, *J. Chem. Inf. Comput. Sci.*, 43(6), 2081-2092. IF₂₀₀₃=2,8 MSWiA₂₀₀₃=32 (IF₂₀₁₃=4,3 MSWiA₂₀₁₃=40 L. cyt.=18)
Mój wkład polegał na opracowaniu koncepcji badań, komputerowej implementacji oraz praktycznym przeprowadzeniu symulacji, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu. Mój udział procentowy szacuję na 50%.
- [4]. J. Polanski, R. Gieleciak, T. Magdziarz, **A. Bak**, **2004**, The GRID Formalism for the Comparative Molecular Surface Analysis: Application to the CoMFA Benchmark Steroids, Azo dyes and HEPT derivatives, *J. Chem. Inf. Comput. Sci.*, 44(4), 1423-1435. IF₂₀₀₄=2,8 MSWiA₂₀₀₄=32 (IF₂₀₁₃=4,3 MSWiA₂₀₁₃=40 L. cyt.=19)
Mój wkład polegał na opracowaniu koncepcji badań, komputerowej implementacji oraz praktycznym przeprowadzeniu symulacji, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu, końcowym przygotowaniu manuskryptu do druku. Mój udział procentowy szacuję na 30%.
- [5]. J. Polanski, R. Gieleciak, **A. Bak**, **2004**, Probability Issues in Molecular Design. Predictive and modeling ability in 3D-QSAR schemes, *Comb. Chem. High Throughput Screen.*, 7(8), 793-807. IF₂₀₀₄=2,1 MSWiA₂₀₀₄=32 (IF₂₀₁₃=2,0 MSWiA₂₀₁₃=25 L. cyt.=16)

Mój wkład polegał na komputerowej implementacji oraz praktycznym przeprowadzeniu części symulacji, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu. Mój udział procentowy szacuję na 30%.

- [6]. J. Polanski, **A. Bak**, R. Gieleciak, T. Magdziarz, **2004**, Self-organizing Neural Networks for Modeling Robust 3D and 4D QSAR: Application to Dihydrofolate Reductase Inhibitors, *Molecules*, 9(12), 1148-1159. IF₂₀₀₄=0,7 MSWiA₂₀₁₂=15 (IF₂₀₁₃=2,4 MSWiA₂₀₁₃=30 L. cyt.=11)

Mój wkład polegał na opracowaniu koncepcji badań, komputerowej implementacji oraz praktycznym przeprowadzeniu symulacji, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu, odpowiedzi na uwagi recenzentów, końcowym przygotowaniu manuskryptu do druku. Mój udział procentowy szacuję na 40%.

Łączny IF₂₀₀₂₋₂₀₀₄=**11,2** MSWiA₂₀₀₂₋₂₀₀₄=**143** (IF₂₀₁₃=**18,0** MSWiA₂₀₁₃=**190**)

Publikacje poza cyklem habilitacyjnym po doktoracie

- [1]. R. Gieleciak, T. Magdziarz, **A. Bak**, J. Polanski, **2005**, Modeling Robust QSAR. 1. Coding Molecules in 3D-QSAR - from a Point to Surface Sectors and Molecular Volumes, *J. Chem. Inf. Model.*, 45(5), 1447-1455. IF₂₀₀₅=2,9 MSWiA₂₀₀₅=32 (IF₂₀₁₃=4,3 MSWiA₂₀₁₃=40 L. cyt.=5)

Mój wkład polegał na komputerowej implementacji oraz praktycznym przeprowadzeniu części symulacji, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu. Mój udział procentowy szacuję na 30%.

