

# **Autoreferat**

**Opracowanie narzędzi obliczeniowych opartych o metody uczenia  
maszynowego i ich wykorzystanie w komputerowo wspomaganym  
projektowaniu leków**

**Dr Sabina Podlewska**



**Instytut Farmakologii  
im. Jerzego Maja  
Polskiej Akademii Nauk**

**Kraków 2022**

## Spis treści

1. Imię i nazwisko .....	3
2. Posiadane dyplomy, stopnie naukowe .....	3
3. Informacja o dotychczasowym zatrudnieniu w jednostkach naukowych .....	3
4. Omówienie osiągnięć, o których mowa w art. 219 ust. 1 pkt. 2 Ustawy .....	4
4.1. Tytuł osiągnięcia naukowego .....	4
4.2. Zestawienie publikacji wchodzących w skład osiągnięcia naukowego.....	4
4.3. Opis wkładu własnego w publikacje wchodzące w skład osiągnięcia naukowego .....	6
4.4. Omówienie celu naukowego przedstawionego osiągnięcia naukowego oraz otrzymanych wyników .....	11
4.4.1. Wprowadzenie .....	11
4.4.2. Cel badawczy .....	16
4.4.3. Omówienie wyników badań.....	17
4.4.4. Podsumowanie .....	44
4.4.5. Bibliografia.....	45
5. Informacja o wykazywaniu się istotną aktywnością naukową albo artystyczną realizowaną w więcej niż jednej uczelni, instytucji naukowej lub instytucji kultury, w szczególności zagranicznej.....	48
5.1. Aktywność naukowa przed uzyskaniem stopnia naukowego doktora.....	48
5.2. Aktywność naukowa po uzyskaniu stopnia naukowego doktora.....	51
6. Informacja o osiągnięciach dydaktycznych, organizacyjnych oraz popularyzujących naukę lub sztukę .....	57
6.1. Działalność dydaktyczna.....	57
6.2. Działalność popularyzacyjna .....	58
6.3. Działalność organizacyjna .....	58
7. Pozostałe informacje .....	58
7.1. Udział w projektach badawczych .....	58
7.1.1. Kierowanie projektami badawczymi.....	58
7.1.2. Uczestnictwo w projektach badawczych w charakterze wykonawcy .....	59
7.2. Nagrody i wyróżnienia .....	60

## 1. Imię i nazwisko

Sabina Podlewska (nazwisko panińskie: Smusz, zmiana nazwiska w 2015 r.)

## 2. Posiadane dyplomy, stopnie naukowe

- 13.10.2016** doktor nauk chemicznych  
Wydział Chemii Uniwersytetu Jagiellońskiego  
**Tytuł rozprawy doktorskiej:** "Development of machine learning-based tools for computer-aided drug design"  
**Promotor:** prof. dr hab. Andrzej J. Bojarski  
Rozprawie doktorskiej przyznano wyróżnienie
- 03.11.2014** magister matematyki  
Wydział Matematyki, Fizyki i Chemii Uniwersytetu Śląskiego  
**Tytuł pracy magisterskiej:** „Modele Markowa w procedurach projektowania substancji aktywnych biologicznie”  
**Promotor:** dr hab. Marta Tyran-Kamińska
- 19.06.2012** magister chemii  
Wydział Chemii Uniwersytetu Jagiellońskiego  
**Tytuł pracy magisterskiej:** „Ewaluacja różnych metod uczenia maszynowego pod kątem ich zastosowania w wirtualnym skriningu”  
**Promotor:** dr hab. Andrzej J. Bojarski, prof. IF

## 3. Informacja o dotychczasowym zatrudnieniu w jednostkach naukowych

- 01.04.2015–obecnie** Instytut Farmakologii im. Jerzego Maja PAN, Zakład Chemii Leków, stanowisko: asystent (21.12.2015–01.03.2016, 08.10.2018–14.10.2021 – urlop bezpłatny; 11.12.2016–09.12.2017 – urlop macierzyński, rodzicielski)
- 08.10.2018–14.10.2021** Wydział Farmaceutyczny Uniwersytetu Jagiellońskiego Collegium Medicum, Katedra i Zakład Technologii i Biotechnologii Środków Leczniczych, stanowisko: asystent (12.12.2018–10.12.2019 – urlop macierzyński, rodzicielski)
- 01.10.2010–31.03.2015** Instytut Farmakologii PAN, Zakład Chemii Leków, stanowisko: pracownik inżynieryjno-techniczny

**02.08.2010–30.09.2010** Instytut Farmakologii PAN, Zakład Chemii Leków, stanowisko: stażysta

#### 4. Omówienie osiągnięć, o których mowa w art. 219 ust. 1 pkt. 2 Ustawy

##### 4.1. Tytuł osiągnięcia naukowego

Opracowanie narzędzi obliczeniowych opartych o metody uczenia maszynowego i ich wykorzystanie w komputerowo wspomaganym projektowaniu leków.

##### 4.2. Zestawienie publikacji wchodzących w skład osiągnięcia naukowego

Przedstawione osiągnięcie naukowe wynikające z art. 219, ust. 1, pkt. 2b ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce stanowi cykl 11 powiązanych tematycznie artykułów naukowych, z których 10 stanowią prace oryginalne, natomiast 1 artykuł jest publikacją przeglądową. Prace te zostały opublikowane w latach 2017–2022. Sumaryczny Impact Factor (IF) prac składających się na osiągnięcie naukowe wynosi **53,248**, natomiast łączna liczba punktów MEiN wynosi **1130**.

##### Zestawienie publikacji wchodzących w skład osiągnięcia naukowego

Lp.	Dane bibliograficzne publikacji	Punktacja	
		IF <sub>rok_wydania</sub>	MEiN
H1	Podlewska, S.; Czarnecki, W.; Kafel, R.; Bojarski, A. J. Creating the New from the Old: Combinatorial Libraries Generation with Machine-Learning-Based Compound Structure Optimization. <i>J. Chem. Inf. Model.</i> <b>2017</b> , <i>57</i> , 133–147.	3,804	40
H2	Podlewska, S.; Kafel, R.; Lacivita, E.; Satała, G.; Kooistra, A. J.; Vass, M.; de Graaf, C.; Leopoldo, M.; Bojarski, A. J.; Mordalski, S. Structural insights into serotonin receptor ligands polypharmacology. <i>Eur. J. Med. Chem.</i> <b>2018</b> , <i>151</i> , 797–814.	4,833	40
H3	Vass, M. <sup>‡</sup> ; Podlewska, S. <sup>‡</sup> ; de Esch, I. J. P.; Bojarski, A. J.; Leurs, R.; Kooistra, A. J.; de Graaf, C. Aminergic GPCR–Ligand Interactions: A Chemical and Structural Map of Receptor Mutation Data. <i>J. Med. Chem.</i> <b>2019</b> , <i>62</i> , 3784–3839.	6,205	200
H4	Leśniak, D.; Podlewska, S.*; Jastrzębski, S.; Sieradzki, I.; Bojarski, A. J.; Tabor, J. Development of new methods needs proper evaluation – benchmarking sets for machine learning experiments for class A GPCRs. <i>J. Chem. Inf. Model.</i> <b>2019</b> , <i>59</i> , 4974–4992.	4,549	100

<b>H5</b>	Jastrzębski, S.; Szymczak, M.; Pocha, A.; Mordalski, S.; Tabor, J.; Bojarski, A. J.; <b>Podlewska S.*</b> Emulating Docking Results Using a Deep Neural Network: A New Perspective for Virtual Screening. <i>J. Chem. Inf. Model.</i> <b>2020</b> , <i>60</i> , 4246–4262.	<b>4,956</b>	<b>100</b>
<b>H6</b>	<b>Podlewska, S.</b> ; Bugno, R.; Kudla, L.; Bojarski, A. J.; Przewlocki, R. Molecular Modeling of $\mu$ Opioid Receptor Ligands with Various Functional Properties: PZM21, SR-17018, Morphine, and Fentanyl—Simulated Interaction Patterns Confronted with Experimental Data. <i>Molecules</i> <b>2020</b> , <i>25</i> , 4636.	<b>4,411</b>	<b>140</b>
<b>H7</b>	<b>Podlewska, S.</b> , Latacz, G.; Łażewska, D.; Kieć-Kononowicz, K.; Handzlik, J. <i>In silico</i> and <i>in vitro</i> studies on interaction of novel non-imidazole histamine H3R antagonists with CYP3A4. <i>Bioorg. Med. Chem. Lett.</i> <b>2020</b> , <i>30</i> , 127147.	<b>2,823</b>	<b>70</b>
<b>H8</b>	<b>Podlewska, S.*</b> ; Bugno, R.; Lacivita, E.; Leopoldo, M.; Bojarski, A. J.; Handzlik, J.* Low Basicity as a Characteristic for Atypical Ligands of Serotonin Receptor 5-HT <sub>2</sub> . <i>Int. J. Mol. Sci.</i> <b>2021</b> , <i>22</i> , 1035.	<b>5,924</b>	<b>140</b>
<b>H9</b>	Wojtuch, A.; Jankowski, R.; <b>Podlewska, S.*</b> How can SHAP values help to shape metabolic stability of chemical compounds? <i>J. Cheminf.</i> <b>2021</b> , <i>13</i> , 74.	<b>5,514</b>	<b>100</b>
<b>H10</b>	Baltrukевич, H.; <b>Podlewska, S.*</b> From data to knowledge—systematic review of tools for automatic analysis of molecular dynamics output. <i>Front. Pharmacol.</i> <b>2022</b> , <i>13</i> , 844293.	<b>5,811</b>	<b>100</b>
<b>H11</b>	Penna, E.†; Niso, M.; <b>Podlewska, S.†</b> ; Volpicelli, F.; Crispino, M.; Perrone-Capano, C.; Bojarski, A.J.; Lacivita, E.; Leopoldo, M. In vitro and in silico analysis of the residence time of serotonin 5-HT <sub>7</sub> receptor ligands with arylpiperazine structure: a structure-kinetics relationship study. <i>ACS Chemical Neurosci.</i> <b>2022</b> , <i>13</i> , 497–509.	<b>4,418</b>	<b>100</b>

† - równorzędny wkład

\* - autor korespondencyjny

Badania zrealizowane przez habilitantkę, które zostały przedstawione w publikacjach wchodzących w skład osiągnięcia naukowego finansowano z następujących źródeł:

- „Poszukiwanie nowych struktur wiodących dla ligandów receptora serotoninowego 5-HT<sub>7</sub> o zwiększonej stabilności metabolicznej” – Narodowe Centrum Nauki: HARMONIA Nr 2015/18/M/NZ7/00377, wysokość dofinansowania: 243 120 PLN, okres realizacji: 12.04.2016–11.05.2018, charakter udziału habilitantki: kierownik projektu [**H1–H3**]

- „Rozwój grupy niezasadowych ligandów receptora serotoninowego 5-HT<sub>7</sub>, pochodnych 2-(6-okso-3-fenilo-1,6-dihydropirydazyn-1-ylo)-N-feniloacetamidu, o zwiększonej stabilności metabolicznej” – Narodowe Centrum Nauki: SONATINA Nr 2018/28/C/NZ7/00145, wysokość dofinansowania: 790 877 PLN, okres realizacji: 15.10.2018–14.10.2021, charakter udziału habilitantki: kierownik projektu [H4, H7–H9]
- „Opracowanie kompleksowego zestawu narzędzi obliczeniowych do poszukiwania ligandów określonego receptora na przykładzie receptora serotoninowego 5-HT<sub>7</sub>” – Narodowe Centrum Nauki: OPUS Nr 2018/31/B/NZ2/00165, wysokość dofinansowania: 645 200 PLN, okres realizacji: 26.11.2019–obecnie, charakter udziału habilitantki: kierownik projektu [H5, H10]
- “Racjonalne projektowanie i ocena funkcjonalnie stronniczych agonistów receptorów opioidowych” – Narodowe Centrum Nauki: OPUS Nr 2018/31/B/NZ7/03954, wysokość dofinansowania: 2 495 020 PLN, okres realizacji: 24.07.2019–obecnie, kierownik projektu: prof. dr hab. Ryszard Przewłocki, charakter udziału habilitantki: wykonawca [H6]

### 4.3. Opis wkładu własnego w publikacje wchodzące w skład osiągnięcia naukowego

**H1. Podlewska, S.; Czarnecki, W.; Kafel, R.; Bojarski, A. J. Creating the New from the Old: Combinatorial Libraries Generation with Machine-Learning-Based Compound Structure Optimization. *J. Chem. Inf. Model.* **2017**, *57*, 133–147.**

Mój wkład w publikację polegał na realizacji następujących zadań: zaprojektowanie metodologii do generowania nowych związków w oparciu o podstruktury optymalne dla aktywności związków, przygotowanie zestawów danych z bazy ChEMBL dla 8 celów biologicznych, reprezentacja danych w postaci fingerprintów podstrukturalnych, identyfikacja fragmentów strukturalnych występujących w ligandach – elementów wejściowych do generowania nowych związków, opracowanie metodologii łączenia optymalnych fragmentów w nowe potencjalne ligandy danego receptora, weryfikacja skuteczności opracowanego podejścia do generowania nowych związków, analiza wygenerowanych bibliotek związków (przygotowanie rozkładów wybranych własności, analiza podobieństwa do istniejących ligandów), interpretacja uzyskanych wyników, przygotowanie tekstu manuskryptu oraz figur (za wyjątkiem jednej, figury nr 3, której autorem jest Wojciech Czarnecki).

**H2. Podlewska, S.;** Kafel, R.; Lacivita, E.; Satała, G.; Kooistra, A. J.; Vass, M.; de Graaf, C.; Leopoldo, M.; Bojarski, A. J.; Mordalski, S. Structural insights into serotonin receptor ligands polypharmacology. *Eur. J. Med. Chem.* **2018**, *151*, 797–814.

Mój wkład w publikację był związany z opracowaniem koncepcji badań oraz przeprowadzeniem eksperymentów modelowania molekularnego, tj. nałożeniem i porównaniem sekwencji aminokwasowych rozpatrywanych receptorów serotoninowych wraz z identyfikacją pozycji mogących odpowiadać za selektywność, przygotowaniem zestawów związków selektywnych i nieselektywnych na podstawie danych z bazy ChEMBL, przeprowadzeniem procedury modelowania homologicznego receptorów 5-HT<sub>1A</sub>, 5-HT<sub>2A</sub>, 5-HT<sub>6</sub> oraz 5-HT<sub>7</sub>, przeprowadzeniem dokowania odpowiednich zestawów związków wraz ze statystyczną analizą częstości występowania poszczególnych oddziaływań w rozpatrywanych grupach ligandów, przeprowadzeniem procedury wirtualnego przesiewania bazy ZINC pod kątem poszukiwania nowych selektywnych ligandów, wyborem związków do zakupu oraz analizą wyników badań powinowactwa *in vitro* do rozpatrywanych receptorów. Ponadto, przeprowadziłam konfrontację wyników eksperymentalnej weryfikacji powinowactwa zakupionych związków z rezultatami modelowania molekularnego oraz przygotowałam wstępną wersję manuskryptu wraz z figurami.

**H3. Vass, M.†; Podlewska, S.‡;** de Esch, I. J. P.; Bojarski, A. J.; Leurs, R.; Kooistra, A. J.; de Graaf, C. Aminergic GPCR–Ligand Interactions: A Chemical and Structural Map of Receptor Mutation Data. *J. Med. Chem.* **2019**, *62*, 3784–3839.

Moją rolą w przedstawionej pracy było dokonanie przeglądu literaturowego w zakresie danych mutagenetycznych dla receptorów serotoninowych, dopaminowych, muskarynowych oraz alfa-adrenergicznych oraz wykonanie analizy statystycznej uzyskanych danych. Ponadto, opracowałam protokół obliczeniowy do przewidywania efektu mutagenetycznego dla danej pary: mutacja w określonej pozycji-ligand, który zastosowałam do analizy informacji mutagenetycznej dla wszystkich receptorów aminergiczych rozpatrywanych w publikacji (34 podtypy). Brałam również udział w przygotowaniu tekstu publikacji oraz części figur (fig. 4, fig. 5, fig. 8, fig. 11, fig. 14, fig. 17, fig. 21).

**H4.** Leśniak, D.; Podlewska, S.\*; Jastrzębski, S.; Sieradzki, I.; Bojarski, A. J.; Tabor, J. Development of new methods needs proper evaluation – benchmarking sets for machine learning experiments for class A GPCRs. *J. Chem. Inf. Model.* **2019**, *59*, 4974–4992.

W publikacji **H4** pełniłam rolę koordynatora projektu, brałam udział w opracowywaniu koncepcji badań oraz przeprowadziłam doświadczenia wstępne na ograniczonym zbiorze danych. Brałam również udział w analizie wyników oraz dokonałam ich interpretacji. Przygotowałam także wstępną wersję manuskryptu wraz z figurami (jestem autorką następujących figur: 1, 2, 6). Ponadto, prowadziłam korespondencję z edytorem *Journal of Chemical Information and Modeling* oraz przygotowałam odpowiedzi na uwagi recenzentów.

**H5.** Jastrzębski, S.; Szymczak, M.; Pocha, A.; Mordalski, S.; Tabor, J.; Bojarski, A. J.; Podlewska S.\* Emulating Docking Results Using a Deep Neural Network: A New Perspective for Virtual Screening. *J. Chem. Inf. Model.* **2020**, *60*, 4246–4262.

Publikacja **H5** jest rezultatem zaprojektowanych przeze mnie badań związanych z opracowaniem protokołu do przewidywania wyników dokowania. Poza zaplanowaniem i koordynowaniem prac, przygotowałam zestawy związków (ligandów i związków nieaktywnych), przeprowadziłam procedury dokowania do odpowiednich receptorów w programie Glide, przekształciłam wyniki dokowania do postaci zero-jedynkowej umożliwiającej przygotowanie modeli uczenia maszynowego oraz dokonałam oceny skuteczności predykcji opracowanego protokołu. Ponadto, brałam udział w analizie wyników i ich interpretacji. Przygotowałam także tekst manuskryptu wraz z figurami (z wyłączeniem fig. 10 i fig. 12). Prowadziłam także korespondencję z edytorem *Journal of Chemical Information and Modeling* oraz przygotowałam odpowiedzi na uwagi recenzentów.

**H6.** Podlewska, S.; Bugno, R.; Kudła, L.; Bojarski, A. J.; Przewlocki, R. Molecular Modeling of  $\mu$  Opioid Receptor Ligands with Various Functional Properties: PZM21, SR-17018, Morphine, and Fentanyl—Simulated Interaction Patterns Confronted with Experimental Data. *Molecules*, **2020**, *25*, 4636.

Mój wkład pracy w publikację polegał na przeprowadzeniu badań *in silico*, tj. charakterystyce dostępnych kryształów receptorów opioidowych, zadokowaniu badanych związków (PZM21, SR-17018, morfiny, fentanylu) oraz związków referencyjnych (BF0, BU72) do wybranych struktur krystalicznych receptora opioidowego (4DKL, 5C1M, 6DDF),



analizie uzyskanych pól oraz przygotowaniu i przeprowadzeniu eksperymentów dynamiki molekularnej. Opracowałam także metodę analizy wyników dynamiki molekularnej polegającą na wyznaczeniu korelacji pomiędzy częstością oddziaływania związków z poszczególnymi aminokwasami a wynikami doświadczeń *in vitro* i zastosowałam ją do otrzymanych danych. Brałam również udział w opracowaniu koncepcji badań, interpretacji uzyskanych wyników i przygotowaniu manuskryptu (fragmenty tekstu odnoszące się do badań obliczeniowych oraz figury 1–9).

**H7. Podlewska, S., Latacz, G.; Łażewska, D.; Kieć-Kononowicz, K.; Handzlik, J. *In silico* and *in vitro* studies on interaction of novel non-imidazole histamine H3R antagonists with CYP3A4. *Bioorg. Med. Chem. Lett.* 2020, 30, 127147.**

W publikacji **H7** pełniłam rolę pomysłodawcy i koordynatora projektu. Opracowałam koncepcję i plan badań, przeprowadziłam dokowanie oraz wykonałam symulacje dynamiki molekularnej dla wybranych ułożeń związków. Przeprowadziłam również statystyczną analizę uzyskanych wyników wraz z ich interpretacją i wizualizacją. Ponadto, brałam udział w przygotowaniu tekstu manuskryptu, a także jestem autorką wszystkich figur znajdujących się w przedstawionym artykule.

**H8. Podlewska, S.\*; Bugno, R.; Lacivita, E.; Leopoldo, M.; Bojarski, A. J.; Handzlik, J.\* Low Basicity as a Characteristic for Atypical Ligands of Serotonin Receptor 5-HT2 *Int. J. Mol. Sci.* 2021, 22, 1035.**

Mój wkład pracy w publikację polegał na przygotowaniu danych o ligandach o niskiej zasadowości na podstawie bazy ChEMBL, analizie statystycznej rozkładu parametru  $pK_a$  dla ligandów rozpatrywanych receptorów, analizie selektywności dla wybranych ligandów niskozasadowych, analizie zasadowości i aktywności poszczególnych grup strukturalnych ligandów, przygotowaniu modeli homologicznych rozpatrywanych receptorów (natywnych oraz ze zmutowaną resztą kwasu asparaginowego D3.32), przeprowadzeniu procedury dokowania oraz analizie statystycznej oddziaływań występujących dla rozpatrywanych grup związków (zasadowe/niezasadowe) z poszczególnymi aminokwasami. Dokonałam także interpretacji uzyskanych wyników, wykonałam ich wizualizacje (wszystkie figury przedstawione w publikacji są mojego autorstwa) oraz brałam udział w przygotowaniu tekstu publikacji.

**H9.** Wojtuch, A.; Jankowski, R.; **Podlewska, S.\*** How can SHAP values help to shape metabolic stability of chemical compounds? *J. Cheminf.* **2021**, *13*, 74.

W badaniach, których przeprowadzenie zaowocowało publikacją **H9** pełniłam rolę pomysłodawcy oraz koordynatora. Moim zadaniem było również przygotowanie odpowiednich zestawów danych do konstrukcji modeli predykcyjnych przewidujących stabilność metaboliczną wraz z ich wstępną analizą i przekształceniem do odpowiedniej postaci, wykonanie analizy i interpretacji uzyskanych wyników, a także zaprojektowanie serwisu online. Ponadto, przygotowałam manuskrypt (tekst oraz figury), prowadziłam korespondencję z edytorem *Journal of Cheminformatics*, jak również jestem autorką odpowiedzi na uwagi recenzentów.

**H10.** Baltrukevich, H.; **Podlewska, S.\*** From data to knowledge: systematic review of tools for automatic analysis of molecular dynamics output. *Front. Pharmacol.* **2022**, *13*, 844293 (praca przeglądowa).

Moją rolą w publikacji było dokonanie przeglądu literatury w zakresie automatycznych podejść do analizy wyników dynamiki molekularnej, przygotowanie następujących rozdziałów: 3, 4, 5 oraz współtworzenie rozdziału 6 i wniosków. Dokonałam również ostatecznej edycji całego manuskryptu, prowadziłam korespondencję z edytorem *Frontiers in Pharmacology* oraz wprowadziłam poprawki po uwagach recenzentów.

**H11.** Penna, E.‡; Niso, M.; **Podlewska, S.‡**; Volpicelli, F.; Crispino, M.; Perrone-Capano, C.; Bojarski, A.J.; Lacivita, E.; Leopoldo, M. In vitro and in silico analysis of the residence time of serotonin 5-HT<sub>7</sub> receptor ligands with arylpiperazine structure: a structure-kinetics relationship study. *ACS Chemical Neurosci.* **2022**, *13*, 497–509.

Mój wkład pracy w publikację był związany z przeprowadzeniem badań modelowania molekularnego, tj. opracowaniem koncepcji badań pozwalających na wyjaśnienie obserwowanych parametrów kinetycznych związków, analizą wartości różnych deskryptorów molekularnych w odniesieniu do parametrów kinetycznych, dokowaniem związków do receptora serotoninowego 5-HT<sub>7</sub>, przeprowadzeniem symulacji dynamiki molekularnej i analizą uzyskanych wyników. Zaproponowałam również wytłumaczenie obserwowanych zależności kinetycznych oraz przygotowałam fragment manuskryptu dotyczący modelowania molekularnego wraz z figurami (fig. 3, 4, 5).

## 4.4. Omówienie celu naukowego przedstawionego osiągnięcia naukowego oraz otrzymanych wyników

### 4.4.1. Wprowadzenie

Proces projektowania nowych leków jest niezwykle złożony, długotrwały i kosztowny.<sup>1,2</sup> Jednak rosnące zapotrzebowanie na różnego rodzaju medykamenty sprawia, że mimo trudności i konieczności ponoszenia znacznych nakładów finansowych, liczba przedsięwzięć podejmowanych w celu odkrycia nowych substancji leczniczych ciągle rośnie.<sup>3</sup> Poszukiwanie związków o określonej aktywności biologicznej nie jest jednak działaniem charakterystycznym dla współczesności, lecz towarzyszy człowiekowi od zarania dziejów. Początkowo, leków poszukiwano w świecie fauny i flory, jednak proces ten był w głównej mierze oparty na intuicji, szczęściu i przypadku.<sup>4</sup> Mimo to, już starożytne cywilizacje Majów, Azteków i Egipcjan posiadają udokumentowane stosowanie różnorodnych mikstur leczniczych, co czyni farmakologię najstarszą (a jednocześnie najbardziej dynamicznie rozwijającą się) dziedziną medycyny.<sup>5,6</sup>

Doświadczenia oraz wiedza medyczna i farmakologiczna przekazywane kolejnym pokoleniom przyczyniły się z czasem do zrewolucjonizowania procesu poszukiwania nowych leków. Obecnie jesteśmy uzbrojeni w wiedzę farmakologiczną, medyczną, biologiczną, fizyczną i chemiczną z jednej strony, a z drugiej możemy korzystać z rosnącej nieustannie mocy obliczeniowej komputerów i nowych technologii. Odkrywane są mechanizmy molekularne chorób, poznawane są fizyczne i chemiczne procesy, którym podlegają ksenobiotyki po wprowadzeniu do organizmu, a także mamy dostęp do coraz doskonalszych i dokładniejszych procedur diagnostycznych. Wszystko to sprawia, że nowe substancje lecznicze nie są już pozyskiwane w sposób przypadkowy, lecz są rezultatem przemyślanego i skrupulatnie zaplanowanego procesu. Współcześnie, nowe terapie są opracowywane w procesie tzw. racjonalnego projektowania leków, składającego się z siedmiu podstawowych etapów: identyfikacji celu molekularnego (najczęściej są nim białka receptorowe i enzymatyczne), weryfikacji celu biologicznego, znalezienia struktury wiodącej, optymalizacji tej struktury, badań przedklinicznych, badań klinicznych (faza I, II i III) oraz wprowadzenia leku na rynek.<sup>7</sup>

Każdy z etapów rozwoju leku produkuje ogromne ilości danych, dlatego różnorodne narzędzia statystyczne i technologie bazodanowe są niezbędne, aby móc nimi odpowiednio zarządzać i korzystać z nich w celu wspomaganie procesów decyzyjnych i analitycznych. Narzędzia obliczeniowe odgrywają jednak istotną rolę już na etapie poszukiwania nowych kandydatów na leki, w ramach tzw. komputerowo wspomaganego projektowania leków

(ang. *computer-aided drug design, CADD*). Wykorzystanie narzędzi *in silico* nie ogranicza się tylko i wyłącznie do identyfikacji związków posiadających potencjalnie pożądaną profil aktywności biologicznej, lecz obejmuje także analizę i optymalizację ich własności fizykochemicznych, farmakokinetycznych, a także potencjalnej toksyczności.<sup>8-11</sup>

Wśród szerokiego wachlarza metod stosowanych w obszarze obliczeniowej chemii medycznej możemy wyróżnić techniki oparte tylko i wyłącznie o strukturę i właściwości znanych ligandów o określonym profilu (ang. *ligand-based drug design, LBDD*),<sup>12</sup> jak i podejścia wykorzystujące informacje o strukturze przestrzennej celu biologicznego (ang. *structure-based drug design, SBDD*).<sup>13</sup> Poza najpopularniejszymi metodologiami, takimi jak modelowanie farmakoforowe<sup>14</sup> czy dokowanie,<sup>15</sup> coraz większą popularność w zadaniach komputerowo wspomaganego projektowania leków zaczynają zdobywać algorytmy uczenia maszynowego, pozwalające na kompleksową i względnie szybką analizę bardzo dużych ilości danych.<sup>16</sup>

### Metody uczenia maszynowego

Algorytmy uczenia maszynowego są obecnie wykorzystywane niemal we wszystkich obszarach życia, pozwalając na uzyskanie informacji i zależności pomiędzy danymi, których identyfikacja jest poza zasięgiem ludzkiego umysłu. Najbardziej spektakularnymi przykładami wykorzystania tego rodzaju narzędzi są autonomiczne samochody, technologia rozpoznawania twarzy czy samozarządzające się magazyny. Warta podkreślenia jest również niezwykła zdolność algorytmów uczenia maszynowego do szybkiego osiągnięcia poziomu arcymistrzowskiego w grze w szachy czy starochińskiej grze planszowej Go (system AlphaGo pokonał w 2016 roku najlepszego gracza na świecie), gdzie coraz częściej bitwę człowiek kontra maszyna wygrywa właśnie ta druga.

Dwie główne klasy problemów, które można rozpatrywać przy wykorzystaniu metod uczenia maszynowego to klasyfikacja (w przypadku związków chemicznych najczęściej pod kątem aktywności (aktywny/nieaktywny) wobec określonego celu biologicznego) oraz regresja (przewidywanie konkretnej wartości numerycznej danego parametru, np. rozpuszczalności, wartości współczynnika podziału oktanol-woda czy czasu półtrwania w przypadku analizy stabilności metabolicznej).<sup>17</sup> Uczenie maszynowe stosuje się również do poszukiwania wzorców danych i grupowania ich w podzbiory według określonych cech (klasteryzacja), co w przypadku związków chemicznych może oznaczać np. automatyczną identyfikację związków posiadających ten sam rdzeń strukturalny.<sup>18</sup>

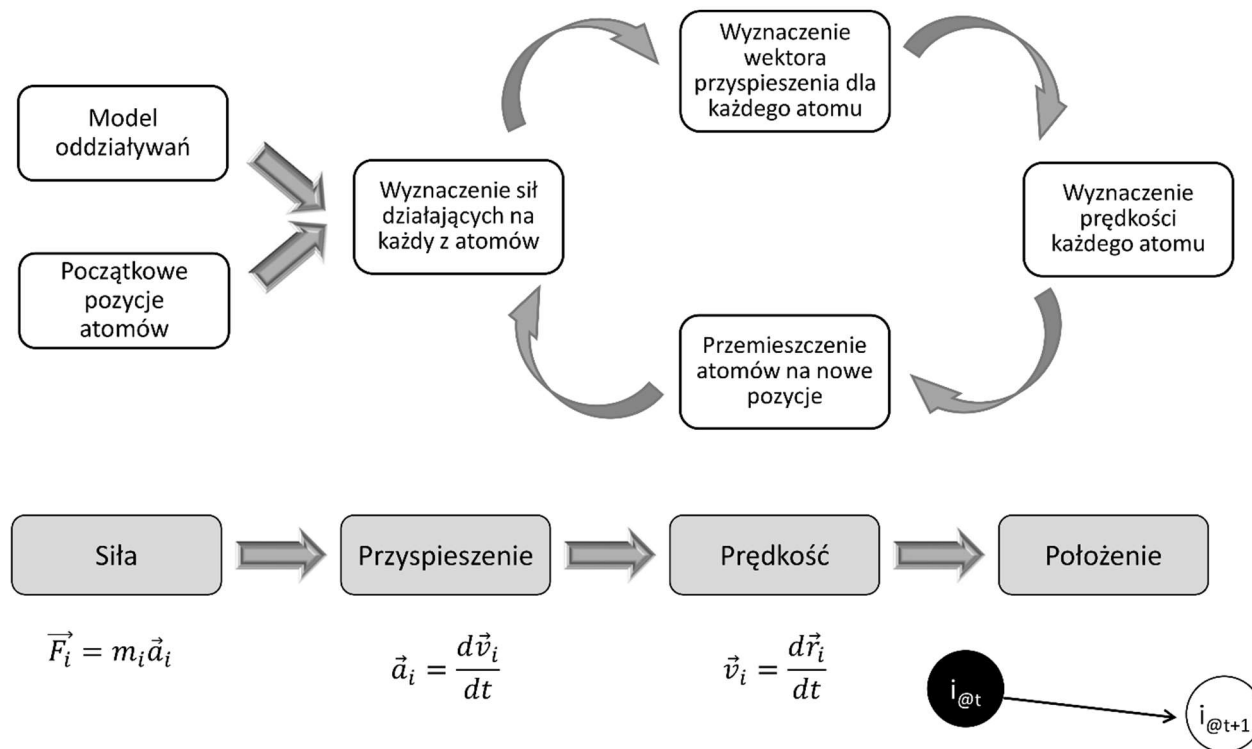
W ostatnich latach, wykorzystanie uczenia maszynowego w komputerowo wspomaganym projektowaniu leków przestało ograniczać się tylko i wyłącznie do oceny potencjalnej aktywności lub innych własności związków,<sup>19-23</sup> lecz zaczęto je również wykorzystywać do proponowania struktur nowych potencjalnych ligandów przy zastosowaniu tzw. metod generatywnych.<sup>24-26</sup>

### Dynamika molekularna

Przewidywania sposobu oddziaływania liganda z receptorem dokonuje się zwykle przy pomocy dokowania. W procesie tym, poszukuje się bioaktywnej konformacji molekuly w miejscu wiążącym białka.<sup>27,28</sup> Dokowanie przeprowadza się najczęściej w układzie labilny ligand-sztywne białko, tak więc zmiany konformacyjne zachodzą jedynie w obrębie związku małowcząsteczkowego. Pomimo opracowania metod pozwalających na uwzględnienie również zmian w ułożeniu atomów białka (dokowanie w układzie labilny ligand-labilne białko), nie istnieje podejście, które brałoby pod uwagę całą dynamikę procesu dokowania molekularnego, a informacje uzyskiwane w wyniku realizacji tej procedury dotyczą jedynie określonego momentu czasu.

Znacznie więcej danych o potencjalnym wzajemnym zachowaniu liganda i białka można uzyskać poprzez symulacje dynamiki molekularnej, które pozwalają na śledzenie ewolucji układu ligand-białko w czasie.<sup>29</sup> Badanie analizowanego układu w funkcji czasu można przeprowadzić korzystając z zasad dynamiki Newtona, określających związek pomiędzy ruchem ciała, a działającymi na niego siłami. Pełnoatomowe symulacje są niezwykle kosztowne obliczeniowo, gdyż wymagają określenia pozycji i oddziałujących sił dla każdego atomu w układzie. Z drugiej strony, moc obliczeniowa komputerów nieustannie rośnie, a ponadto, do prowadzenia symulacji dynamiki molekularnej możliwe jest również wykorzystanie kart graficznych (ang. *Graphical Process Unit, GPU*).<sup>30,31</sup> Pozwala to obecnie na prowadzenie symulacji dynamiki molekularnej w skali mikro- lub nawet milisekundowej (w przypadku, gdy rozważane są symulacje pełnoatomowe). Dla układów uproszczonych, możliwe jest prowadzenie symulacji trwających nawet kilka sekund. Dlatego też, przy zastosowaniu odpowiednich ustawień, skuteczne symulowanie procesu wiązania ligandów (skala mikro- do milisekundowa), czy zwijania białek (milisekundy – sekundy) jest obecnie w zasięgu dostępnej mocy obliczeniowej, a dynamika molekularna jest coraz częściej stosowana w zadaniach związanych z komputerowo wspomaganym projektowaniem leków.<sup>32,33</sup>

Symulacje dynamiki molekularnej rozpoczyna się od zdefiniowania położenia wszystkich atomów w układzie. Następnie wyznacza się działające na nie siły oraz rozwiązywane są równania ruchu, pozwalające na wyznaczenie nowych położenia atomów w kolejnym kroku czasowym (Rysunek 1).



**Rysunek 1.** Przebieg procesu symulacji dynamiki molekularnej.

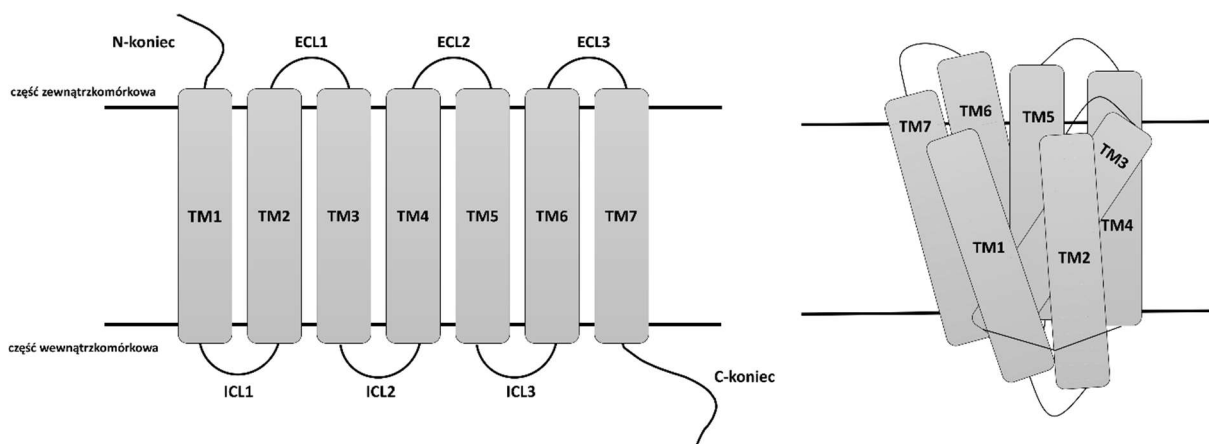
Do wyznaczenia sił działających na każdy z atomów układu wykorzystuje się funkcję, która reprezentuje jego energię potencjalną. Jest ona zdefiniowana przez pole siłowe, które uwzględnia przyczynki energii potencjalnej związane ze zmianami długości wiązań chemicznych, kątów i kątów torsyjnych, a także oddziaływaniami elektrostatycznymi i van der Waalsa.<sup>34</sup>

### Receptory sprzężone z białkiem G

Receptory sprzężone z białkiem G (ang. *G protein-coupled receptors, GPCRs*) są największą i najbardziej zróżnicowaną grupą białek sygnałowych w ludzkim genomie. Receptory GPCR są jedną z głównych klas celów biologicznych dla różnorodnych kampanii projektowania leków, z uwagi na ich udział w znakomitej większości procesów fizjologicznych zachodzących w organizmach żywych.<sup>35,36</sup> Ważkość prac nad receptorami GPCR potwierdza przyznanie w 2012 roku Nagrody Nobla dla Roberta Lefkowitza i Briana Kobilki za badania w obrębie tej grupy receptorów.

Analiza podobieństwa strukturalnego i funkcjonalnego pozwoliła na wyróżnienie 6 głównych klas GPCR (A–F), przy czym trzy najlepiej zbadane obejmują: klasę A (receptory rodopsyno-podobne, ponad 80% wszystkich sekwencji GPCR), klasę B (receptory sekretynopodobne) i klasę C (tworzoną m.in. przez receptory dla glutaminianu). Jednymi z przedstawicieli GPCRów klasy A są receptory aminergiczne, do których zaliczamy receptory serotoninowe, muskarynowe, adrenergiczne, dopaminowe, histaminowe oraz receptory amin śladowych.

Receptory GPCR są białkami transbłonowymi, których szkielet tworzy 7 hydrofobowych helis zanurzonych w błonie komórkowej (TM1–TM7), połączonych trzema pętlami zewnątrzkomórkowymi (ECL1–ECL3) oraz trzema pętlami wewnątrzkomórkowymi (ICL1–ICL3). Wolna grupa aminowa (tzw. N-koniec receptora) znajduje się na zewnątrz komórki, podczas gdy C-koniec jest elementem wewnątrzkomórkowym (Rysunek 2).<sup>37</sup> Szeroki ligandów aktywujących receptory GPCR obejmuje m.in. fotony, jony, lipidy, neuroprzekaźniki, hormony i małe białka. wachlarz ligandów aktywujących (ICL1–ICL3). Wolna grupa aminowa (tzw. N-koniec receptora) znajduje się na zewnątrz komórki, podczas gdy C-koniec jest elementem wewnątrzkomórkowym (Rysunek 2).<sup>37</sup> Szeroki wachlarz ligandów aktywujących receptory GPCR obejmuje m.in. fotony, jony, lipidy, neuroprzekaźniki, hormony i małe białka.



**Rysunek 2.** Budowa ogólna receptorów sprzężonych z białkiem G.

### [Bazy danych](#)

Rosnąca ilość danych chemicznych i farmakologicznych generuje potrzebę ich usystematyzowania i uporządkowania, co ułatwia zarządzanie oraz wyszukiwanie potrzebnych informacji. Dla badaczy poszukujących nowych leków, jednymi z najważniejszych typów danych są wiadomości o aktywności związków wobec określonego celu biologicznego. Tego rodzaju informacji dostarcza m.in. baza ChEMBL



(<https://www.ebi.ac.uk/chembl/>),<sup>38</sup> która gromadzi i standaryzuje dane literaturowe o powinowactwie związków, ich aktywności funkcjonalnej, a także własnościach ADMET.

Do przewidywania sposobu wiązania liganda z celem biologicznym, potrzebna jest znajomość jego trójwymiarowej struktury. Dane krystaliczne o strukturach białkowych gromadzone są w bazie Protein Data Bank (PDB, <https://www.rcsb.org/>),<sup>39</sup> która obecnie zawiera ponad 180 tys. rekordów.

Jedną z technik wykorzystywaną w procesie komputerowo wspomaganego projektowania leków, jest ich poszukiwanie w komercyjnych bazach związków. Firmy dostarczające związki chemiczne do badań posiadają swoje własne katalogi oferowanych produktów, jednak badacze najczęściej korzystają ze skonsolidowanych źródeł potencjalnych związków do zakupu, jak np. z bazy ZINC (<https://zinc20.docking.org/>).<sup>40</sup>

Zbiorami danych, które również cieszą się dużą popularnością wśród badaczy z dziedziny projektowania leków są m.in. DrugBank (baza informacji o lekach, dostarczająca ich kompleksowej charakterystyki fizykochemicznej i farmakologicznej; <https://go.drugbank.com/>)<sup>41</sup>; PubChem (katalog substancji chemicznych w odniesieniu do ich aktywności biologicznej; <https://pubchem.ncbi.nlm.nih.gov/>)<sup>42</sup> oraz GPCRdb (wszechstronne źródło wiedzy i narzędzi do badań nad receptorami GPCR; <https://gpcrdb.org/>).<sup>43</sup>

#### 4.4.2. Cel badawczy

Celem badawczym prac wchodzących w skład dzieła habilitacyjnego było opracowanie narzędzi obliczeniowych pozwalających na przyspieszenie i zwiększenie efektywności kampanii poszukiwań nowych substancji o potencjalnym zastosowaniu terapeutycznym. Procedury te z założenia miały być oparte o wykorzystanie metod statystycznych i uczenia maszynowego, zarówno w podejściu *ligand-*, jak i *structure-based*. W tym ostatnim przypadku były one projektowane zarówno pod kątem zastosowania w dokowaniu, jak i do analizy wyników dynamiki molekularnej. Opracowane podejścia są z założenia uniwersalne, jednak ich przydatność do zastosowań w komputerowo wspomaganym projektowaniu leków i optymalizację prowadzono głównie na przykładzie aminergicznych receptorów GPCR.