- [2]. T. Magdziarz, B. Lozowicka, R. Gieleciak, **A. Bak**, J. Polanski, Z. Chilmonczyk, **2006**, 3D QSAR study of hypolipidemic asarones by comparative molecular surface analysis, *Bioorg. Med. Chem.*, 14, 1630-1643. IF₂₀₀₆=2,3 MSWiA₂₀₀₆=32 (IF₂₀₁₃=2,9 MSWiA₂₀₁₃=30 L. cyt.=6)

Mój wkład polegał na komputerowej implementacji oraz praktycznym przeprowadzeniu części symulacji, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu. Mój udział procentowy szacuję na 20%.

- [3]. **A. Bak**, J. Polanski, **2006**, A 4D-QSAR study on anti-HIV HEPT analogues, *Bioorg. Med. Chem.*, 14, 273-279. IF₂₀₀₆=2,3 MSWiA₂₀₀₆=32 (IF₂₀₁₃=2,9 MSWiA₂₀₁₃=30 L. cyt.=26)

Mój wkład polegał na opracowaniu koncepcji badań, komputerowej implementacji oraz praktycznym przeprowadzeniu symulacji, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu, odpowiedzi na uwagi recenzentów, końcowym przygotowaniu manuskryptu do druku. Mój udział procentowy szacuję na 80%.

- [4]. J. Polanski, **A. Bak**, R. Gieleciak, T. Magdziarz, **2006**, Modeling robust QSAR, *J. Chem. Inf. Model.*, 46(6), 2310-2318. IF₂₀₀₆=3,4 MSWiA₂₀₀₆=32 (IF₂₀₁₃=4,3 MSWiA₂₀₁₃=40 L. cyt.=36)
Mój wkład polegał na opracowaniu koncepcji badań, komputerowej implementacji oraz praktycznym przeprowadzeniu symulacji, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu, odpowiedzi na uwagi recenzentów, końcowym przygotowaniu manuskryptu do druku. Mój udział procentowy szacuję na 40%.
- [5]. H. Niedbala, J. Polanski, R. Gieleciak, R. Musiol, D. Tabak, B. Podeszwa, **A. Bak**, A. Palka, J.-F. Mouscadet, J. Gasteiger, M. Le Bret, **2006**, Comparative molecular surface analysis (CoMSA) for virtual combinatorial library screening of styrylquinoline HIV-1 blocking agents, *Comb. Chem. High Throughput Screen.*, 9, 753-770. IF₂₀₀₆=2,5 MSWiA₂₀₀₆=32 (IF₂₀₁₃=2,0 MSWiA₂₀₁₃=25 L. cyt.=8)
Mój wkład polegał na komputerowej implementacji oraz praktycznym przeprowadzeniu części symulacji, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu. Mój udział procentowy szacuję na 20%.
- [6]. A. Smolinski, P. Kuna-Gwozdziwicz, M. Jacek, **A. Bak**, **2014**, Study of the polycyclic aromatic hydrocarbons content in gas released from burning mine waste dump, *Acta Chromatographica*, (w druku), IF₂₀₁₃=0,6 MSWiA₂₀₁₃=20
Mój wkład polegał na komputerowej implementacji oraz praktycznym przeprowadzeniu części symulacji, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu. Mój udział procentowy szacuję na 20%.
- [7]. J. Bogocz, **A. Bak**, J. Polanski, **2014**, No free lunches in Nature? An analysis of the regional distribution of the affiliations of Nature publications, *Scientometrics*, (w druku), IF₂₀₁₃=2,1 MSWiA₂₀₁₃=35
Mój wkład polegał na opracowaniu koncepcji badań, komputerowej implementacji oraz praktycznym przeprowadzeniu symulacji, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu, odpowiedzi na uwagi recenzentów, końcowym przygotowaniu manuskryptu do druku. Mój udział procentowy szacuję na 50%.

Łączny IF₂₀₀₅₋₂₀₁₃=**16,1** MSWiA₂₀₀₅₋₂₀₁₃=**215** (IF₂₀₁₃=**19,1** MSWiA₂₀₁₃=**220**)

Podsumowanie:

Całkowity IF₂₀₀₂₋₂₀₁₃=**55,3** MSWiA₂₀₀₂₋₂₀₁₃=**697** (IF₂₀₁₃=**65,4** MSWiA₂₀₁₃=**740**)

5B. Wynalazki oraz wzory użytkowe i przemysłowe, które uzyskały ochronę i zostały wystawione na międzynarodowych lub krajowych wystawach i targach
 – nie dotyczy

5C. Monografie, publikacje naukowe w czasopismach międzynarodowych lub krajowych innych niż znajdujące się w bazie, o której mowa w punkcie 5A

- [1] J. Polanski, R. Gieleciak, **A. Bak**, **2003**, The comparative molecular surface analysis (CoMSA): Theoretical advantages and practical application in drug design, *Pol. J. Pharmacol.*, **55**, 508.

Mój wkład polegał na komputerowej implementacji oraz praktycznym przeprowadzeniu części symulacji, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu. Mój udział procentowy szacuję na 30%.

- [2] **A. Bak**, R. Gieleciak, T. Magdziarz, J. Polanski, **2006**, Self-organizing neural networks for modeling robust 3D and 4D-QSAR: Application to dihydrofolate reductase (DHR) inhibitors, *Comput. Aided Drug Des. Dev. Soc. (Turkey)*, *QSAR and molecular modeling in rational design of bioactive molecules*, eds. E. Aki(Sener) and I. Yalcin, pp.121-122, ISBN 975-00782-0-9

Mój wkład polegał na opracowaniu koncepcji badań, komputerowej implementacji oraz praktycznym przeprowadzeniu symulacji, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu, odpowiedzi na uwagi recenzentów, końcowym przygotowaniu manuskryptu do druku. Mój udział procentowy szacuję na 80%.

- [3] **A. Bak**, R. Gieleciak, T. Magdziarz, J. Polanski, **2006**, Application of the self-organizing neural networks for modeling steric and electronic effects in 4D-QSAR, *Comput. Aided Drug Des. Dev. Soc. (Turkey)*, *QSAR and molecular modeling in rational design of bioactive molecules*, eds. E. Aki(Sener) and I. Yalcin, pp.119-120, ISBN 975-00782-0-9

Mój wkład polegał na opracowaniu koncepcji badań, komputerowej implementacji oraz praktycznym przeprowadzeniu symulacji, wstępnym opracowaniu wyników, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu, odpowiedzi na uwagi recenzentów, końcowym przygotowaniu manuskryptu do druku. Mój udział procentowy szacuję na 80%.

- [4]. J. Polanski, **A. Bak**, **2010**, Podstawy chemoinformatyki leków, Oficyna Wydawnicza, Katowice, ISBN 978-83-60743-30-0

Mój wkład polegał na opracowaniu koncepcji manuskryptu, wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu, odpowiedzi na uwagi recenzentów, końcowym przygotowaniu manuskryptu do druku. Mój udział procentowy szacuję na 50%.

- [5]. A. Kurczyk, **A. Bak**, T. Magdziarz, P. Mazur, J. Polanski, **2010**, Polifarmakologia i Lekotypia Związków Azanaftalenowych, Na Pograniczu Chemii i Biologii, Wydawnictwo Naukowe UAM, Poznań, ISBN 978-83-232-2254-5

Mój wkład polegał na wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu, końcowym przygotowaniu manuskryptu do druku. Mój udział procentowy szacuję na 30%.

5D. Opracowanie zbiorowe, katalogi zbiorów, dokumentacja prac badawczych, ekspertyz, utworów i dzieł artystycznych

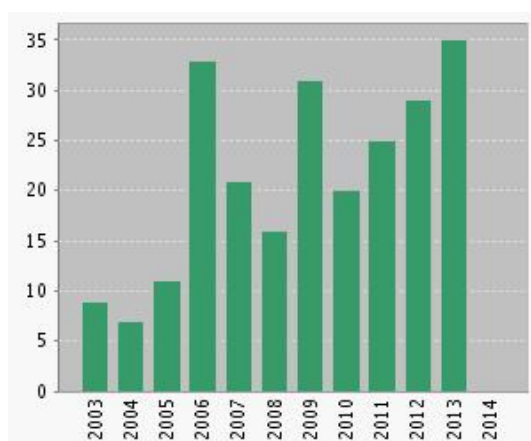
[1]. Laboratorium projektowania molekularnego. Materiały do ćwiczeń. Praca zbiorowa pod redakcją A. Bąk, Katowice, 2010.