#### Szczegółowe cele badawcze obejmują:

- Konstrukcję zestawu danych odniesienia do oceny efektywności nowo powstających narzędzi bazujących na metodach uczenia maszynowego [H4]



- Opracowanie procedury generowania struktur nowych potencjalnych ligandów na podstawie optymalnego zestawu podstruktur wyznaczanego przez metody uczenia maszynowego [H1]
- Określenie determinantów odpowiedniego profilu aktywności (selektywność) i własności (zasadowość) dla ligandów receptorów serotoninowych [H2, H8]
- Przygotowanie zestawu danych mutagenetycznych dla aminergicznych receptorów GPCR oraz opracowanie procedury pozwalającej na przewidywanie wpływu podmiany określonego aminokwasu na powinowactwo danego liganda [H3]
- Opracowanie metody pozwalającej na przewidywanie wyniku dokowania z dwuwymiarowej struktury związku chemicznego [H5]
- Konstrukcja narzędzi do automatycznej analizy wyników dynamiki molekularnej [H6, H7, H10, H11]
- Opracowanie podejścia do interpretowania przewidywań własności związków uzyskiwanych przy wykorzystaniu metod uczenia maszynowego, umożliwiającego dokonanie optymalizacji związku chemicznego pod kątem rozpatrywanego parametru [H9]

#### 4.4.3. Omówienie wyników badań

Realizacja postawionych celów badawczych zaowocowała powstaniem zestawu narzędzi i protokołów obliczeniowych wspierających poszukiwanie nowych kandydatów na leki zarówno w ścieżce *ligand-*, jak i *structure-based*.

#### [Konstrukcja referencyjnych zbiorów danych i przeprowadzenie eksperymentów odniesienia na potrzeby oceny narzędzi opartych o metody uczenia maszynowego](#)

Problemem, który występuje podczas opracowywania narzędzi obliczeniowych do zastosowań w komputerowo wspomaganym projektowaniu leków jest odpowiednia ocena skuteczności skonstruowanych narzędzi. Dzięki współpracy z Grupą Uczenia Maszynowego z Wydziału Matematyki i Informatyki Uniwersytetu Jagiellońskiego możliwe było przeprowadzenie eksperymentów referencyjnych, które pozwoliły na zbadanie wielu czynników wpływających na uzyskiwane wyniki [H4]. Aspektami, które poddano badaniu były: sposób ekstrakcji danych z bazy zawierającej informacje o aktywności związków (w naszym przypadku korzystaliśmy z bazy ChEMBL),<sup>38</sup> reprezentacja związków (skonfrontowano 16 reprezentacji generowanych zarówno z dwu-, jak i trójwymiarowej

struktury związku), sposób podziału związków na zbiór uczący i testowy, rodzaj zastosowanego algorytmu uczącego, a także rodzaj parametru oceniającego skuteczność predykcji.

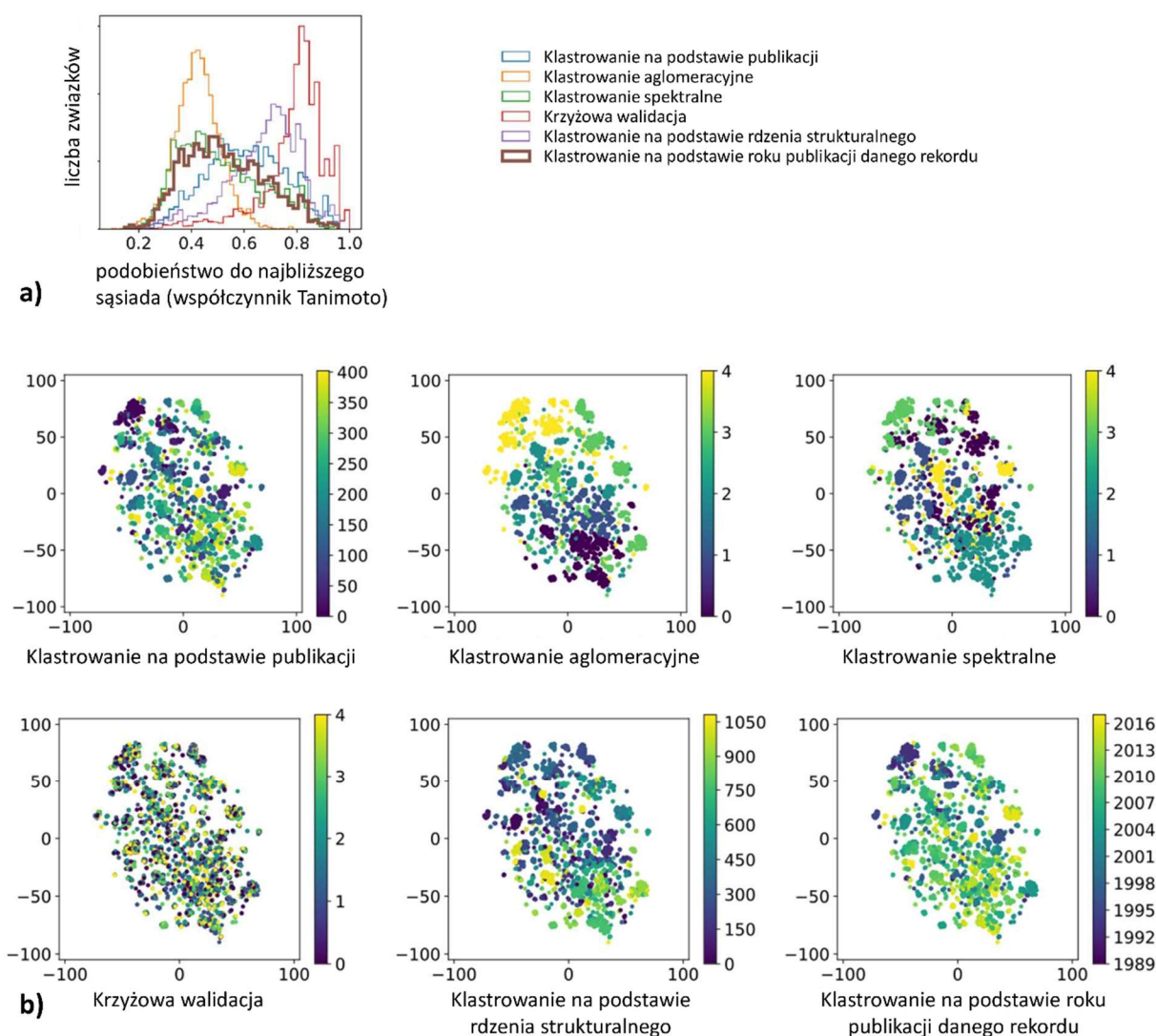
Pomimo tego, że aspekty takie jak wpływ reprezentacji związków na uzyskiwane predykcje czy rodzaj algorytmu uczącego są relatywnie często przedmiotem różnego rodzaju badań porównawczych, rola innych czynników, takich jak sposób konstrukcji zbioru danych czy też sposób podziału na zbiór uczący i testowy, nie został do czasu podjęcia naszych badań w tym zakresie dogłębnie zbadany.

Procedura przygotowania zbioru danych wymaga podjęcia wielu (mogłoby się wydawać mało znaczących) decyzji związanych z włączaniem lub wyłączeniem określonych rekordów ze zbioru. W naszych badaniach wzięliśmy pod uwagę takie aspekty jak rozpatrywanie różnych parametrów aktywności związków (np. porównano zbiory danych oparte tylko o parametr  $K_i$  oraz takie, które łączyły dane o  $K_i$  oraz  $IC_{50}$ ), relacje pomiędzy parametrem aktywności, a jego wartością („<”, „>”, „=”), pracę na wartościach zlogarytmowanych lub oryginalnych, usuwanie duplikatów (bądź branie pod uwagę wszystkich dostępnych danych dla danego związku) oraz ewentualne odrzucanie rekordów, których poprawność nie została ręcznie zweryfikowana. Pomimo tego, że zestawy danych generowane dla różnych kombinacji wymienionych wyżej aspektów różniły się znacznie pod względem ilościowym, nie prowadziły one jednak do znacznych różnic w skuteczności predykcji realizowanej przez metody uczenia maszynowego.

Czynnikiem, który miał zdecydowanie największy wpływ na uzyskiwane wyniki ewaluacji różnych algorytmów uczących był sposób podziału zbioru danych na część uczącą i testową. Wizualizacja przestrzeni chemicznej oraz analiza podobieństwa do najbardziej zbliżonego związku dla różnych podziałów (na podstawie różnych sposobów klastrowania) pokazuje, jak bardzo niemiernodajne wyniki można uzyskać przy zastosowaniu nieodpowiedniego podejścia do wyodrębnienia zbioru uczącego (Rysunek 3).

Analiza rozkładów współczynników podobieństwa pomiędzy zestawem uczącym i testowym dla różnych podziałów (Rysunek 3a) pokazuje, jak zmienia się trudność zadania predykcji aktywności związków w zależności od zastosowanego podejścia. W przypadku walidacji krzyżowej, gdzie zbiór danych dzielony jest w sposób losowy, współczynniki podobieństwa pomiędzy elementami zbioru uczącego i testowego przyjmują wyższe wartości. Oznacza to, że zadanie predykcyjne jest istotnie ułatwione i oceniana metoda może dawać fałszywie dobre wyniki, podczas gdy w realnych zastosowaniach w procesie poszukiwania nowych terapeutyków jej przydatność będzie znikoma (w zadaniach wirtualnego przesiewania

zwykle ocenie podlegają związki reprezentujące szeroką przestrzeń chemiczną, a tym samym istotnie różne od przykładów znajdujących się w zbiorze uczącym). Sytuację najwierniej oddającą rzeczywiste wyzwania jakie stoją przed metodami uczenia maszynowego w projektowaniu leków zapewnia wyodrębnianie zbioru uczącego w oparciu o klastrowanie aglomeracyjne (indukcyjne; punkt startowy: jednoelementowe klastry w liczbie równej całkowitej liczbie obiektów, w kolejnych krokach obiekty te są łączone w większe skupiska na podstawie przyjętej metryki odległości). Pomimo tego, że z reguły w takim trybie oceny retrospektywnej uzyskiwane są znacznie gorsze wyniki niż w przypadku pozostałych podejść (a w szczególności w porównaniu do krzyżowej walidacji), to właśnie ono zapewnia najbardziej realną ewaluację możliwości predykcyjnych danej metody. Oceniana jest bowiem zdolność do poprawnego przewidywania własności nowych strukturalnie związków. Analiza współczynników podobieństwa (Rysunek 3a) przekłada się również na wnioski wypływające z wizualizacji przestrzeni chemicznej (Rysunek 3b), gdzie poszczególne klastry (reprezentowane przez różne kolory) znajdują się w osobnych obszarach tej przestrzeni.



**Rysunek 3.** Analiza podziałów zbioru danych na zbiór uczący i testowy opartych o różnego rodzaju klawrowania na przykładzie ligandów receptora serotoninowego 5-HT<sub>1A</sub>; a) rozkład współczynników podobieństwa pomiędzy wszystkimi elementami zbioru testowego, a uczącego dla różnych podziałów przekładający się na trudność zadania klasyfikacji; b) wizualizacja przestrzeni chemicznej ligandów dla różnych podziałów za pomocą techniki tSNE - stochastycznej metody porządkowania sąsiadów w oparciu o rozkład t [H4].

Z uwagi na fakt, że ocena podobieństwa jest silnie uzależniona od sposobu reprezentacji związków, wprowadzono również nowy sposób oceny, uwzględniający dwie różne reprezentacje struktur chemicznych:

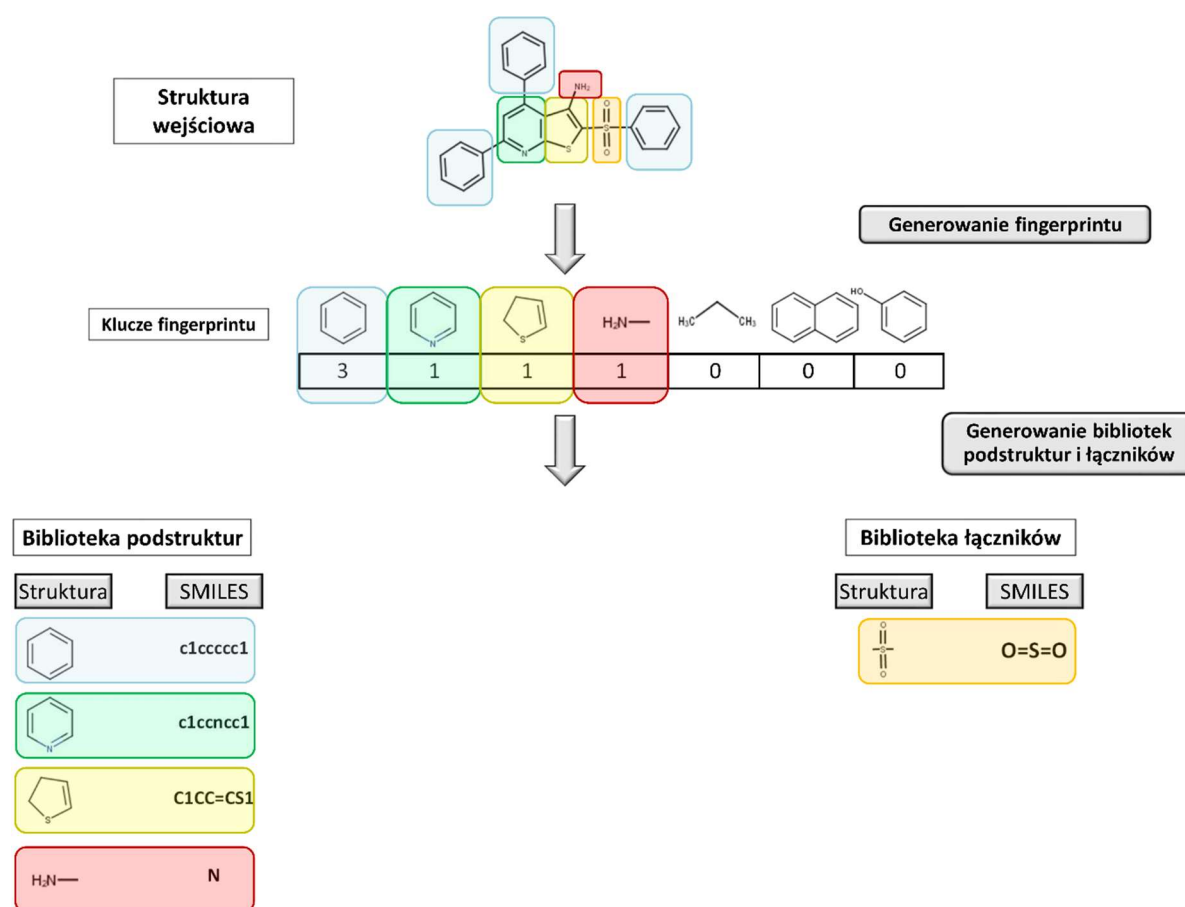
$$\text{podobieństwo} = \frac{\text{wsp.Tanimoto (Morgan)} + \text{wsp.Tanimoto (AtomPairs)}}{2},$$

gdzie  $\text{wsp.Tanimoto(Morgan)}$  jest współczynnikiem Tanimoto<sup>44</sup> wyznaczonym na fingerprintie Morgana,<sup>45</sup> natomiast  $\text{wsp.Tanimoto(AtomPairs)}$  jest współczynnikiem Tanimoto wyznaczonym na fingerprintie AtomPairs.<sup>46</sup>

Poza omówionymi analizami, przygotowano także dane uczące dla 100 receptorów GPCR z klasy A (na podstawie danych dostępnych w bazie ChEMBL) oraz udostępniono kod źródłowy pozwalający na przeprowadzenie analogicznych eksperymentów do tych, które zaprezentowano w publikacji.

[Opracowanie metodologii generowania nowych potencjalnych ligandów danego receptora w oparciu o optymalne fragmenty strukturalne wyznaczone przez metody uczenia maszynowego](#)

Jedną z podstawowych ścieżek poszukiwania kandydatów na nowe terapeutyki jest identyfikacja potencjalnie aktywnych związków w różnorodnych bibliotekach struktur chemicznych. Zestawy komercyjnie dostępnych związków zawierają obecnie kilkadziesiąt milionów przykładów,<sup>47</sup> są one jednak już dość intensywnie przebadane w różnorodnych kampaniach poszukiwania nowych leków. Kolejnym źródłem potencjalnych ligandów są biblioteki wirtualne, których przykładem są biblioteki kombinatoryczne. Podstawą

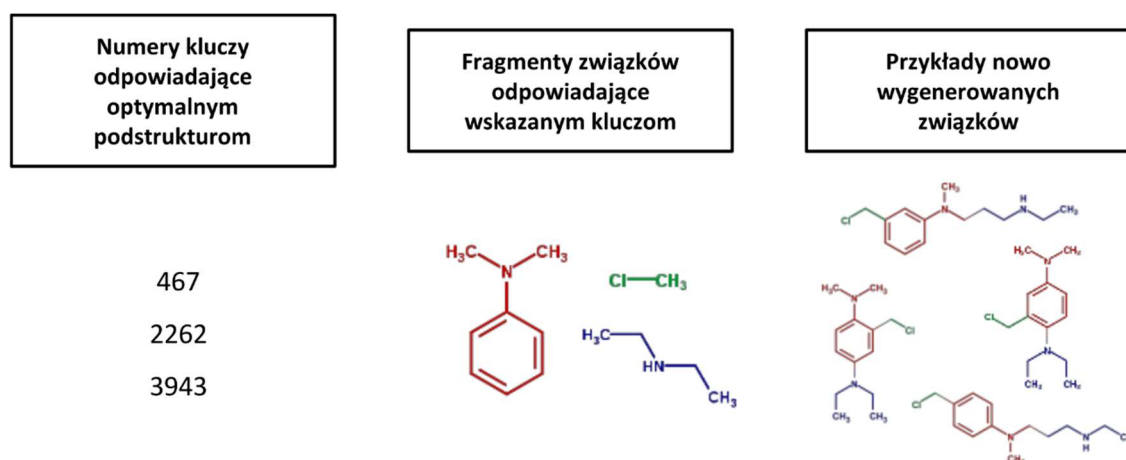


**Rysunek 4.** Schemat powstawania bibliotek podstruktur i łączników na podstawie wygenerowanej reprezentacji związków [H1].

ich generowania jest systematyczna podmiana określonych ugrupowań chemicznych, przy wykorzystaniu wiedzy chemicznej, jak to ma miejsce np. w podstawieniu bioizosterycznym,<sup>48</sup> podmianie rdzenia związku (ang. *scaffold hopping*)<sup>49</sup> czy hybrydyzacji znanych ligandów.<sup>50</sup>

W swoich badaniach, opracowałam nowatorski sposób konstruowania bibliotek związków, wykorzystujący jako punkt wyjścia optymalną reprezentację struktur chemicznych [H1]. Polega on na przekształceniu zestawu związków o znanej aktywności (zarówno aktywnych, jak i nieaktywnych) wobec rozpatrywanego celu biologicznego do formy fingerprintu podstrukturalnego, a następnie optymalizacji otrzymanego ciągu bitowego przy wykorzystaniu metod uczenia maszynowego. Pierwszy etap stanowi przygotowanie zbioru fragmentów odpowiadających poszczególnym bitom zastosowanych fingerprintów (Rysunek 4). Z uwagi na niecałkowite pokrycie cząsteczek przez wykorzystane reprezentacje, tworzone są również niezależne zbiory tzw. łączników tak, aby wszystkie fragmenty strukturalne ligandów znalazły się w zestawach elementów do konstrukcji nowych bibliotek.

Następnie, dzięki zastosowaniu metod uczenia maszynowego, dla każdej reprezentacji generowany jest ciąg zero-jedynkowy, odpowiadający związkowi o optymalnej aktywności (optymalizację fingerprintów przeprowadził dr Wojciech Czarnecki z Wydziału Matematyki i Informatyki Uniwersytetu Jagiellońskiego). Ze względu na stosunkowo dużą liczbę fragmentów strukturalnych wskazywanych jako preferowanych dla aktywności związku (~25–40 w zależności od reprezentacji) niemożliwe jest przygotowanie bibliotek ligandów zawierających wszystkie zidentyfikowane podstruktury. Z zestawu optymalnych fragmentów tworzone są więc podzbiory trój-, cztero- i pięcioelementowe, na podstawie których generowane są następnie propozycje nowych ligandów (Rysunek 5).



**Rysunek 5.** Przykład procesu generowania nowych potencjalnych ligandów receptora serotoninowego 5-HT<sub>6</sub> [H1].



Za pomocą tak przygotowanego protokołu, wygenerowałam zestawy związków o potencjalnej aktywności wobec 8 białek: receptor 5-HT<sub>2A</sub>, 5-HT<sub>2C</sub>, 5-HT<sub>6</sub>, 5-HT<sub>7</sub>, β<sub>2</sub>-adrenergiczny, D<sub>2</sub>, katepsyna B oraz kinaza SRC. Utworzone w ten sposób biblioteki poddałam ocenie pod kątem aktywności względem danego celu biologicznego przy pomocy maszyny wektorów nośnych (ang. *Support Vector Machines, SVM*) dla dwóch reprezentacji: fingerprintu haszowanego (Extended Fingerprinter z pakietu PaDEL-Descriptor)<sup>51</sup> oraz fingerprintu Klekota-Roth. Ponadto, sprawdziłam jaki odsetek wygenerowanych związków jest obecny w bazie ZINC oraz czy nie przewiduje się, że zaproponowane struktury będą niespecyficycznie oddziaływać z wieloma celami biologicznymi (filtr PAINS: *Pan-assay interference compounds*)<sup>52</sup> – Tabela 1.