Mój wkład polegał na wstępnym oraz ostatecznym przygotowaniu tekstu manuskryptu, końcowym przygotowaniu manuskryptu do druku. Mój udział procentowy szacuję na 60%.

5E. Sumaryczny *impact factor* według listy Journal Citation Reports (JCR): **65.4**

5F. Liczba cytowań publikacji według bazy Web of Science (WoS): **237**

bez autocytowań **175**



5G. Indeks Hirscha według bazy Web of Science (WoS): **9**

5H. Kierowanie międzynarodowymi i krajowymi projektami badawczymi oraz udział w takich projektach

a. finansowanych ze środków zewnętrznych

1. KBN 4T09A 088 25, 2005, kierownik projektu
2. 3T09A 01127, grant zespołowy, wykonawca projektu
3. N N519 575638, grant zespołowy, wykonawca projektu
4. 1787B/PO1/2008/35 grant zespołowy, wykonawca projektu

b. finansowanych ze środków wewnętrznych (uczelnia, wydział):

1. Dotacja z Rezerwy Rektora 2010, Wydział Mat. Fiz. Chem. UŚ, kierownik projektu
2. Dotacja dla Młodych Naukowców 2011, Wydział Mat. Fiz. Chem. UŚ, kierownik projektu
3. Projekt UPGOW, 2012, Wydział Mat. Fiz. Chem. UŚ, wykonawca projektu
4. Projekt iCSE, 2012, Wydział Mat. Fiz. Chem. UŚ, wykonawca projektu

5I. Międzynarodowe i krajowe nagrody za działalność naukową albo artystyczną

1. Stypendium krajowe dla młodych uczonych, program START, Fundacja na Rzecz Nauki Polskiej, 2006
2. Stypendium krajowe dla młodych uczonych, program START, Fundacja na Rzecz Nauki Polskiej, 2007

5J. Wygłoszenie referatów na międzynarodowych i krajowych konferencjach tematycznych

1. **Bak A.**, Magdziarz T., Kurczyk A., Polanski J., Mapping Drug Architecture by MoStBioDat – Rapid Screening of Catechol Motifs, The 33-th Symposium “Chromatographic Methods of Investigating the Organic Compounds”, **2010**, Szczyrk, wykład na zaproszenie
2. **Bak A.**, Pharmacophore-Based Virtual Screening in Drug Design, Seminarios de Investigation, **2012**, Salamanka, Hiszpania, wystąpienie

6. Dorobek dydaktyczny i popularyzatorski oraz informacja o współpracy międzynarodowej habilitanta

6A. Uczestnictwo w programach europejskich oraz innych programach międzynarodowych i krajowych – brak