**Tabela 1.** Analiza skonstruowanych bibliotek związków.

Cel biologiczny	Liczba wygenerowanych związków	Liczba związków wskazanych jako aktywne w ocenie przy wykorzystaniu fingerprintu haszowanego (% wszystkich wygenerowanych związków)	Liczba związków wskazanych jako aktywne w ocenie przy wykorzystaniu fingerprintu Klekota-Roth (% wszystkich wygenerowanych związków)	Liczba związków obecnych w bazie ZINC (% wszystkich wygenerowanych związków)	Liczba związków, które uzyskały pozytywną ocenę w filtrze PAINS (% wszystkich wygenerowanych związków)
Receptor 5-HT <sub>2A</sub>	141 625	84 295 (60%)	105 938 (75%)	3 648 (3%)	139 901 (99%)
Receptor 5-HT <sub>2C</sub>	683 639	209 370 (31%)	305 122 (45%)	10 638 (2%)	653 117 (96%)
Receptor 5-HT <sub>6</sub>	240 910	119 995 (50%)	98 786 (41%)	6 430 (3%)	234 818 (97%)
Receptor 5-HT <sub>7</sub>	230 346	122 618 (53%)	170 091 (74%)	3 064 (1%)	228 884 (99%)
Receptor β <sub>2</sub> -adrenergiczny	67 856	8 052 (12%)	11 912 (18%)	2 508 (4%)	66 996 (99%)
Katepsyna B	8 569	468 (5%)	163 (2%)	669 (8%)	8467 (99%)
Receptor D <sub>2</sub>	194 422	56 779 (29%)	69 616 (36%)	7 092 (4%)	187 599 (96%)
Kinaza Src	27 642	3 510 (13%)	3713 (13%)	1 374 (5%)	27 639 (100%)

Wstępna ocena nowych zestawów struktur potwierdza ich wysoki potencjał aktywności w stosunku do poszczególnych celów biologicznych. Jest to zgodne z założeniami przyjętymi podczas opracowywania koncepcji projektu, tj. nowe propozycje związków są tworzone z fragmentów już istniejących ligandów. Powstałe zbiory związków są więc niejako

bibliotekami celowanymi, skupionymi na optymalizacji struktur pod kątem aktywności wobec konkretnego receptora. Ponadto, należą one do nieeksplorowanej dotychczas przestrzeni chemicznej (przynajmniej w zakresie komercyjnie dostępnych związków), z uwagi na obecność w bazie ZINC jedynie kilku procent wygenerowanych struktur. Jednocześnie, zaproponowane związki charakteryzuje niskie ryzyko niespecyficznego oddziaływania z wieloma białkami, gdyż praktycznie wszystkie z nich są pozytywnie oceniane w tym zakresie przez filtr PAINS.

Poza charakterystyką utworzonych bibliotek związków wykazałam również, że wykorzystanie metod uczenia maszynowego do wskazania najkorzystniejszych elementów strukturalnych prowadzi do istotnie różnych wyników niż w przypadku zastosowania podejścia struktur uprzywilejowanych.<sup>53</sup> Wynika to z faktu, że w procesie optymalizacji metodami uczenia maszynowego uwzględniane są zarówno związki aktywne, jak i nieaktywne wobec danego receptora, co pozwala na identyfikację podstruktur różnicujących te dwie grupy związków.

#### Określenie determinantów odpowiedniego profilu aktywności (selektywność) i własności (zasadowość) dla ligandów receptorów serotoninowych

Podczas poszukiwania nowych kandydatów na leki w obrębie określonej grupy receptorów, istotnym elementem jest zdefiniowanie wymagań, które powinien posiadać związek, aby móc aspirować do stania się przyszłym terapeutą. Kryteria te mogą być związane zarówno z występowaniem w strukturze określonych ugrupowań chemicznych, posiadaniem ściśle zdefiniowanego profilu własności fizykochemicznych, jak również tworzeniem ustalonych oddziaływań z celem biologicznym.

Celem moich badań było wyznaczenie determinantów do posiadania przez związek aktywności wobec receptora 5-HT<sub>7</sub>, przy jednoczesnym braku powinowactwa do pozostałych podtypów receptora serotoninowego, w szczególności 5-HT<sub>1A</sub>, 5-HT<sub>1B</sub>, 5-HT<sub>2A</sub>, 5-HT<sub>2B</sub>, oraz 5-HT<sub>6</sub> [H2].

Prace rozpoczęłam od porównania sekwencji wszystkich analizowanych białek wykorzystując narzędzia z bazy GPCRdb.<sup>43</sup> Największe podobieństwo do sekwencji receptora 5-HT<sub>7</sub> wykazały receptory 5-HT<sub>1A</sub> i 5-HT<sub>1B</sub>, dlatego też zadanie znalezienia związków selektywnych pomiędzy parami receptorów 5-HT<sub>7</sub>-5HT<sub>1A</sub> oraz 5-HT<sub>7</sub>-5HT<sub>1B</sub> oceniono za najtrudniejsze. Porównanie sekwencji badanych receptorów w obrębie miejsca wiążącego pozwoliło mi na wskazanie pozycji, w których znajdują się aminokwasy o różnym charakterze dla poszczególnych białek, mogące stanowić potencjalne źródło selektywności. Wykazano m.in., że dla receptora 5-HT<sub>7</sub> w pozycji 1x36 (numerowanie aminokwasów zgodnie



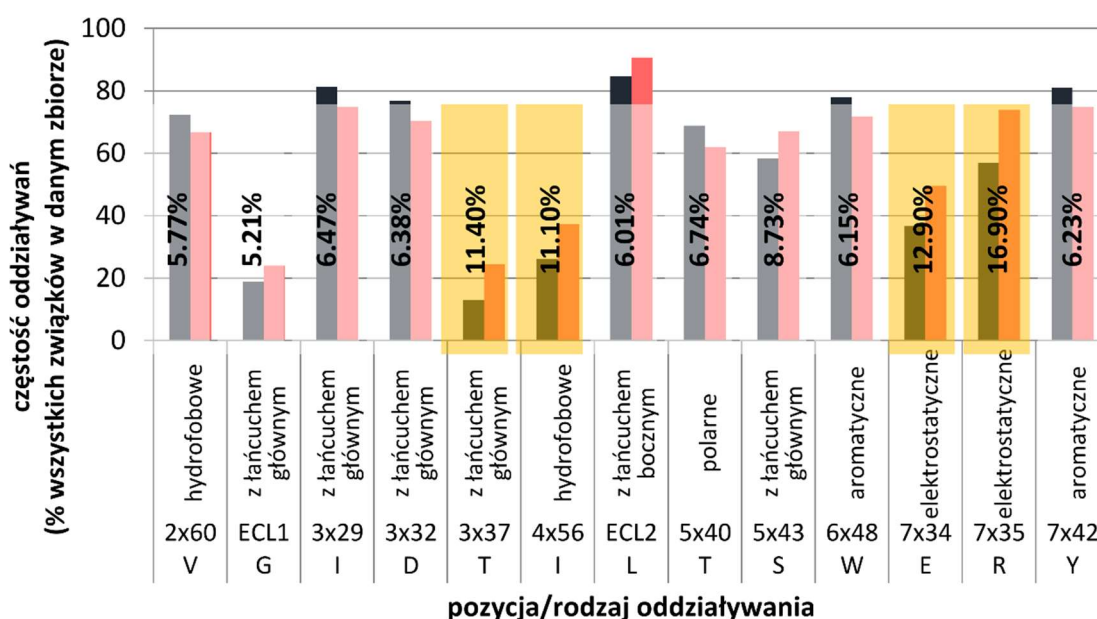
z numeracją z GPCRdb) występuje glicyna (aminokwas o niewielkich rozmiarach, łańcuch boczny stanowi tylko atom wodoru), podczas gdy dla pozostałych podtypów, w pozycji 1x36 obecne są aminokwasy o bardziej rozbudowanych łańcuchach bocznych (alanina, walina, seryna). Ponadto, receptor 5-HT<sub>7</sub> jest jedynym podtypem, dla którego w pozycji 2x64 występuje ujemnie naładowany kwas asparaginowy, a w pozycji 3x24 aminokwas o charakterze aromatycznym (fenyloalanina). Dodatkowo, pozycja 6x58 jest w receptorze 5-HT<sub>7</sub> zajmowana przez argininę z dodatnio naładowanym łańcuchem bocznym (dla pozostałych badanych podtypów, na tej pozycji występują aminokwasy obojętne); również na pozycjach 7x32, 7x34 oraz 7x35 występują aminokwasy charakterystyczne tylko dla receptora 5-HT<sub>7</sub> (odpowiednio: tryptofan, kwas glutaminowy, arginina).

Kolejnym krokiem było przygotowanie zbiorów ligandów rozpatrywanych receptorów (źródło danych: baza ChEMBL). Analiza liczności zestawów związków o poszczególnych profilach aktywności potwierdziła trudność zadania uzyskania selektywności pomiędzy receptorem 5-HT<sub>7</sub>, a 5-HT<sub>1A</sub>: pomimo największej ilości danych dostępnych dla receptora 5-HT<sub>1A</sub> (4632 związki o K<sub>i</sub> poniżej 100 nM i 1347 o K<sub>i</sub> powyżej 1000 nM, stanowiące odpowiednio zestawy związków aktywnych i nieaktywnych wobec tego receptora), jedynie 44 związki charakteryzowały się selektywną aktywnością 5-HT<sub>7</sub>/5-HT<sub>1A</sub>, podczas gdy związków o dualnej aktywności (5-HT<sub>7</sub> i 5-HT<sub>1A</sub>) było aż 341.

Następnie przeprowadziłam dokowanie przygotowanych zestawów związków do struktur krystalicznych (podczas realizacji badań dostępne były struktury krystaliczne receptorów 5-HT<sub>1B</sub> i 5-HT<sub>2B</sub>) oraz modeli homologicznych (dla receptorów 5-HT<sub>1A</sub>, 5-HT<sub>2A</sub>, 5-HT<sub>6</sub>, 5-HT<sub>7</sub>) rozpatrywanych białek. Otrzymane kompleksy ligand-receptor zakodowałam w postaci zero-jedynkowej przy wykorzystaniu fingerprintu oddziaływań strukturalnych (ang. *Structural Interaction Fingerprint, SIFt*),<sup>54</sup> co umożliwiło statystyczną analizę oddziaływań dla poszczególnych grup związków. W pierwszym kroku, skonfrontowałam częstość oddziaływań pomiędzy związkami aktywnymi, a nieaktywnymi wobec danego receptora. Następnie, analogiczne statystyki wyznaczyłam dla związków selektywnych i nieselektywnych (przykładowe wyniki badań selektywności pomiędzy receptorem 5-HT<sub>7</sub> a 5-HT<sub>6</sub> przedstawiono na Rysunku 6).

Wykorzystując uzyskane wyniki, przeprowadziłam procedurę wirtualnego poszukiwania nowych ligandów receptora 5-HT<sub>7</sub> w komercyjnych bazach związków (filtrowaniu poddano bazę ZINC zawierającą 13 211 398 struktur), wykorzystując zarówno narzędzia ze ścieżki *ligand-based* (uczenie maszynowe, 3 różne reprezentacje związków), jak i *structure-based* (dokowanie). Ostatecznie zakupiono 12 związków, z których dwa

charakteryzowały się powinowactwem do receptora 5-HT<sub>7</sub> (wyrażonym przez parametr  $K_i$ ) poniżej 5  $\mu$ M. Dodatkowo, jeden ze znalezionych ligandów ( $K_{i\ 5\text{-HT}_7\text{R}} = 3.7\ \mu\text{M}$ ), posiadał znaczną selektywność w stosunku do receptora 5-HT<sub>1A</sub> (współczynnik selektywności równy 54,  $K_{i\ 5\text{-HT}_{1A}\text{R}} = 195.6\ \mu\text{M}$ ). Całość badań przeprowadzono w interdyscyplinarnym zespole złożonym z naukowców z Instytutu Farmakologii im. Jerzego Maja PAN (Sabina Podlewska, Rafał Kafel, Grzegorz Satała, Andrzej J. Bojarski, Stefan Mordalski), Wydziału Farmaceutycznego Uniwersytetu w Bari (Enza Lacivita, Marcello Leopoldo) oraz z Katedry Chemii i Nauk Medycznych Wydziału Nauk Przyrodniczych działającego na Wolnym Uniwersytecie w Amsterdamie (Albert J. Kooistra, Márton Vass, Chris de Graaf).



**Rysunek 6.** Różnica w częstości oddziaływań z poszczególnymi aminokwasami dla związków o dualnej aktywności 5-HT<sub>7</sub>-5-HT<sub>6</sub> (kolor granatowy), a związkami selektywnie aktywnymi wobec receptora 5-HT<sub>7</sub> (kolor czerwony). Kolorem pomarańczowym podświetlono pozycje, dla których różnica w częstości oddziaływania była większa niż 10% [H2].

Określenie determinantów aktywności przeprowadziłam również dla bardzo szczególnej grupy ligandów receptorów serotoniny, a mianowicie tych związków, które charakteryzują się niską zasadowością [H8]. Parametr ten jest niezwykle istotny z punktu widzenia projektowania leków działających poprzez modulowanie aktywności aminergicznych receptorów GPCR, gdyż umożliwia pozbycie się problemu kardiotoksyczności, która niezwykle często charakteryzuje typowe ligandy aminowe. Zarówno dane mutagenetyczne, jak i doświadczenia modelowania molekularnego, wskazały serię wymagań strukturalnych do posiadania przez związek aktywności wobec receptorów serotoniny, wśród których znajduje się m.in. obecność zasadowego atomu azotu oddziałującego w sprotonowanej formie

z resztą kwasu asparaginowego na pozycji 3x32.<sup>55</sup> Pomimo tego, udało się zidentyfikować również aktywnych związków, które owych wymagań nie spełniają. Zachęca to do prowadzenia poszukiwań niestandardowych ligandów tych białek, charakteryzujących się brakiem zasadowego atomu azotu w strukturze związku.

Aby uniknąć niepewności związanej z modelowaniem homologicznym białek, w przedstawionych badaniach skupiłam się na tych podtypach, dla których w momencie prowadzenia eksperymentów były dostępne struktury krystaliczne, tj. 5-HT<sub>1B</sub>, 5-HT<sub>2A</sub>, 5-HT<sub>2B</sub>, 5-HT<sub>2C</sub>.

W pierwszym etapie, dla utworzonych zestawów związków aktywnych wobec rozpatrywanych receptorów ( $K_i$  poniżej 1000 nM), przeprowadziłam analizę ich zasadowości. Za związki o niskiej zasadowości uznano związki, dla których  $pK_a$  przewidziane w programie InstantJChem<sup>56</sup> przyjmowało wartości poniżej 6, aby zapewnić niski ułamek protonowanych cząsteczek w fizjologicznym pH. Zestawienie liczby ligandów poszczególnych receptorów o różnym charakterze zestawiono w Tabeli 2.

**Tabela 2.** Zestawienie liczby ligandów charakteryzujących się niską ( $pK_a < 6$ ) i wysoką ( $pK_a > 8$ ) zasadowością. W nawiasach podano procent jaki stanowi określona grupa związków z całkowitej liczby ligandów danego receptora.

	5-HT <sub>1B</sub>	5-HT <sub>2A</sub>	5-HT <sub>2B</sub>	5-HT <sub>2C</sub>
<b>Niska zasadowość</b>	8 (1,04%)	228 (7,66%)	229 (22,50%)	329 (17,06%)
<b>Wysoka zasadowość</b>	584 (75,75%)	2005 (67,35%)	613 (60,22%)	1258 (65,25%)
<b>Całkowita liczba ligandów</b>	771	2977	1018	1928

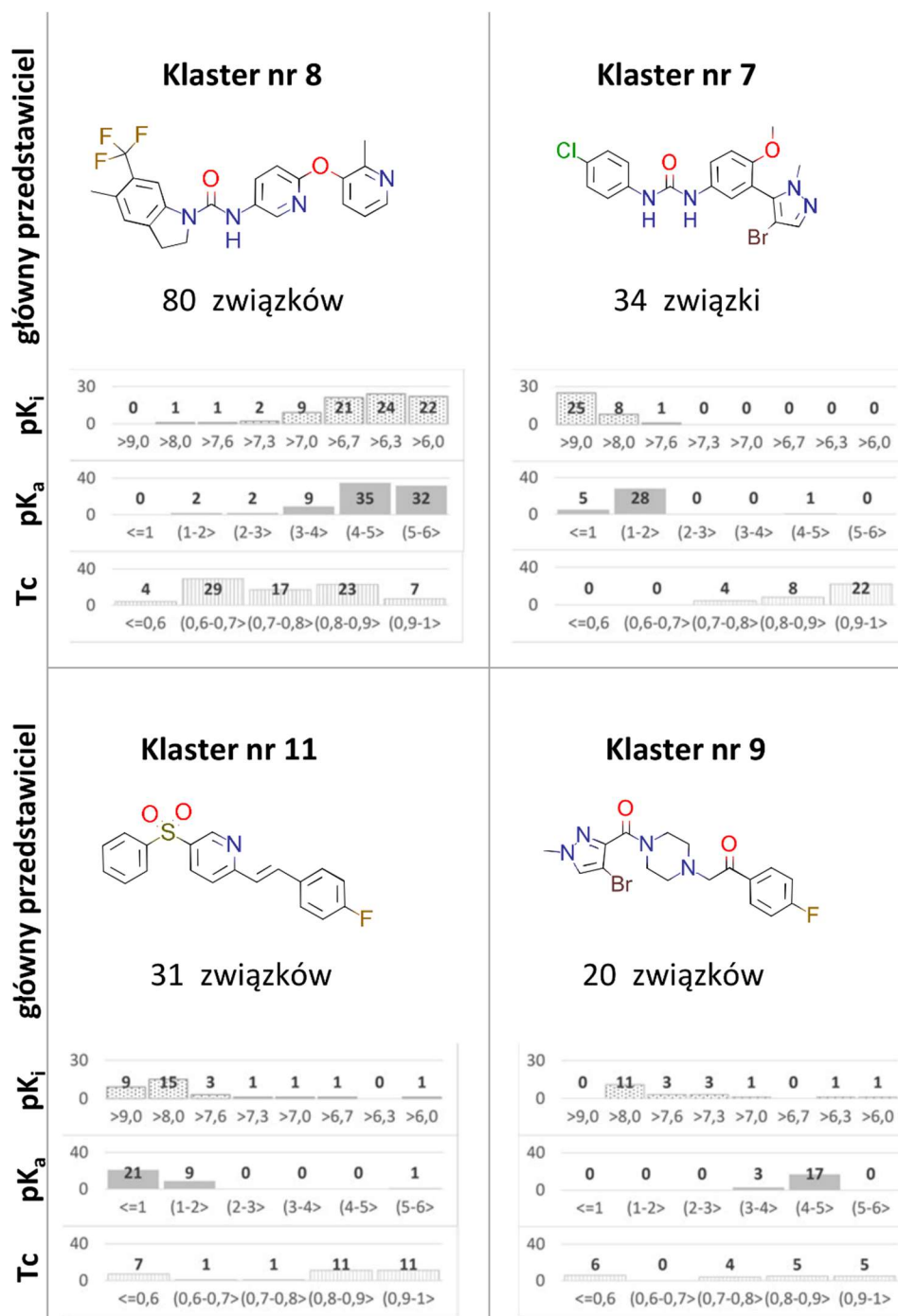
Związki niskozasadowe stanowią stosunkowo duży odsetek ligandów receptora 5-HT<sub>2B</sub> (22,5%) oraz 5-HT<sub>2C</sub> (17,06%), choć w przypadku porównywania liczb bezwzględnych, niskozasadowych ligandów receptora 5-HT<sub>2C</sub> jest dokładnie o 100 więcej niż analogicznych związków wiążących się z receptorem 5-HT<sub>2B</sub> (329 vs. 229).