6B. Aktywny udział w międzynarodowych i krajowych konferencjach naukowych

1. **Bak A.**, Kucharski S., Częstotliwość harmoniczna i podstawowa dla jonu CN- w oparciu o zaawansowane obliczenia korelacyjne, Katowice, **2001**, PTChem, Polska
2. Ratajczak A., Łysiak V., **Bak A.**, Projektowanie i synteza nowych kwasów arylo-sulfonylocykloalkanokarboksylowych i badanie zależności struktura-aktywność biologiczna, Kraków, **2002**, PTChem, Polska
3. Polański J., Gieleciak R., Wyszomirski M., **Bak A.**, Niedbała H., Musiał R., Zemanem A., Pałka A., Jarzembek K., The Comparative Molecular Surface Analysis (CoMSA) a New 3D QSAR Metod for Drug Design, 1-st Central European Conference „Chemistry Towards Biology”, **2002**, Portoż, Słowenia
4. Polański J., Gieleciak R., **Bak A.**, Wyszomirski M., The Comparative Molecular Surface Analysis (CoMSA) a New Efficient Technique for Drug Design, Trzecia Interdyscyplinarna Konferencja Nauki o Leku (MKNOL), **2002**, Piła, Polska
5. Magdziarz T., Rogoz R., Gieleciak R., **Bak A.**, Polański J., The Grid Version of the Comparative Molecular Surface Analysis. Modeling of Azodye-cellulose Affinities, 2003, Zakopane, Polska
6. **Bak A.**, Magdziarz T., Gieleciak R., Polański J., Self-Organizing Neural Networks (SOM) for Modeling Robust 3D and 4D QSAR. Applications to Dihydrofolate Reductase Inhibitors. The 15-th European Symposium on Quantitative Structure-Activity Relationships and Molecular Modeling (EuroQSAR), **2004**, Istanbul, Turkey
7. **Bak A.**, Magdziarz T., Gieleciak R., Polański J., Applications of the Self-Organizing Neural Networks for Modeling Steric and Electronic Effects in 4D-QSAR Schemes. The 15-th European Symposium on Quantitative Structure-Activity Relationships and Molecular Modeling, **2004**, Istanbul, Turkey
8. Gieleciak R., Magdziarz T., **Bak A.**, Polański J., The Comparative Molecular Surface Analysis: New Formalizm and Application in 3D QSAR Studies. The 15-th European

- Symposium on Quantitative Structure-Activity Relationships and Molecular Modeling, **2004**, Istanbul, Turkey
9. Stockner T., **Bąk A.**, Klebe G., Jakusch M., Towards an Automated Procedure to Model Estrogen Receptor-Ligand Complexes Based on the MOBILE Approach, 15-th International Conference on Intelligent Systems for Molecular Biology and 6-th European Conference on Computational Biology (ISMB/ECCB), **2007**, Vienna, Austria
 10. Polański J., Musioł R., Serafin K., Niedbała H., Finster J., Mazur P., Magdziarz T., **Bąk A.**, Design and synthesis of HIV integrase inhibitors, 38th Conference Synteza i Analiza Leciwników, **2009**, Hradec Kralove, Czech Republic
 11. Kurczyk A., **Bąk A.**, Magdziarz T., Polański J., MoStBioDat – Molecular and Structural Bioinformatics Database, **2009**, Gliwice, Polska
 12. **Bąk A.**, Magdziarz T., Kurczyk A., Polański J., Mapping Drug Architecture by MoStBioDat – Rapid Screening of Intramolecular Hydrogen Bonded Motifs in Catechols, 19-th European Symposium on Quantitative Structure-Activity Relationships and Molecular Modeling, **2010**, Rhodes, Greece
 13. Kurczyk A., Mazur P., Janik B., **Bąk A.**, Magdziarz T., Polanski J., Mining databases for the analysis of catechol and azanaphthalene polypharmacology, 5-th Central European Conference “Chemistry toward Biology”, 2010, Primosten, Croatia
 14. **Bąk A.**, Magdziarz T., Kurczyk A., Polański J., Mapping Fragmental Drug-likeness in the MoStBioDat Environment – Intramolecular Hydrogen Bonding in β -Ketoenols, The 34-th Symposium “Chromatographic Methods of Investigating the Organic Compounds”, 2011, Szczyrk, Polska
 15. Kurczyk A., **Bąk A.**, Magdziarz T., Polański J., Knowledge Discovery in Molecular and Structural Chemical Databases, The 34-th Symposium “Chromatographic Methods of Investigating the Organic Compounds”, 2011, Szczyrk, Polska
 16. **Bąk A.**, Magdziarz T., Kurczyk A., Jarzembek K., Kozik V., Polanski J., Mining Chemical Databases for Mapping Fragmental Drug-Likeness in MoStBioDat, 6-th International Symposium on Computational Methods in Toxicology and Pharmacology Integrating Internet Resources, 2011, Maribor, Slovenia
 17. **Bąk A.**, Jarzembek K., Kozik V., Magdziarz T., Polanski J., Pharmacophore-Based Database Mining for Probing Fragmental Drug-Likeness of Diketo Acid Analogues, The 35-th Symposium “Chromatographic Methods of Investigating the Organic Compounds”, **2012**, Szczyrk, Polska
 18. **Bąk A.**, Magdziarz T., Polanski J., Kozik V., Jarzembek K., Mapping Fragmental Drug-Likeness of Diketo Acid Analogues Using Pharmacophore-Based Database

- Miting, The 19-th EuroQSAR – Knowledge Enabling Ligand Design, 2012, Vienna, Austria
19. Polański J., **Bąk A.**, Mining Molecular Databases for Fragonomics Based Exploration of the Architecture of Drugs, Modeling and Design of Molecular Materials **2012**, Wrocław, Poland
 20. **Bąk A.**, Daszykowski M., Kozik V., Jarzembek K., Kamiński Z., Kiec-Kononowicz K., Kuder K., Fraczyk J., Kolesinska B., Ciosek P., Polanski J., Probing an Artificial Polypeptide Receptor Library Rusing a Serie sof Novel Histamine H3 Receptor Ligands, The 36-th Symposium “Chromatographic Methods of Investigating the Organic Compounds”, 2013, Szczyrk, Polska
 21. Matussek M., Kozik V., Jarzembek K., Pytlakowska K., Michalik S., **Bąk A.**, Rojkiewicz M., Synthesis new terephtalamides and structure characterization, The 36-th Symposium “Chromatographic Methods of Investigating the Organic Compounds”, 2013, Szczyrk, Polska
- 6C. Udział w komitetach organizacyjnych międzynarodowych konferencji i krajowych konferencji naukowych
1. Chemistry towards Biology, 2014, Katowice, Polska, organizator konferencji
- 6D. Otrzymane nagrody i wyróżnienia inne niż wymienione w punkcie 5I – brak
- 6E. Udział w konsorcjach i sieciach badawczych - brak
- 6F. Kierowanie projektami realizowanymi we współpracy z naukowcami z innych ośrodków polskich i zagranicznych oraz we współpracy z przedsiębiorcami, innymi niż wymienione w punkcie 5H.
1. Projekt ekoStaż 22/2013 realizowany we współpracy z firmą i-Petrol w Katowicach
- 6G. Udział w komitetach redakcyjnych i radach naukowych czasopism – brak
- 6H. Członkostwo w międzynarodowych i krajowych organizacjach oraz towarzystwach naukowych
1. Członek Komisji Nauk Chemicznych PAN o. Katowice
- 6I. Osiągnięcia dydaktyczne i w zakresie popularyzacji nauki i sztuki
1. Ekspert z zakresu chemii w ramach projektu „Z angielskim do źródeł wiedzy” organizowanym przez II LO Rybnik oraz Towarzystwo Przyjaciół UŚ, 2009-2010
 2. Materiały dydaktyczne w ramach projektu UPGOW, 2010
 3. Materiały dydaktyczne w ramach projektu iCSE realizowanego we współpracy z Uniwersytetem w Oslo, 2012-2013
 4. Zajęcia laboratoryjne dla szkół gimnazjalnych, 2013
 5. Wykłady popularno-naukowe dla szkół średnich w ramach projektu Uniwersytet Śląski Maturzystów, 2014