Z uwagi na relatywnie dużą różnorodność strukturalną ligandów wchodzących w skład podzbiorów związków o niskiej zasadowości, zostały one poklastrowane (klastrowanie ręczne przeprowadzone przez dr. Ryszarda Bugno), a analizę ich własności ( $pK_i$ , przewidziane  $pK_a$  oraz współczynnik podobieństwa Tanimoto do głównego reprezentanta klastra) przeprowadziłam dla poszczególnych klastrów niezależnie (Rysunek 7 przedstawia przykład dla najliczniejszych klastrów receptora 5-HT<sub>2A</sub>). Wykazałam, że spójne strukturalnie grupy

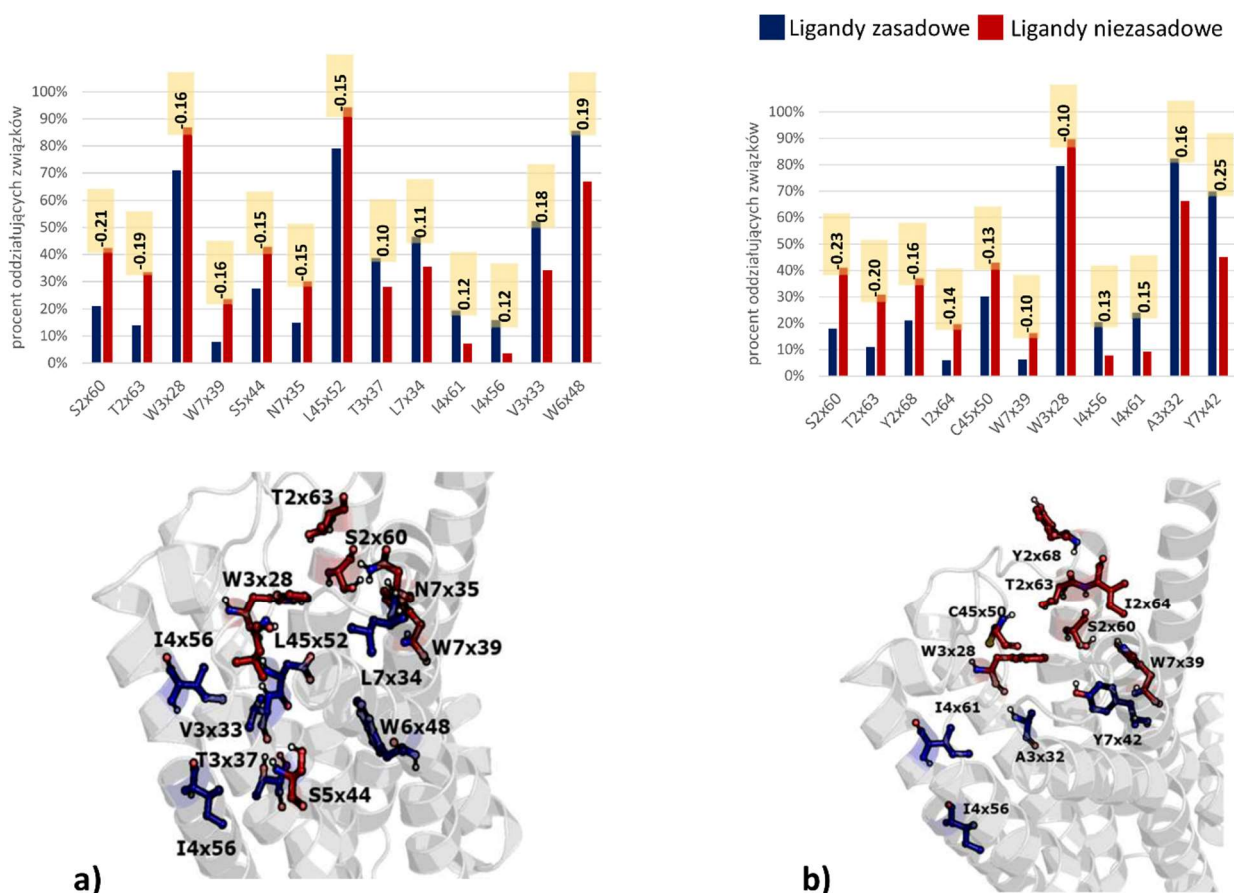
związków są również spójne pod względem własności i aktywności wobec receptora 5-HT<sub>2A</sub>. Przykładowo, związki należące do klastra nr 8, mają w znakomitej większości pK<sub>i</sub> poniżej 6,7, podczas gdy w przypadku klastra nr 8, wartości te są powyżej 7,6 dla wszystkich rozpatrywanych struktur. Podobne zależności widoczne są podczas analizy pK<sub>a</sub> – dla klastra nr 8 większość związków ma przewidziane pK<sub>a</sub> w przedziale 4–6, podczas gdy bardzo niska zasadowość związków należących do klastra nr 7 jest wyrażona przez przyjmowanie przewidywanych wartości pK<sub>a</sub> poniżej 2. Ponadto, związki należące do klastra nr 7 są znacznie bardziej spójne strukturalnie niż te, które reprezentują klaster nr 8 (wartości współczynnika Tanimoto w klastrze nr 7 są dla większości przykładów w przedziale 0,9–1, a dla klastra nr 8: 0,6–0,8).

Kolejnym etapem prac było przeprowadzenie procedury dokowania, którą zrealizowano na dwa sposoby: z wykorzystaniem struktur krystalicznych rozpatrywanych receptorów oraz używając przygotowanych przez mnie zmutowanych wersji każdego z tych białek (podmiana D<sup>3x32</sup>A). Zastosowana mutacja punktowa wynika z charakteru oddziaływania związków zasadowych, które w sprotonowanej formie tworzą wiązanie wodorowe z przeniesieniem ładunku z kwasem asparaginowym D3x32.

Dokonałam reprezentacji wyników dokowania w postaci fingerprintów typu SIFt oraz przeprowadziłam szczegółową analizę częstości oddziaływań poszczególnych grup ligandów z aminokwasami celu biologicznego (Rysunek 8). Dzięki tej procedurze, wskazałam aminokwasy, które istotnie częściej oddziałują z ligandami zasadowymi lub niezasadowymi, co stanowi cenną wiedzę do projektowania nowych struktur o określonym profilu aktywności i niskiej zasadowości. W przypadku receptora 5-HT<sub>2A</sub>, pozycje, które najlepiej rozróżniają w dokowaniu ligandy zasadowe od niezasadowych, zarówno dla niezmutowanego, jak i zmutowanego białka to S2x60, T2x63, W3x28, I4x51, 4x56, oraz W7x39.



**Rysunek 7.** Rozkład wartości pK<sub>i</sub>, przewidzianego pK<sub>a</sub>, oraz współczynnika Tanimoto (T<sub>c</sub>) do głównego przedstawiciela klastra dla najliczniejszych niezasadowych klastrów receptora 5-HT<sub>2A</sub> wraz z podaniem liczby związków o wartościach pK<sub>i</sub>, pK<sub>a</sub> i współczynnika Tanimoto do centroidy klastra w określonym przedziale.



**Rysunek 8.** Analiza częstości oddziaływań zasadowych oraz niezasadowych ligandów receptora 5-HT<sub>2A</sub> z poszczególnymi aminokwasami dla a) receptora natywnego, b) receptora z wprowadzoną podmianą D<sup>3x32</sup>A. Różnice w częstości oddziaływań badanych grup związków zaznaczono kolorem pomarańczowym, położenie najbardziej dyskryminujących aminokwasów zaznaczono na kryształach oraz na modelu homologicznym receptora.

[Przygotowanie zestawu danych mutagenetycznych dla aminergicznych receptorów GPCR oraz opracowanie procedury pozwalającej na przewidywanie efektu podmiany aminokwasu dla danego liganda](#)

Podczas pobytu w grupie Medycznej Chemii Komputerowej kierowanej przez Dr. Chrisa de Graafa (02.11.2015–01.03.2016, Wolny Uniwersytet w Amsterdamie), rozpocząłam prace nad bazą danych mutagenetycznych odnoszących się do aminergicznych receptorów GPCR, które kontynuowałam jeszcze przez okres około 2.5 roku po zakończeniu stażu.

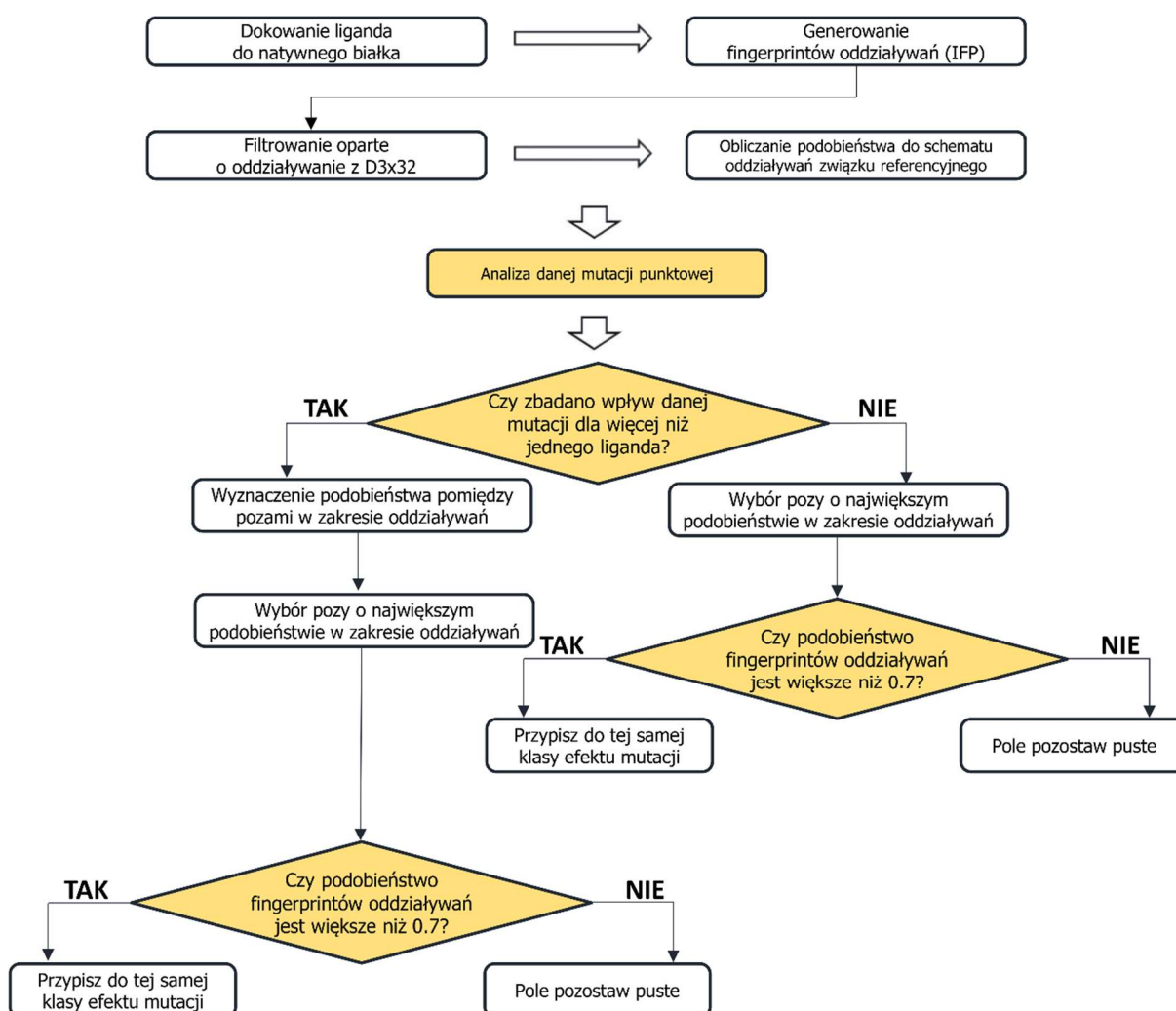
Pierwszym etapem badań było zebranie obecnych w literaturze danych mutagenetycznych w obrębie rozpatrywanych receptorów, tj. wpływu podmiany danego aminokwasu na aktywność określonego liganda wobec rozpatrywanego białka [H3]. Prace



te realizowałam wspólnie z Dr. Mártonem Vassem z Amsterdamu, a moją rolą było przygotowanie informacji mutagenetycznych dla następujących grup receptorów: serotoninowe, dopaminowe, muskarynowe oraz alfa-adrenergiczne. Przygotowany zestaw danych dla wszystkich rozpatrywanych białek zawierał 5445 rekordów mutacja-ligand. Zebrane informacje poddano różnym analizom statystycznym i wizualizacjom. Aby ułatwić interpretację wyników, efekty mutacji (wyrażone jako stosunek: aktywność liganda w zmutowanym białku/aktywność liganda w białku natywnym) podzielono na cztery grupy w zależności od wielkości i kierunku zmian indukowanych przez daną podmianę aminokwasową:  $< 0,3$  (wzrost powinowactwa),  $0,3-3$  (brak wyraźnego wpływu),  $3-10$  (obniżenie powinowactwa),  $> 10$  (znaczne obniżenie powinowactwa).

Z uwagi na fakt, że zebrane dane były stosunkowo wybiórcze, tj. wpływ danej mutacji zbadano w sposób doświadczalny jedynie dla wybranych ligandów, opracowałam protokół do przewidywania efektu mutagenetycznego dla brakujących danych (Rysunek 9, **H3**). Opiera się on na dokowaniu badanych związków do rozpatrywanego receptora oraz na reprezentacji uzyskanych kompleksów ligand-receptor w postaci fingerprintu oddziaływań (ang. *Interaction Fingerprint, IFP*).<sup>57</sup> Następnie, z uwagi na występowanie w przygotowanej bazie danych mutagenetycznych głównie ligandów zasadowych, przeprowadzono filtrowanie pól w oparciu o występowanie oddziaływania z resztą kwasu asparaginowego w pozycji D3x32 (odrzucono te pozycje, w których nie ma obecnej interakcji ligand-D3x32). Jeśli wpływ danej mutacji punktowej zbadano jedynie dla jednego liganda (stanowiącego tzw. ligand referencyjny) sprawdzano, czy podobieństwo ułożenia badanego związku i tego liganda jest powyżej ustalonego poziomu odcięcia (podobieństwo pól oceniane jako współczynnik Tanimoto wyznaczany dla wygenerowanych fingerprintów oddziaływań). W przypadku podobieństwa większego niż ustalony próg odcięcia (0,7), efekt mutagenetyczny badanego związku przypisywano do tej samej klasy efektu (wzrost powinowactwa, brak wyraźnego wpływu, obniżenie powinowactwa, znaczne obniżenie powinowactwa), co związek odniesienia. Z kolei gdy podobieństwo wyznaczone na fingerprintach oddziaływań było niższe niż 0,7, pole pozostawiano puste (stwierdzano, że wpływ danej mutacji nie może być oceniony z należytą wiarygodnością). W sytuacji gdy dla danej podmiany aminokwasowej dostępne są dane o jej wpływie na powinowactwo więcej niż jednego związku, wybierany jest ligand o największym podobieństwie w sposobie interakcji z celem biologicznym (współczynnik Tanimoto dla reprezentacji IFP) do badanego przykładu i przypisywana jest jego klasa wpływu mutacji (z zastrzeżeniem o konieczności przyjmowania przez współczynnik podobieństwa wartości wyższych niż 0,7, w przeciwnym wypadku, pole pozostaje puste).

Otrzymane rezultaty stanowią cenną wiedzę do analizy wyników uzyskiwanych w ścieżce *structure-based* i pozwalają na identyfikację pozycji aminokwasowych kluczowych do wiązania ligandów z określonym białkiem. Opracowany protokół do przewidywania efektu mutagenetycznego umożliwia z kolei zbadanie wpływu mutacji dla związków pokrywających szeroką przestrzeń chemiczną. Dzięki temu możliwe jest dostosowanie wymagań w zakresie oddziaływań z danym receptorem do struktury badanych kandydatów na nowe ligandy [H3]. Przygotowany zestaw danych mutagenetycznych stał się częścią bazy GPCRdb.<sup>43</sup>



**Rysunek 9.** Schemat protokołu do przewidywania efektu mutagenetycznego dla danej pary mutacja-ligand [H3].



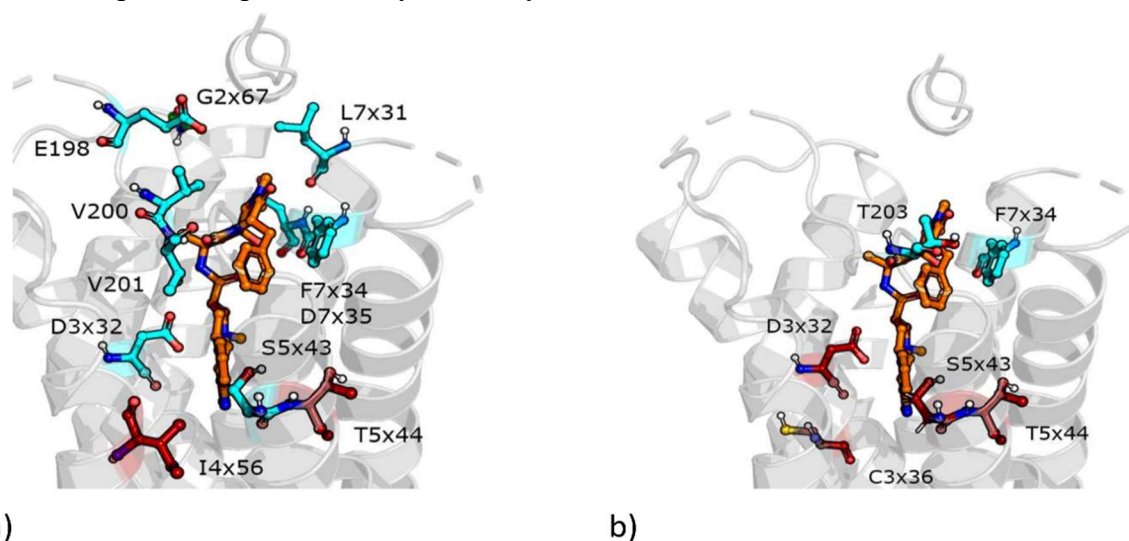
## Opracowanie metody pozwalającej na przewidywanie wyniku dokowania z dwuwymiarowej struktury związku chemicznego

Dokowanie jest jedną z najważniejszych procedur stosowanych podczas poszukiwania nowych kandydatów na leki. Polega ono na poszukiwaniu optymalnego ułożenia badanego związku w miejscu wiążącym celu biologicznego. Możliwość przewidywania wzajemnej konformacji liganda i białka oraz badanie potencjalnych oddziaływań występujących między nimi stanowi niezwykle ważny wkład w proces podejmowania decyzji odnośnie traktowania związku jako aspiranta do bycia przyszłym terapeutycznym. Pomimo znacznej ilości informacji dostarczanej przez dokowanie, jego istotną wadą jest konsumpcja sporych zasobów obliczeniowych, co skutecznie ogranicza jego zastosowanie do oceny wielomilionowych bibliotek związków.

Aby zniwelować zasoby obliczeniowe niezbędne do przeprowadzenia dokowania, a jednocześnie zachować najważniejsze informacje dostarczane w tym procesie, opracowałam procedurę przewidywania oddziaływań występujących pomiędzy ligandem a białkiem przy wykorzystaniu jedynie dwuwymiarowej struktury związku chemicznego [H5]. W tym celu, przygotowałam zestawy związków uczących dla białek, dla których testowano skuteczność skonstruowanych narzędzi, tj. receptory: 5-HT<sub>1B</sub>, 5-HT<sub>2B</sub>, M<sub>2</sub>, oraz  $\beta$ 2-adrenergiczny, a także 4 podtypy cytochromu P450: CYP2C8, CYP2C9, CYP2D6 oraz CYP3A4. Następnie, przeprowadziłam dokowanie tych związków w programie Glide<sup>58</sup> (mgr Maciej Szymczak wykonał analogiczne dokowanie w programie SMINA).<sup>59</sup> Kompleksy ligand-białko otrzymane w obydwu procedurach przekształciłam do postaci zero-jedynkowej, z wykorzystaniem fingerprintu IFP.<sup>57</sup> Uzyskane w ten sposób ciągi bitowe stanowiły zestaw danych wejściowych do trenowania modelu uczenia maszynowego. W badaniach zastosowano 4 różne algorytmy: regresję logistyczną, perceptron wielowarstwowy (ang. *Multilayer Perceptron, MLP*),<sup>60</sup> grafowe sieci konwolucyjne (ang. *Graph Convolutional Networks, GCNs*) oraz sieci grafowe typu EAGCN (*edge-attention graph convolutional network*).<sup>61</sup> Regresja logistyczna oraz MLP wykorzystuje w procesie uczenia struktury związków reprezentowane przez fingerprint Morgana, z kolei sieci grafowe bazują na związku chemicznym przedstawionym jako graf molekularny (trenowanie modeli przeprowadzili badacze z Wydziału Matematyki i Informatyki Uniwersytetu Jagiellońskiego: Stanisław Jastrzębski, Maciej Szymczak, Agnieszka Pocha). Informacją zwracaną ostatecznie przez każdy z modeli jest ciąg bitowy, będący symulowanym fingerprintem interakcji dla danego związku. Poprzez analizę kolejnych pozycji w takim zero-

jedynkowym ciągu, można prześledzić oddziaływania, których obecność jest przewidywana w kompleksie danego związku z białkiem.

Skuteczność uzyskanych modeli przetestowano w retrospektywnych badaniach na przygotowanych uprzednio zestawach danych, w trybie 5-krotnej walidacji krzyżowej (podział danych w oparciu o klastrowanie na podstawie struktury chemicznej związku). Dokonując analizy wyników predykcji wykazałam, że największą efektywnością charakteryzowały się metody bazujące na grafach, tj., GCN oraz EAGCN, osiągając wartości AUROC w przedziale  $\sim 0,7-0,8$ . Dla każdego białka wskazałam również aminokwasy, dla których najlepiej oraz najgorzej przewidywano oddziaływania z ligandami (na Rysunku 10 przedstawiono przykład takiej analizy dla receptora 5-HT<sub>1B</sub>). Ponadto, sprawdziłam trudność przewidywania poszczególnych typów oddziaływań – badane metody wykazywały się największą skutecznością w identyfikacji potencjalnych oddziaływań jonowych. Jest to zgodne z intuicją, gdyż do wystąpienia tego rodzaju oddziaływania jest wymagana obecność konkretnych elementów strukturalnych, zarówno w białku, jak i ligandzie, które są stosunkowo łatwe do rozpoznania przez metody automatyczne.



**Rysunek 10.** Wizualizacja aminokwasów w receptorze 5-HT<sub>1B</sub>, dla których najlepiej (cyjan) i najgorzej (czerwony) przewidywano oddziaływania z ligandami dla dokowania przeprowadzonego a) w programie Glide, b) w programie SMINA; kolorem pomarańczowym zaznaczono ligand obecny w kryształach 4IAQ [H5].