6. Współpraca w ramach projektu „Praktyczny kurs nowoczesnych metod analizy instrumentalnej”, 2010-2011
- 6J. Opieka naukowa nad studentami i lekarzami w toku specjalizacji
 1. Prace magisterskie, 2010-2014, Uniwersytet Śląski, 8 osób
 2. Prace licencjackie, 2013, Uniwersytet Śląski, 1 osoba
 3. Zajęcia (seminaria/wykłady) na studiach I i II stopnia z następujących przedmiotów: laboratorium chemii organicznej, wstęp do projektowania leków, chemoinformatyka, bioinformatyka, technologie informatyczne, sieci neuronowe
 4. Opiekun studentów V roku, 2013-2014
- 6K. Opieka naukowa nad doktorantami w charakterze opiekuna naukowego lub promotora pomocniczego – brak
- 6L. Staże w zagranicznych i krajowych ośrodkach naukowych lub akademickich
 1. **15.12.2002–22.12.2002** Computer-Chemie-Centrum, University of Erlangen, Erlangen, Niemcy, staż naukowy
 2. **01.02.2007–31.12.2008** Austrian Institute of Technology, Seibersdorf, Austria, staż podoktorski
 3. **01.02.2012–01.03.2012** Biochemistry Department, University of Oxford, Oxford, Anglia, staż naukowy
 4. **02.09.2012–09.09.2012** Instytut Fizyczno-Chemiczny, Ukraińska Akademia Nauk, Odessa, Ukraina, staż naukowy
 5. **17.10.2012–24.10.2012** Wydział Farmaceutyczny, University of Salamanca, Hiszpania, staż naukowy
- 6M. Wykonane ekspertyzy lub inne opracowania na zamówienie
 1. Konsultacja słownika wydawnictwa TERKA w zakresie chemii, Bielsko-Biała, 2010
- 6N. Udział w zespołach eksperckich i konkursowych
 1. Ekspert z zakresu chemii w ramach projektu „Z angielskim do źródeł wiedzy” organizowanym przez II LO Rybnik oraz Towarzystwo Przyjaciół UŚ, 2009-2010
- 6O. Recenzowanie projektów międzynarodowych i krajowych – brak
- 6P. Recenzowanie publikacji w czasopismach międzynarodowych i krajowych
 1. Chemometric and Intelligent Laboratory Systems, 2010, 1 publikacja
 2. Journal of Molecular Modeling, 2010, 1 publikacja
 3. SAR&QSAR in Environmental Research, 2014, 1 publikacja
- 6Q. Inne osiągnięcia, nie wymienione w punkcie 6A-P.
 - a. Certyfikaty
 1. First Certificate in English (FCE), Cambridge English
 2. Certificate in Advanced English (CAE), Cambridge English

Podsumowanie dorobku naukowo-dydaktycznego

Liczba wszystkich publikacji **32** (po doktoracie **25**)

Liczba publikacji wchodzących w skład habilitacji **12** (w **9** habilitant jest pierwszym autorem)

Monografie lub rozdziały w monografiach **2**

Artykuły konferencyjne (pełnotekstowe) **2**

Udział w konferencjach krajowych i międzynarodowych **21**

Recenzje dla czasopism: **3**

Summary IF₂₀₀₂₋₂₀₁₃=**55,3** (IF₂₀₁₃=**65,4**)

Summary IF publikacji wchodzących w skład habilitacji IF₂₀₀₇₋₂₀₁₃=**28** (IF₂₀₁₃=**28,3**)

Summary IF pozostałych publikacji (przed doktoratem IF₂₀₀₂₋₂₀₀₄=**11,2** IF₂₀₁₃=**18**)

IF₂₀₀₂₋₂₀₁₃=**27,3** (IF₂₀₁₃=**37,1**)

Liczba cytowań publikacji według bazy Web of Science (WoS) bez autocytowań **175**

Indeks Hirscha według bazy Web of Science (WoS) **9**

Zgodnie z kryteriami MNiSW na rok 2013 dorobek naukowy oceniany jest na **740** punktów