Dla potwierdzenia skuteczności opracowanego podejścia w identyfikacji aktywnych związków, przeprowadziłam symulację procesu wirtualnego przesiewania bibliotek chemicznych, wykorzystując 500 000 losowo wybranych związków z bazy ZINC oraz 100 000 losowo wybranych struktur z komercyjnej biblioteki związków (ChemBridge). Zbiory związków z bazy ZINC oraz ChemBridge połączono ze znanymi związkami aktywnymi wobec rozpatrywanych celów biologicznych i sprawdzano efektywność identyfikacji ligandów

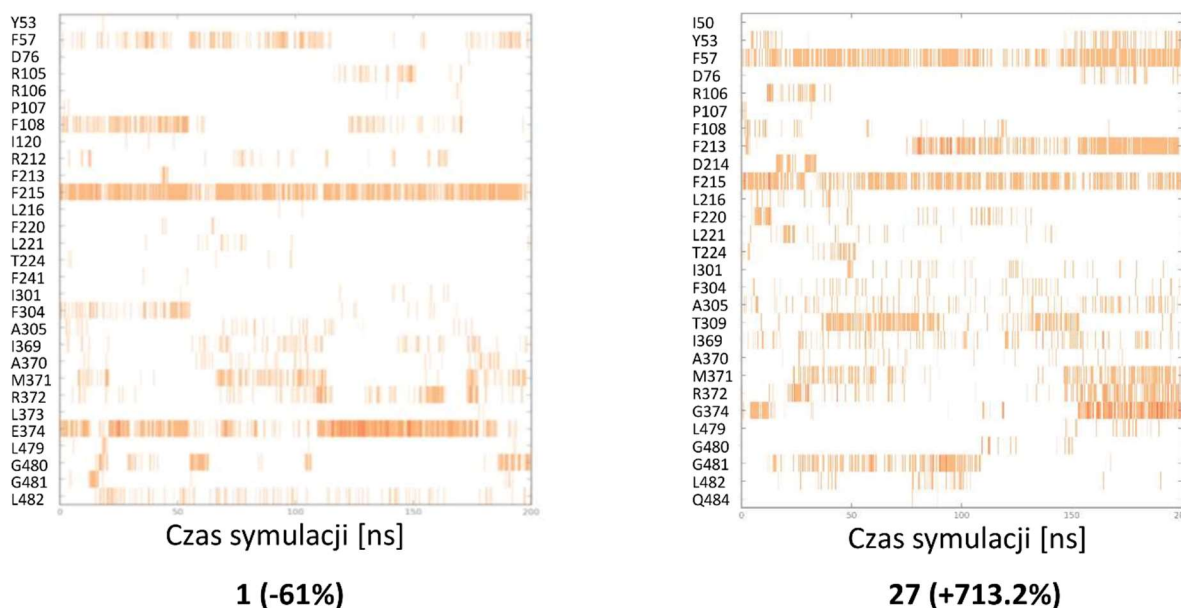
na podstawie przewidzianych fingerprintów oddziaływań (ocena związków przez model przeprowadzona przez Macieja Szymczaka i Stanisława Jastrzębskiego). Wykazano nie tylko bardzo wysoką skuteczność prowadzonej w ten sposób wirtualnej oceny (współczynnik wzbogacenia przyjmował wartości nawet powyżej 60 w przypadku eksperymentów ze związkami z bazy ChemBridge), lecz również znacznie krótszy czas obliczeń w porównaniu z klasycznym dokowaniem (dokowanie związków w bazy ChemBridge zajęło ok. 188 godzin, podczas gdy ich ocena przez modele GCN i EAGCN trwała ok. 12 godzin, przy wykorzystaniu tych samych jednostek obliczeniowych).

### Konstrukcja narzędzi do automatycznej analizy wyników dynamiki molekularnej

Pomimo ograniczeń wynikających z wymagań z zakresu mocy obliczeniowej, dynamika molekularna jest coraz częściej wykorzystywaną procedurą w procesie projektowania nowych leków. Choć wykorzystanie jej na szeroką skalę w przesiewowych kampaniach poszukiwania nowych kandydatów na leki nie jest możliwe, symulacje dynamiki molekularnej służą często wyjaśnianiu aktywności związków poprzez dostarczanie szczegółowych informacji odnośnie sposobu ich oddziaływania z celem biologicznym.

Rezultatem symulacji dynamiki molekularnej jest bardzo duża ilość danych, dlatego też istnieje wiele programów umożliwiających przeprowadzenie analizy jej wyników. W swoich badaniach skupiłam się na opracowaniu metodologii do poszukiwania zależności pomiędzy zachowaniem związków podczas symulacji dynamiki molekularnej, a ich aktywnością i własnościami.

Prace w tym zakresie rozpocząłam od analizy serii nieimidazolowych antagonistów receptora histaminowego H<sub>3</sub> w kontekście ich oddziaływania z cytochromem CYP3A4 [H7]. Badany zestaw związków był złożony z 27 reprezentantów, charakteryzujących się różnym wpływem na CYP3A4 – od inhibicji (-61% aktywności referencyjnej) do aktywacji (+713% aktywności referencyjnej). Dla każdego ze związków przeprowadziłam dokowanie do miejsca aktywnego CYP3A4, a otrzymane kompleksy stanowiły punkt wyjścia do symulacji dynamiki molekularnej (czas trwania: 200 ns). Analiza uzyskanych diagramów interakcji (przykładowe analizy dla dwóch związków o skrajnych aktywnościach wobec CYP3A4 przedstawiono na Rysunku 11) nie pozwoliła na powiązanie sposobu interakcji z poszczególnymi aminokwasami z obserwowanymi aktywnościami *in vitro*. Pomimo zauważalnej różnicy w oddziaływaniach związku **1** i **27** z CYP3A4 (np. znacznie bardziej intensywny kontakt związku **27** z F213, a także częstsze oddziaływanie **27** z F57, T309 oraz G481 w porównaniu



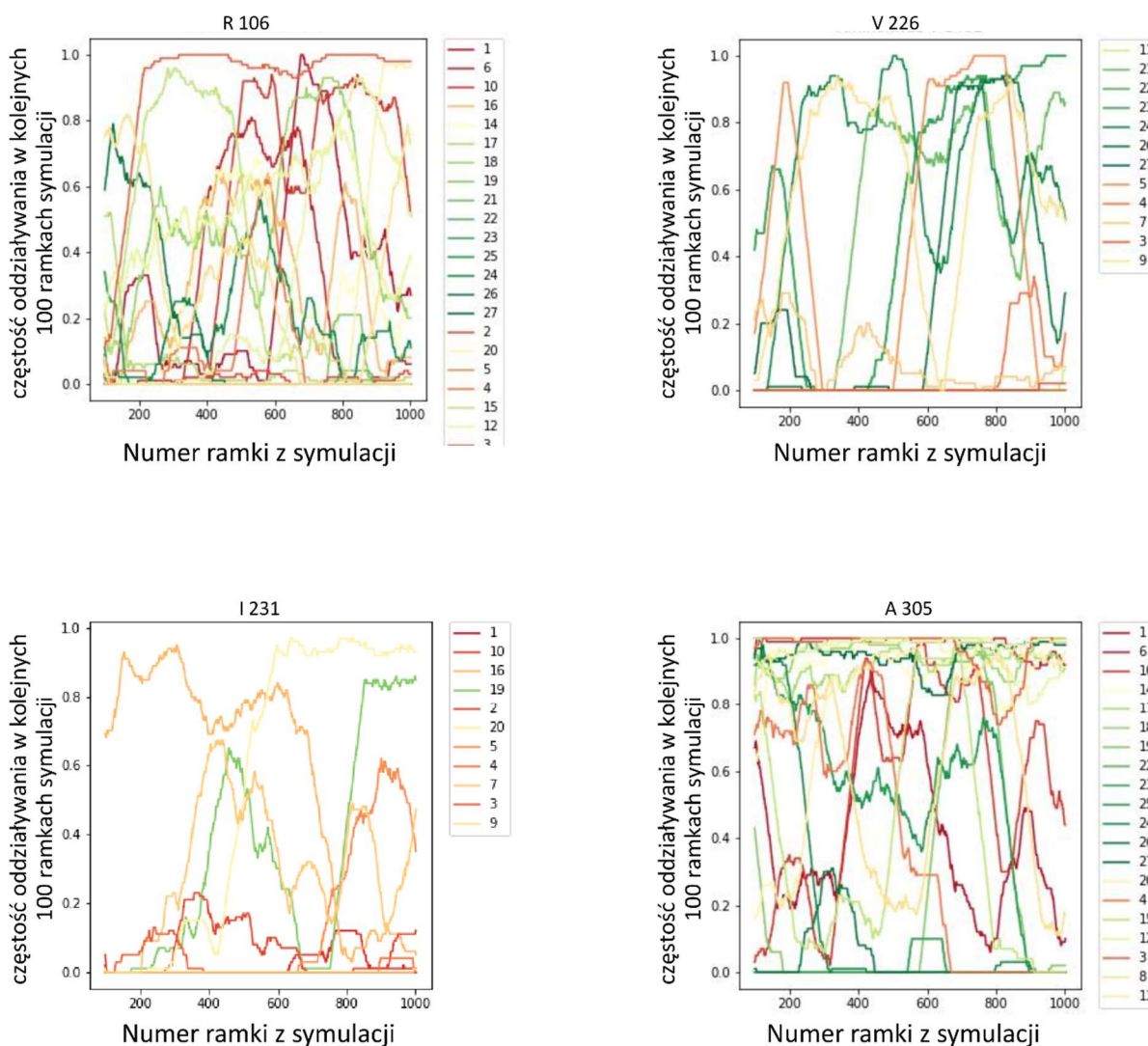
**Rysunek 11.** Przykład oddziaływania dwóch badanych związków z CYP3A4 o różnej aktywności wobec celu biologicznego (–61% inhibicji aktywności referencyjnej oraz wzrost aktywności o 713.2% w porównaniu z aktywnością referencyjną) w czasie symulacji [H7].

z **1**; z drugiej strony, częstsza interakcja związku **1** z GLU374), nie było możliwe określenie zależności schemat oddziaływań-aktywność biologiczna dla całego zestawu badanych związków.

Kolejnym etapem analizy było wyznaczenie korelacji pomiędzy częstością oddziaływań związków z poszczególnymi aminokwasami CYP3A4, a ich wpływem na aktywność tego enzymu. Dla reszt aminokwasowych o najwyższych współczynnikach korelacji (tj. R106, V226, I231 oraz A305), wykonałam szczegółowe wizualizacje zmian w przebiegu ich oddziaływania z ligandami w czasie (Rysunek 12).

Uzyskane wyniki pozwolą na optymalizację kolejnych związków o strukturach chemicznych podobnych do analizowanych związków pod kątem aktywności wobec CYP3A4, co ma szczególne znaczenie w kontekście ich stabilności metabolicznej.

Powiązanie intensywności oddziaływań ligandów z poszczególnymi aminokwasami z aktywnością związków było również podstawą badań, które prowadziłam w obszarze receptorów opioidowych [H6]. Podczas tych prac, eksperymentom poddawałam 4 związki: PZM21 oraz SR-17018 (agoniści receptora  $\mu$ -opiodowego działający przez selektywną aktywację białka G), niewykazująca selektywności funkcjonalnej morfina oraz fentanyl (działający poprzez selektywną aktywację beta-arestyny). Oddziaływania tych ligandów z receptorem  $\mu$ -opiodowym konfrontowałam z wynikami eksperymentów *in vitro*, jednak tym razem nie były to doświadczenia weryfikujące powinowactwo do celu biologicznego, lecz



**Rysunek 12.** Częstość oddziaływania badanych związków (numery związków po prawej stronie) w czasie kolejnych 100 ramek symulacji. Kolorem czerwonym zaznaczono związki hamujące aktywność CYP3A4, kolorem zielonym – związki zwiększające aktywność tego enzymu [H8].

następujące testy funkcjonalne: aktywacja Gai2, inhibicja cAMP, rekrutacja beta-arestyny 2, monitorowanie GTPazy Rab5, aktywacja GIRK, rekrutacja GRK2. W badaniach wykorzystałam 3 struktury krystaliczne receptora  $\mu$ -opiodowego: 4DKL (przedstawiający receptor w konformacji nieaktywnej), 5C1M oraz 6DDF (obydwa wykrystalizowane z agonistami i reprezentujące receptor w jego aktywnej konformacji).

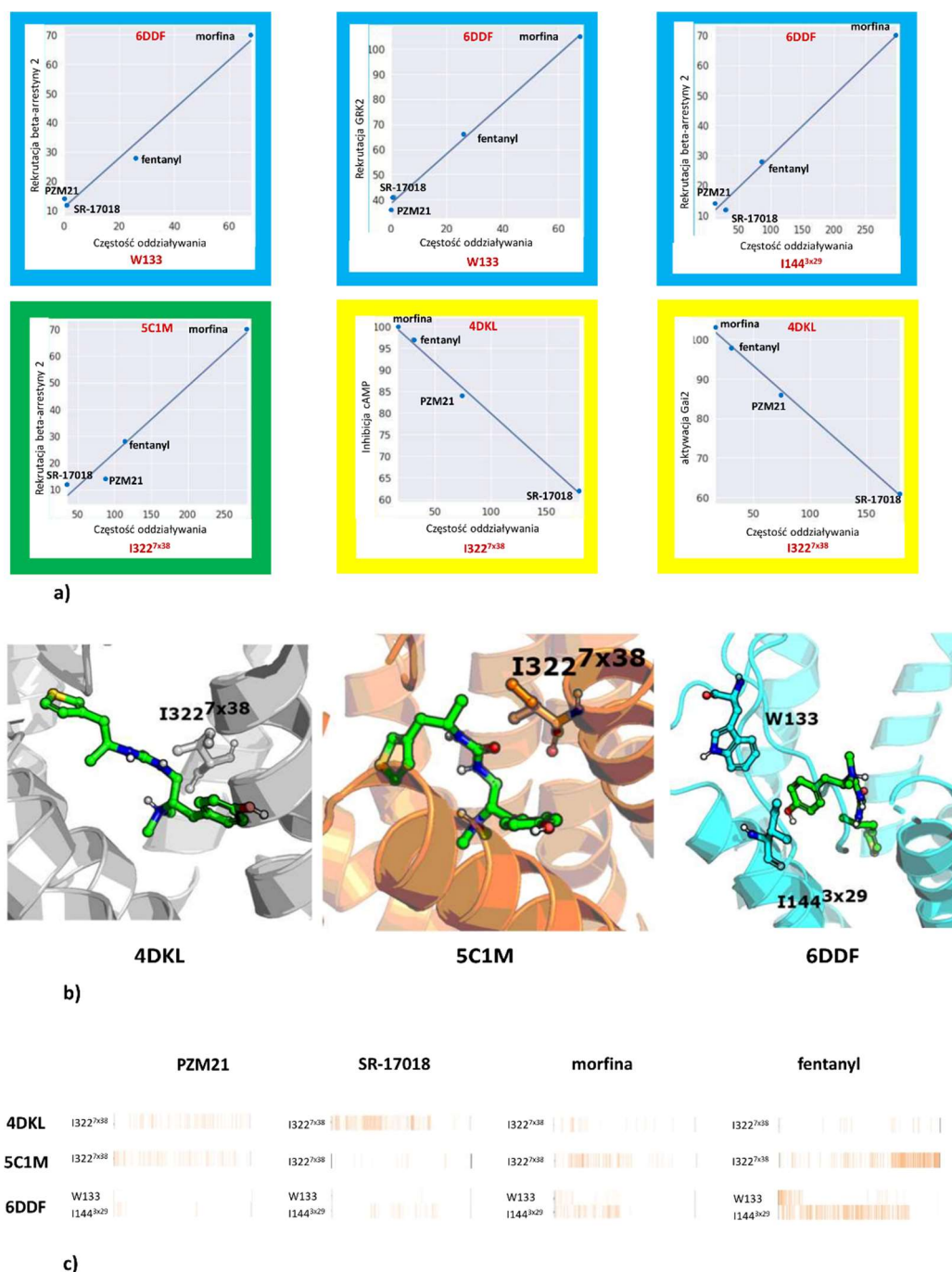
Podobnie jak w przypadku badań z CYP3A4, punkty startowe do symulacji dynamiki molekularnej stanowiły kompleksy ligand-receptor, które uzyskałam w procedurze dokowania. Długość symulacji w przypadku badań receptorów opiodowych była jednak znacznie większa, gdyż każda z symulacji obrazowała oddziaływania liganda z receptorem w czasie 2000 ns. Następnie, dla interakcji ligandów z każdym aminokwasem receptora wyznaczyłam korelacje



pomiędzy częstotliwością kontaktu w trakcie symulacji, a rezultatami badań *in vitro*. Dla kilku pozycji uzyskano bardzo wysokie wartości współczynnika korelacji Pearsona, powyżej 0,98 (Rysunek 13). Analizując diagramy oddziaływań dla wybranych pozycji (Rysunek 13c) widać wyraźną tendencję do zmian w częstości oddziaływań z prezentowanymi aminokwasami. W zależności od analizowanego doświadczenia *in vitro* jest to spadek lub wzrost częstości interakcji wraz ze wzrostem wartości wyniku otrzymanego w eksperymencie.

Otrzymane wyniki stanowią cenną wiedzę wspierającą projektowanie kolejnych ligandów działających w obszarze receptorów opioidowych, zwłaszcza takich, które wykazują selektywność funkcjonalną (tj. aktywują tylko białka G lub beta-arestyny). Obecnie poszukiwanie związków działających selektywnie poprzez wybraną ścieżkę sygnałową jest niezwykle pożądane z uwagi na potencjalne możliwości uwolnienia się od efektów ubocznych związanych ze stosowaniem klasycznych opioidowych leków przeciwbólowych (takich jak euforia czy depresja ośrodka oddechowego). Szanse te są skutkiem odkryć poczynionych w ostatnich latach, które wskazują, że działanie przeciwbólne jest wynikiem aktywacji białek G, natomiast efekty niepożądane są rezultatem aktywacji szlaku beta-arestyny.<sup>62</sup>

Symulacje dynamiki molekularnej pozwoliły mi również na przeprowadzenie badań nad kinetyką wybranych ligandów receptora serotoninowego 5-HT<sub>7</sub> [H11]. Podczas tych prac, starałam się powiązać czas przebywania związku w receptorze (ang. *residence time*, *RT*) z jego sposobem interakcji z celem biologicznym. Eksperymentom obliczeniowym poddałam 12 związków wykazujących aktywność wobec receptora serotoninowego 5-HT<sub>7</sub>, pochodnych znanego agonisty tego białka – LP-211; badania *in vitro* przeprowadzili badacze z Wydziału Farmaceutycznego Uniwersytetu w Bari oraz naukowcy z Uniwersytetu Neapolitańskiego im. Fryderyka II.



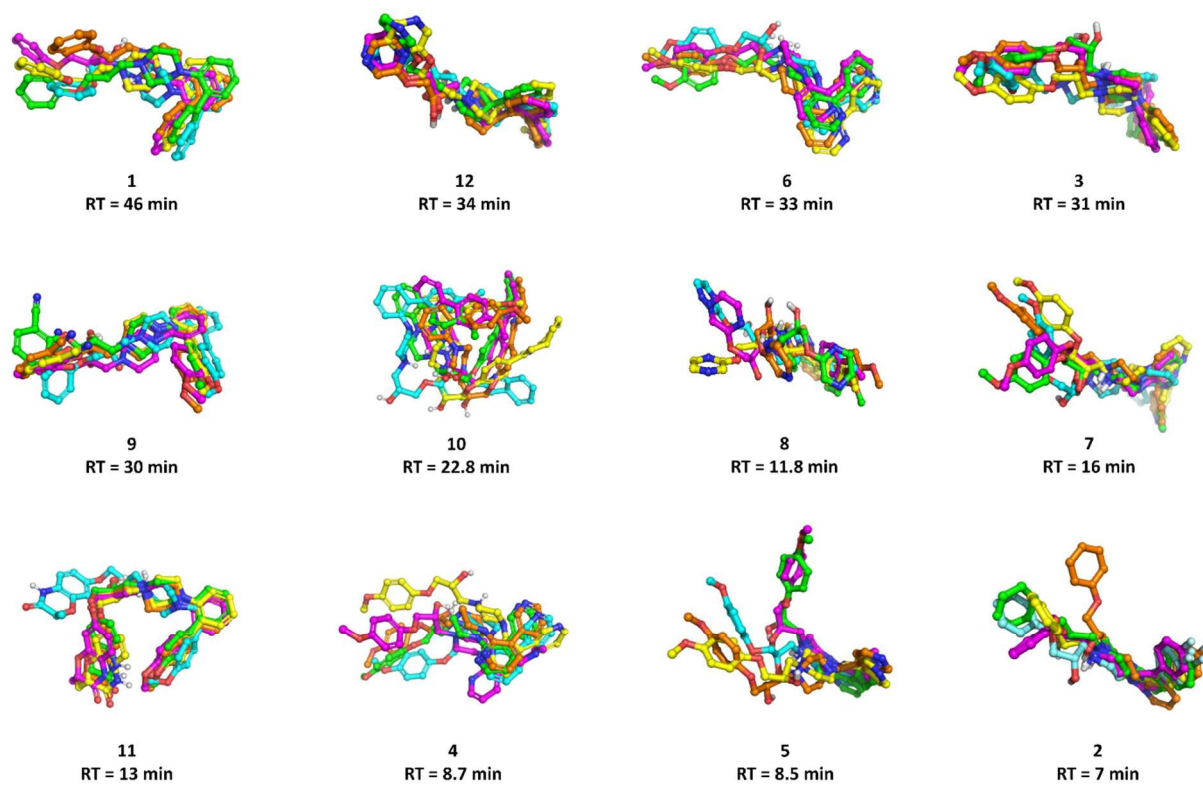
**Rysunek 13.** Wyniki eksperymentów korelacyjnych pomiędzy częstością oddziaływań badanych związków z poszczególnymi aminokwasami a wynikami badań *in vitro* (prezentowane są wyniki o najwyższych wartościach współczynnika korelacji Pearsona): a) wykresy korelacyjne, kolor krawędzi różnicuje poszczególne struktury krystaliczne: 4DKL – żółty, 5C1M – zielony, 6DDF – niebieski; b) wizualizacja pozycji najsilniej skorelowanych reszt aminokwasowych na odpowiednich kryształach, c) wizualizacja diagramów oddziaływań ligand-białko dla najsilniej skorelowanych reszt aminokwasowych

W pierwszym etapie badań sprawdziłam powiązanie pomiędzy wartościami zestawu deskryptorów molekularnych (deskryptory wygenerowałam przy wykorzystaniu oprogramowania PaDEL-Descriptor), a RT związków, jednak nie uzyskałam żadnej statystycznie istotnej korelacji. Następnie przeprowadziłam dokowanie do modeli homologicznych receptora 5-HT<sub>7</sub> (używając zarówno aktywnej, jak i nieaktywnej formy receptora), a otrzymane kompleksy ligand-receptor wykorzystałam jako punkty początkowe do symulacji dynamiki molekularnej o czasie trwania 1000 ns. Podobnie jak w przypadku wcześniejszych badań, wyznaczyłam korelacje pomiędzy częstością oddziaływań ligandów z każdym z aminokwasów białka, a ich RT, jednak nie uzyskałam żadnego istotnego powiązania. Bardziej szczegółowym badaniom poddałam aminokwasy należące do pętli zewnątrzkomórkowych białka, z uwagi na doniesienia o ich wpływie na własności kinetyczne związków,<sup>63-65</sup> jednak zarówno szczegółowe pomiary odległości, kątów, jak i względnego położenia wybranych reszt aminokwasowych również nie przyniosły wyjaśnień dla obserwowanych własności kinetycznych ligandów.

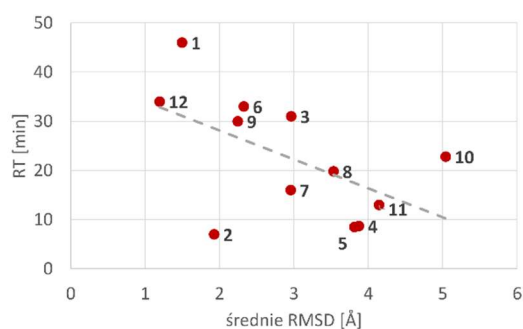
Czas przebywania liganda w miejscu wiążącym receptora okazał się pozostawać natomiast w ścisłym związku ze stabilnością ułożenia związku podczas symulacji. Obserwację tę, początkowo jakościową (Rysunek 14a), potwierdziłam również wyznaczeniem współczynnika korelacji Pearsona pomiędzy RT, a średnim RMSD liganda (Rysunek 14b), który wyniósł  $-0,552$ .

Z uwagi na rosnące znaczenie metod dynamiki molekularnej w procesie projektowania nowych leków, a także konieczności rozwoju narzędzi pozwalających na analizę ogromnej ilości danych produkowanych podczas symulacji, przygotowałam (we współpracy z Panią Hanną Baltrukevich z Wydziału Farmaceutycznego Uniwersytetu Jagiellońskiego) publikację przeglądową na temat zastosowania metod automatycznych w przetwarzaniu wyników dynamiki molekularnej [H10]. Publikacja charakteryzuje zarówno oprogramowanie umożliwiające ekstrakcję różnego rodzaju informacji z symulacji, jak również przybliża i podaje przykłady wykorzystania metod uczenia maszynowego do analiz danych wyjściowych z dynamiki. Opiswane są metody zarówno z podejścia nadzorowanego (ang. *supervised*), jak i nienadzorowanego (ang. *unsupervised*), do którego zalicza się m.in. klastrowanie będące jedną z podstawowych procedur przetwarzania danych wynikowych z symulacji.





a)



b)

**Rysunek 14.** Wyniki analizy stabilności ułożenia ligandów w miejscu wiążącym białka w odniesieniu do ich czasu przebywania w receptorze, a) ułożenie ligandów podczas wybranych momentów symulacji: poza początkowa – cyan, ułożenie po upływie 250 ns – żółty, ułożenie po upływie 500 ns – pomarańczowe, ułożenie po upływie 750 ns – fuksja, ułożenie po upływie 1000 ns – zielony, b) wykres zależności RT związku od jego średniego RMSD podczas symulacji [H11]

Opracowanie podejścia do interpretowania przewidywań uzyskiwanych przy wykorzystaniu metod uczenia maszynowego, umożliwiającego dokonanie optymalizacji związku chemicznego pod kątem rozpatrywanego parametru

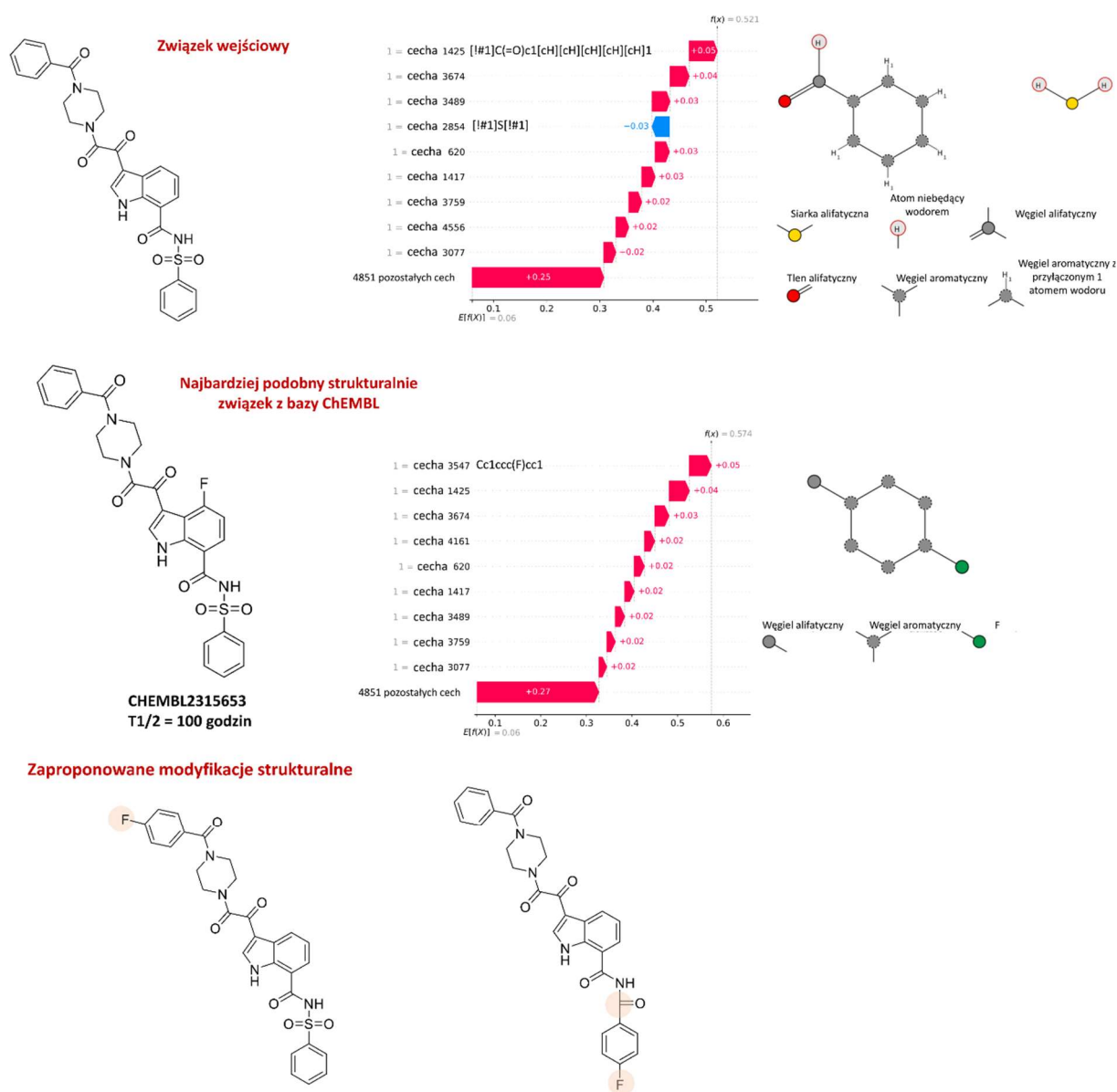
Metody uczenia maszynowego wykorzystuje się nie tylko do oceny potencjalnej aktywności związków wobec rozpatrywanego celu biologicznego, ale również do oceny własności fizykochemicznych, farmakokinetycznych, a także potencjalnej toksyczności. Ponadto, pożądane jest aby różnorodne narzędzia dostarczały nie tylko informacji o wartości ocenianego parametru, ale również podawały wskazówki pozwalające na optymalizację związku pod kątem danej własności.

Podążając za tymi potrzebami, opracowałam metodę do interpretacji wyników przewidywań dostarczanych przez algorytmy uczące [H9]. Skupiłam się w niej nie na ocenie aktywności związków chemicznych, lecz na ocenie stabilności metabolicznej. Wybór ten był podyktowany głównie koniecznością konstruowania nowych narzędzi do oceny tego parametru. Wynika to z faktu, że stabilność metaboliczna jest niezwykle trudną własnością do oceny obliczeniowej, co z kolei jest rezultatem niezwykle złożoności procesów transformacji ksenobiotyków zachodzących w organizmach żywych. Z tego względu, liczba dostępnych metod *in silico* do oceny stabilności związków jest stosunkowo niewielka.

W swoich badaniach zaprojektowałam metodę do oceny stabilności metabolicznej (wyrażonej przez czas półtrwania) związków chemicznych w podejściu *ligand-based*, która wykorzystuje trzy fingerprinty podstrukturalne do reprezentacji związków oraz następujące algorytmy uczenia maszynowego: naiwny model bayesowski, drzewa, oraz maszyny wektorów nośnych. Przygotowane modele (trenowanie modeli przeprowadziła Agnieszka Wojtuch z Wydziału Informatyki i Matematyki Uniwersytetu Jagiellońskiego) były optymalizowane zarówno pod kątem zadań klasyfikacyjnych (przypisywanie związku do „klas stabilnościowych”: stabilny, niestabilny, o średniej stabilności), jak i regresyjnych (przewidywanie konkretnej wartości czasu półtrwania). Przygotowano niezależne modele dla danych o czasie półtrwania związków uzyskanych w badaniach na ludzkich oraz szczurzych mikrosomach.

Ponadto, zapewniłam interpretowalność uzyskiwanych predykcji poprzez wskazywanie cech strukturalnych o największym wpływie na daną odpowiedź modelu. W tym celu wykorzystane są tzw. wartości SHAP (Shapley Additive exPlanations),<sup>66</sup> które przypisywane są do każdej cechy strukturalnej związku dla danej predykcji i wskazują na wagę danej podstruktury w odpowiedzi modelu (implementację wyznaczania wartości SHAP dla predykcji

modelu zaimplementowała Agnieszka Wojtuch). Dodatkowo, dla każdego związku, dla którego przeprowadzane są przewidywania, zwracany jest najbardziej podobny strukturalnie związek (miarą podobieństwa jest współczynnik Tanimoto) ze zbioru uczącego skonstruowanego na podstawie danych z bazy ChEMBL. Dostarczanie tego rodzaju informacji pozwala na zaplanowanie i przeprowadzenie optymalizacji związków pod kątem ocenianego parametru, przykład tego rodzaju analizy zaprezentowano na Rysunku 15.



**Rysunek 15.** Przykład analizy opartej o wartości SHAP wraz z wizualizacją najbardziej istotnych podstruktur dla danej predykcji (ilustrację kluczy strukturalnych wykonano w programie SMARTS Plus; [SMARTS.plus](#)); przedstawiony model skonstruowano na reprezentacji Klekota-Roth na danych odnoszących się do czasu półtrwania wyznaczonego na ludzkich mikrosomach [H9].

Opracowane narzędzie udostępniono szerokiemu gronu odbiorców w formie platformy online; <https://metstab-shap.matinf.uj.edu.pl/> (interfejs graficzny do opracowanej metodologii przygotował Rafał Jankowski).

#### 4.4.4. Podsumowanie

Rosnąca ilość danych również w dziedzinie komputerowo wspomaganego projektowania leków wymaga stosowania różnych metod automatycznych pozwalających na ich skuteczną analizę oraz ekstrakcję istotnych informacji i identyfikację zależności występujących pomiędzy nimi. Pomimo opracowania szeregu narzędzi oraz intensywnej pracy wielu naukowców w obszarze cheminformatyki, potrzeby rozwijania nowych metodologii w tym zakresie nie maleją, co jest związane głównie ze wzrostem oczekiwań stawianych metodom automatycznym i ciągłym definiowaniem nowych zadań dedykowanych do rozwiązania przez algorytmy uczenia maszynowego lub różnego rodzaju podejścia statystyczne.

W badaniach wchodzących w skład opisywanego osiągnięcia naukowego dotykałam problemów zarówno ze ścieżki *ligand-*, jak i *structure-based*, a także tworzyłam narzędzia interdyscyplinarne, stanowiące niejako pomost pomiędzy tymi dwiema grupami metod. Rozwinięte protokoły, które opierały się tylko i wyłącznie o struktury znanych ligandów, dotyczyły przewidywania stabilności metabolicznej związków (wraz z innowacyjnym podejściem zapewniania interpretowalności wyników umożliwiającym optymalizację struktur pod kątem ocenianego parametru oraz konstrukcją platformy online do dającej dostęp do opracowanych narzędzi szerokiemu gronu odbiorców) oraz przygotowania eksperymentów referencyjnych pozwalających na ocenę skuteczności nowo rozwijanych narzędzi opartych o metody uczenia maszynowego.

Ponadto, udało mi się opracować innowacyjny protokół łączący fragmentaryczne podejścia do projektowania nowych ligandów (ang. *fragment-based*) z metodami *ligand-based*, pozwalający na generowanie nowych potencjalnie aktywnych związków na podstawie optymalnego ciągu bitowego. Uzyskiwany zestaw zer i jedynek odpowiada najbardziej pożądanemu składowi strukturalnemu ligandów danego receptora i jest wyznaczany przez algorytmy uczenia maszynowego typu generatywnego.

Uczenie maszynowe pozwoliło mi również na przygotowanie narzędzia do przewidywania efektu mutagenetycznego określonych podmian aminokwasowych, co umożliwia przewidzenie skutków wymiany aminokwasu dla aktywności danego liganda,

a także zdefiniowanie wymagań strukturalnych i w zakresie oddziaływań dla nowo projektowanych związków.

Zastosowanie metod statystycznych do analizy częstości oddziaływań określonych grup ligandów z celem biologicznym stanowiło także podstawę identyfikacji oddziaływań charakterystycznych dla określonych grup związków (wykazujących ściśle zdefiniowany profil aktywności lub o konkretnych własnościach).

Bardzo ważnym osiągnięciem jest również opracowanie protokołu do przewidywania wyników dokowania w zakresie oddziaływań ligand-białko na podstawie dwuwymiarowej struktury związku chemicznego. Zaletą tej metody jest znaczna oszczędność czasu obliczeniowego, a także możliwość szybkiego odfiltrowania struktur nieposiadających wymaganych oddziaływań z celem biologicznym podczas przeprowadzania procedur przesiewowych.

Ponadto, istotnym wkładem w rozwój metod z zakresu komputerowo wspomaganego projektowania leków jest wykorzystanie metod statystycznych do analizy wyników dynamiki molekularnej, co pozwoliło na identyfikację zależności w sposobie oddziaływania ligand-białko, które nie były widoczne przy zastosowaniu standardowych narzędzi obróbki danych z symulacji.

Podsumowując, wszystkie cele naukowe zaplanowane w ramach prezentowanego dzieła habilitacyjnego zostały osiągnięte. Innowacyjność, a zarazem skuteczność opracowanych protokołów z jednej strony, a także rosnące zapotrzebowanie na rozwój narzędzi cheminformacyjnych z drugiej strony, daje nadzieję, na szerokie wykorzystanie przygotowanych metodologii przez społeczność naukową. Mam nadzieję, że rezultaty uzyskane w ramach przeprowadzonych przeze mnie badań przyczynią się do znalezienia nowych, bardziej skutecznych terapeutyków.

#### 4.4.5. Bibliografia

1. Trist, D.G. Scientific process, pharmacology and drug discovery. *Curr. Opin. Pharmacol.* **2011**, *11*, 528–533.
2. Wouters, O.J.; McKee, M.; Luyten, J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA* **2020**, *323*, 844–853.
3. Takebe, T.; Imai, R.; Ono, S. The Current Status of Drug Discovery and Development as Originated in United States Academia: The Influence of Industrial and Academic Collaboration on Drug Discovery and Development. *Clin Transl Sci.* **2018**, *11*, 597-606.
4. Ratti, E.; Trist, D. The continuing evolution of the drug discovery process in the pharmaceutical industry. *Farmaco* **2001**, *56*, 13–19.
5. Metwaly, A.M.; Ghoneim, M.M.; Eissa, I.H.; Elsehemy, I. A., Mostafa, A. E., Hegazy, M. M.; Afifi, W. M.; Dou, D. Traditional ancient Egyptian medicine: A review. *Saudi J Biol Sci.* **2021**, *28*, 5823–5832.

6. Hacker, M. Chapter 1 - History of Pharmacology—From Antiquity to the Twentieth Century In: Pharmacology Principles and Practice; Hacker, M.; Messer, W.; Bachmann, K.; Eds. 2009, Academic Press.
7. Rao, V. S.; Srinivas, K. Modern drug discovery process: An in silico approach. *J. Bioinf. Seq. Anal.* **2011**, *5*, 89–94
8. Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W. Computational Methods in Drug Discovery. *Pharmacol. Rev.* **2014**, *66*, 334–395.
9. Terstappen, G. C.; Reggiani, A. In silico research in drug discovery. *Trends Pharmacol. Sci.* **2001**, *22*, 23–26.
10. Katsila, T.; Spyroulias, G. A.; Patrinos, G. P.; Matsoukas, M.-T. Computational approaches in target identification and drug discovery. *Comput. Struct. Biotechnol. J.* **2016**, *14*, 177–184.
11. Reddy, A. S.; Pati, S. P.; Kumar, P. P.; Pradeep, H. N.; Sastry, G. N. Virtual Screening in Drug Discovery – a Computational Perspective. *Curr. Protein Pept. Sci.* **2007**, *8*, 329–351.
12. Geppert, H.; Vogt, M.; Bajorath, J. Current trends in ligand-based virtual screening: molecular representations, data mining methods, new application areas, and performance evaluation. *J. Chem. Inf. Model.* **2010**, *2*, 205–216.
13. Anderson, A. C. The process of structure-based drug design. *Chem Biol.* **2003**, *9*, 787–797.
14. Yang, S.Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discov. Today* **2010**, *11-12*, 444–450
15. Sousa, F.; Fernandes, P. A.; Ramos, M. J. Protein – ligand docking : current status and future challenges. *Proteins* **2006**, *1*, 15–26
16. Fukunishi, Y. Structure-based drug screening and ligand-based drug screening with machine learning. *Combinatorial Chemistry & High Throughput Screening* **2009**, *12*, 397–408.
17. Kotsiantis, S.B. Supervised Machine Learning : A Review of Classification Techniques. *Informatica* **2007**, *31*, 249–268.
18. Böcker, A.; Derksen, S.; Schmidt, E.; Teckentrup, A.; Schneider, G. A hierarchical clustering approach for large compound libraries. *J. Chem. Inf. Model.* **2005**, *45*, 807–815.
19. Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; von Lilienfeld, O. A.; Müller, K. R.; Tkatchenko, A. Machine learning predictions of molecular properties: accurate many-body potentials and nonlocality in chemical space. *J. Phys. Chem. Lett.* **2015**, *12*, 2326–2331.
20. Schwaighofer, A.; Schroeter, T.; Mika, S.; Blanchard, G. How wrong can we get? A review of machine learning approaches and error bars. *Comb. Chem. High Throughput Screen.* **2009**, *5*, 453–468.
21. Varnek, A.; Baskin, I. Machine learning methods for property prediction in chemoinformatics: quo vadis? *J. Chem. Inf. Model.* **2012**, *6*, 1413–1437.
22. Mitchell, J. B. O. Machine learning methods in chemoinformatics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *5*, 468–481.
23. Melville, J. L.; Burke, E. K.; Hirst, J. D. Machine learning in virtual screening. *Comb. Chem. High Throughput Screen.* **2009**, *4*, 332–343.
24. Schroedl, S. Current methods and challenges for deep learning in drug discovery. *Drug Discov. Today Technol.* **2019**, *32–33*, 9–17.
25. Xia, X.; Hu, J.; Wang, Y.; Zhang, L.; Liu, Z. Graph-based generative models for de Novo drug design. *Drug Discov. Today Technol.* **2019**, *32–33*, 45–53.
26. Chen, H. Can Generative-Model-Based Drug Design Become a New Normal in Drug Discovery? *J. Med. Chem.* **2022**, *65*, 100–102.
27. Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins* **2002**, *4*, 409–443.
28. Sousa, F.; Fernandes, P. A.; Ramos, M. J. Protein – ligand docking : current status and future challenges. *Proteins* **2006**, *1*, 15–26.
29. Hollingsworth, S.A.; Dror, R. O. Molecular dynamics simulation for all. *Neuron*. **2018**, *99*, 1129–1143.
30. Phillips, J.C.; Hardy, D.J.; Maia, J.D.C.; Stone, J.E.; Ribeiro, J.V.; Bernardi, R.C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; McGreevy, R.; Melo, M.C.R.; Radak, B.K.; Skeel, R.D.; Singharoy, A.; Wang, Y.; Roux, B.;



- Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L.V.; Schulten, K.; Chipot, C.; Tajkhorshid, E. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **2020**, *153*, 044130.
31. Loukatou, S.; Papageorgiou, L.; Fakourelis, P.; Filntisi, A.; Polychronidou, E.; Basis, I.; Megalooikonomou, V.; Makałowski, W.; Vlachakis, D.; Kossida, S. Molecular dynamics simulations through GPU video games technologies. *J. Mol. Biochem.* **2014**, *3*, 64–71.
32. Bera, I.; Payghan, P.V. Use of Molecular Dynamics Simulations in Structure-Based Drug Discovery. *Curr. Pharm. Des.* **2019**, *25*, 3339–3349.
33. Zhao, H.; Caflisch, A. Molecular dynamics in drug design. *Eur. J. Med. Chem.* **2015**, *91*, 4–14.
34. Monticelli, L.; Tieleman, D.P. Force fields for classical molecular dynamics. *Methods Mol. Biol.* **2013**, *924*, 197–213.
35. Shonberg, J.; Kling, R. C.; Gmeiner, P.; Löber, S. GPCR crystal structures: medicinal chemistry in the pocket. *Bioorg. Med. Chem.* **2014**, *14*, 3880–3906.
36. Kobilka, B. K. G protein coupled receptor structure and activation. *Biochim. Biophys. Acta* **2007**, *4*, 794–807.
37. Kooistra, A.J.; Leurs, R.; de Esch, I.J.; de Graaf, C. From three-dimensional GPCR structure to rational ligand discovery. *Adv. Exp. Med. Biol.* **2014**, *796*, 129–157.
38. Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **2011**, *40*, D1100–D1107.
39. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
40. Irwin, J. J.; Khanh, G. T.; Young, J.; Dandarchuluun, C.; Wong, B. R.; Khurelbaatar, M.; Moroz, Y. S.; Mayfield, J.; Sayle, R. A. ZINC20 – A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *J. Chem. Inf. Model.* **2020**, *60*, 6065–6073.
41. Wishart, D.S.; Knox, C.; Guo, A.C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **2006**, *34*, D668–D672.
42. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* **2019**, *49*, D1388–D1395.
43. Isberg, V.; Mordalski, S.; Munk, C.; Rataj, K.; Harpsøe, K.; Hauser, A. S.; Vroiling, B.; Bojarski, A. J.; Vriend, G.; Gloriam, D. E. GPCRdb: An information system for G protein-coupled receptors. *Nucleic Acids Res.* **2017**, *45*, 2936.
44. Bajusz, D.; Rácz, A.; Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform.* **2015**, *7*, 20–33.
45. Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures - A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5*, 107–113.
46. Smith, D. H.; Carhart, R. E.; Venkataraghavan, R. Atom Pairs as Molecular Features in Structure-Activity Studies: Definition and Applications. *J. Chem. Inf. Model.* **1985**, *25*, 64–73.
47. Irwin, J.J.; Shoichet, B.K. ZINC--a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177–182.
48. Langdon, S. R.; Ertl, P.; Brown, N. Bioisosteric replacement and scaffold hopping in lead generation and optimization. *Mol. Inf.* **2010**, *5*, 366–385.
49. Böhm, H. J.; Flohr, A.; Stahl, M. Scaffold hopping. *Drug Discov. Today Technol.* **2004**, *3*, 217–224.
50. Pierce, A. C.; Rao, G.; Bemis, G. W. BREED: Generating novel inhibitors through hybridization of known ligands. Application to CDK2, P38, and HIV protease. *J. Med. Chem.* **2004**, *11*, 2768–2775.
51. Yap, C. W. E. I. Software news and update PaDEL-Descriptor : an open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.* **2011**, *7*, 1466–1474.
52. Baell, J. B.; Holloway, G. A. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for their Exclusion in Bioassays. *J. Med. Chem.* **2010**, *53*, 2719–2740.



53. Evans, B.; Rittle, K.; Bock, M.; DiPardo, R.; Freidinger, R.; Whitter, W.; Lundell, G.; Veber, D.; Anderson, P. Methods for Drug Discovery: Development of Potent, Selective, Orally Effective Cholecystokinin Antagonists. *J. Med. Chem.* 1988, 31, 2235–2246
54. Deng, Z.; Chuaqui, C.; Singh, J. Structural interaction fingerprint (SIFt): A novel method for analyzing three-dimensional protein-ligand binding interactions. *J. Med. Chem.* 2004, 47, 337–344.
55. Impellizzeri, A. A. R.; Pappalardo, M.; Basile, L.; Manfra, O.; Andressen, K. W.; Krobert, K. A.; Messina, A.; Levy, F. O.; Guccione, S. Identification of essential residues for binding and activation in the human 5-HT7(a) serotonin receptor by molecular modeling and site-directed mutagenesis. *Front. Behav. Neurosci.* 2015, 9, 1–12.
56. InstantJChem version 15.3.30.0, licensed by ChemAxon. 2015. Available online: [www.chemaxon.com](http://www.chemaxon.com).
57. Marcou, G.; Rognan, D. Optimizing Fragment and Scaffold Docking by Use of Molecular Interaction Fingerprints. *J. Chem. Inf. Model.* 2007, 47, 195–207.
58. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* 2004, 47, 1739–1749.
59. Koes, D. R.; Baumgartner, M. P.; Camacho, C. J. Lessons Learned in Empirical Scoring with Smina from the CSAR 2011 Benchmarking Exercise. *J. Chem. Inf. Model.* 2013, 53, 1893–1904.
60. Castro, W.; Oblitas, J.; Santa-Cruz, R.; Avila-George, H. Multilayer perceptron architecture optimization using parallel computing techniques. *PLoS One.* 2017, 12, e0189369.
61. Shang, C.; Liu, Q.; Chen, K.; Sun, J.; Lu, J.; Yi, J.; Bi, J. Edge Attention-Based Multi-Relational Graph Convolutional Networks. *ArXiv.* 2018. abs/1802.04944.
62. Raehal, K. M. Morphine Side Effects in -Arrestin 2 Knockout Mice. *J. Pharmacol. Exp. Ther.* 2005, 314, 1195–1201.
63. Wacker, D.; Wang, S.; McCorvy, J. D.; Betz, R. M.; Venkatakrishnan, A. J.; Levit, A.; lansu, K.; Schools, Z. L.; Che, T.; Nichols, D. E.; Shoichet, B. K.; Dror, R. O.; Roth, B. L. Crystal Structure of an LSD-Bound Human serotonin receptor. *Cell* 2017, 168, 377–389.
64. Wang, S.; Che, T.; Levit, A.; Shoichet, B. K.; Wacker, D.; Roth, B. L. Structure of the D2 dopamine receptor bound to the atypical antipsychotic drug risperidone. *Nature* 2018, 555, 269–273.
65. Ågren, R.; Zeberg, H.; Stępniewski, T. M.; Free, R. B.; Reilly, S. W.; Luedtke, R. R.; Århem, P.; Ciruela, F.; Sibley, D. R.; Mach, R. H.; Selent, J.; Nilsson, J.; Sahlholm K, K. Ligand with Two Modes of Interaction with the Dopamine D2 Receptor-An Induced-Fit Mechanism of Insurmountable Antagonism. *ACS Chem. Neurosci.* 2020, 11, 3130–3143.
66. Lundberg, S.M.; Lee, S.A. A unified approach to interpreting model predictions. In: Advances in neural information processing systems 30. Guyon, I.; Luxburg, U.V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; Garnett, R., Eds. Red Hook, Curran Associates Inc, 2017, USA, 4765–4774.

## 5. Informacja o wykazywaniu się istotną aktywnością naukową albo artystyczną realizowaną w więcej niż jednej uczelni, instytucji naukowej lub instytucji kultury, w szczególności zagranicznej

### 5.1. Aktywność naukowa przed uzyskaniem stopnia naukowego doktora

Swoją pracę naukową rozpoczęłam tuż po ukończeniu studiów licencjackich na kierunku chemia, w sierpniu 2010 roku, gdy jako stażystka dołączyłam do Zespołu Modelowania Molekularnego, będącego częścią Zakładu Chemii Leków Instytutu Farmakologii PAN. Swoje zatrudnienie w ramach umowy o pracę (stanowisko inżynierijsko-techniczne) rozpoczęłam w dniu 01.10.2010 od udziału w projekcie pt. „Stworzenie platformy do poszukiwania związków działających na układy serotonergiczny lub glutamatergiczny jako potencjalnych leków przeciwdepresyjnych lub przeciwlękowych”. Projekt ten był finansowany

z Norweskiego Mechanizmu Finansowego w ramach Polsko-Norweskiego Funduszu Badań Naukowych. Jednostkami realizującymi badania w ramach tego grantu był Instytut Farmakologii PAN (Lider), Narodowy Instytut Leków w Warszawie oraz Zakład Biologii Medycznej Uniwersytetu w Tromsø. Ciągła współpraca zdalna pomiędzy partnerami projektowymi była wspierana przez regularne wzajemne wizyty badaczy z poszczególnych ośrodków naukowych. Miejscem mojego wyjazdu badawczego był Uniwersytet w Tromsø (pobyt w okresie 12–21.09.2011 r.).

W ramach rozszerzania zastosowań metod uczenia maszynowego do charakterystyki związków chemicznych (początkowo zajmowałam się przewidywaniem aktywności wobec określonego celu biologicznego – publikacje A.11, A.12, A.13 z *Wykazu osiągnięć naukowych*), rozpoczęłam również predykcje związane z własnościami fizykochemicznymi oraz parametrami ADME. Tematyka ta pozostaje w ścisłym obszarze zainteresowań grupy kierowanej przez prof. Marcello Leopoldo z Wydziału Farmaceutycznego w Bari, która posiada duże doświadczenie i sporą ilość danych w zakresie stabilności metabolicznej związków. Od 2013 roku prowadziłam badania we współpracy z grupą naukowców z Bari, opracowując modele predykcyjne pozwalające na ewaluację związków pod kątem tego parametru. W ramach wspólnych działań badawczych, w okresach 17–25.11.2014 oraz 09–16.07.2015 przebywałam w Bari, a efektem początkowego etapu współpracy jest publikacja Lacivita, E.; **Podlewska, S.**; Speranza, L.; Niso, M.; Satała, G.; Perrone, R.; Perrone-Capano, C.; Bojarski, A. J.; Leopoldo, M. Structural modifications of the serotonin 5-HT<sub>7</sub> receptor agonist N-(4-cyanophenylmethyl)-4-(2-biphenyl)-1-piperazinehexanamide (LP-211) to improve in vitro microsomal stability: A case study, *Eur. J. Med. Chem.*, **2016**, *120*, 363–379 (A.1; kolejne wspólne prace opublikowano już po uzyskaniu stopnia naukowego doktora).

Rozszerzając wachlarz narzędzi modelowania molekularnego, które wykorzystywałam w swoich badaniach, zajęłam się również dynamiką molekularną. Aby podnieść swoje umiejętności w tym zakresie, udałam się na tygodniowy staż szkoleniowy (09–13.02.2015) do grupy prof. Sereiny Riniker na Wydziale Chemii i Stosowanych Nauk Biologicznych ETH Zürich – ekspertów w zakresie symulacji dynamiki molekularnej i jej zastosowania w komputerowo wspomaganym projektowaniu leków.

Rozpoczynając swoją pracę badawczą w Instytucie Farmakologii PAN skupiałam się na rozwijaniu narzędzi przede wszystkim pod kątem ich potencjalnych zastosowań do poszukiwania nowych ligandów receptorów serotoninowych, jednak z czasem moje zainteresowania zaczęły się poszerzać o kolejne grupy receptorów. Z uwagi na relatywnie słabą

eksplorację tematu selektywności pomiędzy receptorami serotoninowymi a histaminowymi, postanowiłam zgłębić ten temat zarówno z perspektywy związków chemicznych, jak i struktur białek. Badania prowadziłam we współpracy z grupą prof. Chrisa de Graafa z Wydziału Chemii Wolnego Uniwersytetu w Amsterdamie, ze względu na posiadaną przez nich wiedzę i doświadczenie w prowadzeniu badań nad receptorami histaminowymi. Początkowo, odbyłam dwie krótkoterminowe wizyty na Uniwersytecie w Amsterdamie (23–27.02.2015, 14–20.06.2015), a następnie spędziłam tam cztery miesiące (02.11.2015–01.03.2016), podczas stażu w ramach grantu ETIUDA (2015/16/T/NZ2/00058), który przyznano mi właśnie na realizację badań we współpracy z grupą prof. de Graafa. W okresie odbywania stażu, przeprowadziłam szczegółową analizę problemu selektywności pomiędzy wybranymi parami receptorów serotoninowych i histaminowych, jednak współpraca z grupą prof. de Graafa była kontynuowana także po zakończeniu stażu i po uzyskaniu stopnia naukowego doktora.

Poza kooperacją z ośrodkami zagranicznymi, prowadziłam również badania w ramach licznych współprac krajowych. W lutym 2013 roku, Zespół Modelowania Molekularnego, będący częścią Zakładu Chemii Leków IFPAN nawiązał kontakt z Grupą Uczenia Maszynowego, działającej na Wydziale Matematyki i Informatyki Uniwersytetu Jagiellońskiego. Początkowo, prowadziłam wspólne działania z Wojciechem Czarneckim, a następnie do naszego zespołu dołączyli kolejni badacze (studenci i doktoranci z Instytutu Informatyki UJ). Prace te były ukierunkowane w stronę badań teoretycznych. Rozpoczęto je od opracowania metody pozwalającej na uwzględnienie w modelach predykcyjnych niepewności wyników badań biologicznych (A.9). Tego rodzaju modele są bardzo pożądane z uwagi na stosunkowo częstą obecność w bazach danych (zawierających informacje o aktywności związków chemicznych) niespójnych (a czasem wręcz wykluczających się) wyników badań powinowactwa. Przeprowadziliśmy również optymalizację procesu selekcji parametrów maszyny wektorów nośnych w zadaniach cheminformatycznych (A.4), a także prace zmierzające do wykorzystania w zadaniach związanych z oceną aktywności związków nowych algorytmów klasyfikacyjnych, nierozpatrywanych do tej pory w zagadnieniach dotyczących ewaluacji związków chemicznych, w szczególności algorytmów opartych na procedurze randomizacji (A.3).

Z kolei we współpracy z Katedrą Technologii i Biotechnologii Środków Leczniczych Wydziału Farmaceutycznego CMUJ przeprowadzałam symulacje dynamiki molekularnej pozwalające na zaproponowanie mechanizmu znoszenia lekooporności na antybiotyki beta-laktamowe w metycylinoopornych szczepach gronkowca złocistego (A.5), a także

realizowałam badania obliczeniowe w zakresie modelowania ligandów receptora serotoninowego 5-HT<sub>7</sub>, posiadających rdzeń hydantoiny (A.2).

Owoce prac, zarówno prowadzonych z innymi pracownikami IF PAN, jak również w ramach licznych współprac krajowych i zagranicznych było opublikowanie 13 artykułów naukowych (A.1–A.13). Część z nich stała się przedmiotem rozprawy doktorskiej, obronionej 13.10.2016 na Wydziale Chemii UJ (tytuł rozprawy „Development of machine learning-based tools for computer-aided drug design”). Jej celem było opracowanie zestawu narzędzi opartych o metody uczenia maszynowego do poszukiwania nowych biologicznie aktywnych związków. Cykl 7 publikacji wchodzących w skład rozprawy doktorskiej (A.3, A.4, A.6, A.7, A.9, A.11, A.13) obejmował: ilościową i jakościową optymalizację składu zestawu związków nieaktywnych wykorzystywanych podczas uczenia modeli (A.11, A.13), wspomniany wcześniej protokół uwzględniania niepewności wyników badań biologicznych w modelach predykcyjnych (A.9), nową metodę optymalizacji parametrów maszyny wektorów nośnych (A.4), wykorzystanie nowych metod uczących do poszukiwania związków aktywnych biologicznie (A.3), protokół do automatycznej oceny wyników dokowania oparty o algorytmy uczące (A.7), a także wykorzystanie metod uczenia maszynowego do poszukiwania nowych ligandów receptora serotoninowego 5-HT<sub>6</sub> w procedurze wirtualnego przesiewania dwóch komercyjnie dostępnych baz związków (A.6). Ocenę struktur w procesie wirtualnego skringingu przeprowadziłam w tzw. podejściu fingerprint-consensus, tj. związek był traktowany jako potencjalnie aktywny wtedy i tylko wtedy, gdy został on oceniony jako taki dla wszystkich trzech zastosowanych reprezentacji. Dla dwóch, spośród 22 zamówionych, związków eksperymentalnie wykazano istotną aktywność wobec receptora 5-HT<sub>6</sub> ( $K_i$  równe odpowiednio 119 i 670 nM), przy czym pierwszy posiadał również aktywność do receptora 5-HT<sub>2A</sub> ( $K_i = 296$  nM), natomiast drugi wykazywał selektywność powinowactwa wobec pozostałych badanych receptorów (5-HT<sub>1A</sub>, 5-HT<sub>2A</sub>, 5-HT<sub>7</sub>). Z uwagi na usunięcie z rozpatrywanego zbioru związków podobnych do znanych ligandów 5-HT<sub>6</sub>R, znalezione nowe aktywne związki posiadają unikalne struktury, a ponadto należą do grupy tzw. ligandów niezasadowych, charakteryzujących się niestandardowym schematem oddziaływań z celem biologicznym.

## 5.2. Aktywność naukowa po uzyskaniu stopnia naukowego doktora

Po uzyskaniu stopnia naukowego doktora, złożyłam wniosek o grant badawczy w programie Narodowego Centrum Nauki SONATINA. Projekt otrzymał finansowanie, a jego realizacja miała miejsce na Wydziale Farmaceutycznym Uniwersytetu Jagiellońskiego –

Collegium Medicum, gdzie podjęłam zatrudnienie w dniu 08.10.2018 r. (w okresie zatrudnienia na UJCM, tj. do dnia 14.10.2021, przebywałam na urlopie bezpłatnym w IF PAN).

Celem badań prowadzonych w projekcie SONATINA była konstrukcja i zastosowanie narzędzi obliczeniowych do poszukiwania nowych, niskozasadowych ligandów receptora serotoninowego 5-HT<sub>7</sub>, charakteryzujących się zwiększoną stabilnością metaboliczną. Z uwagi na doświadczenie w zakresie badań nad stabilnością metaboliczną, jak również nad ligandami receptora 5-HT<sub>7</sub> grupy prof. Marcello Leopoldo z Uniwersytetu w Bari, zespół ten pełnił w projekcie rolę partnera zagranicznego. Uniwersytet w Bari był również miejscem odbywania przeze mnie czteromiesięcznego stażu zagranicznego (10.02.2021–09.06.2021), w czasie którego przeprowadziłam optymalizację narzędzia umożliwiającego interpretowalność predykcji dostarczanych przez algorytmy uczenia maszynowego (prace realizowane wspólnie z badaczami z Wydziału Matematyki i Informatyki UJ). Zostało ono udostępnione jako platforma online pozwalająca na ocenę predykcji stabilności metabolicznej związku dostarczonego przez użytkownika oraz jego optymalizację pod kątem tego parametru poprzez wykorzystanie informacji o interpretowalności przewidywań algorytmów uczenia maszynowego (<https://metstab-shap.matinf.uj.edu.pl/>) – B.6 [H9].

Kontynuując zapoczątkowaną w okresie studiów doktoranckich współpracę, określiliśmy również determinanty aktywności dla określonych paneli aktywności w obrębie receptorów serotoninowych, które zostały następnie zastosowane w procesie wirtualnego przesiewania bibliotek związków chemicznych do poszukiwania selektywnych ligandów tych białek (B.24 [H2]).

Ponadto, wraz z partnerami z Bari prowadziłam także badania nad kinetyką wybranej serii ligandów receptora 5-HT<sub>7</sub>, wykorzystując dokowanie oraz dynamikę molekularną do wyjaśnienia obserwowanych zależności struktura-oddziaływanie z receptorem-parametry kinetyczne (B.2 [H11]). Podczas mojego pobytu w Bari, rozpoczęliśmy też wspólne prace związane z poszukiwaniem nowych ligandów receptora purynergicznego P2X<sub>7</sub>.

Po uzyskaniu stopnia naukowego doktora, kontynuowałam również współpracę z grupą prof. de Graafa (Wolny Uniwersytet w Amsterdamie). Jej efektem jest przygotowanie kompletnego zestawu danych mutagenetycznych dla aminergicznych receptorów GPCR wraz z protokołem do przewidywania efekt mutagenetycznego dla danej pary ligand-mutacja (B.22 [H3]) oraz przeprowadzenie procedury wirtualnego przesiewania bibliotek związków chemicznych pod kątem poszukiwania ligandów receptorów serotoninowych o ściśle zdefiniowanym profilu aktywności (B.24 [H2]).

Partnerem zagranicznym realizowanych przeze mnie badań był także Uniwersytet w Katanii (grupa Prof. Emanuele Amaty, współpraca koordynowana przez dr Katarzynę Szczepańską oraz prof. dr hab. Katarzynę Kieć-Kononowicz). Wraz z badaczami należącymi do tego zespołu prowadziłam prace z zakresu modelowania molekularnego (dokowanie oraz symulacje dynamiki molekularnej), zmierzające do wyjaśnienia obserwowanych zależności struktura-aktywność grupy ligandów receptora histaminowego H<sub>3</sub> (wraz z ich jednoczesną aktywnością lub nieaktywnością wobec receptora sigma-1 i sigma-2) o rdzeniu piperazyny oraz piperydyny (B.4).

Niezależnie od realizacji projektu SONATINA, w okresie podoktorskim brałam również udział w projekcie LIDER pt. „Polifarmakologiczna platforma skringowa in silico”, finansowanym przez Narodowe Centrum Badań i Rozwoju w charakterze wykonawcy (LIDER/37/0137/L-9/17/NCBR/2018, kierownik projektu: dr hab. Rafał Kurczab, okres udziału: 01.01.2019–30.06.2021, jednostka realizująca: Instytut Farmakologii im. Jerzego Maja PAN). Celem projektu było przygotowanie serwisu online, pozwalającego na kompleksową ocenę aktywności związków wobec kilkuset celów biologicznych, z wykorzystaniem modeli predykcyjnych opartych o ścieżkę *ligand-based* i *structure-based*. Możliwość optymalizacji skuteczności modeli predykcyjnych *ligand-based*, przy wykorzystaniu wiedzy z eksperymentów *structure-based* stała się przedmiotem publikacji (B.9).

Po uzyskaniu stopnia naukowego doktora, pełniłam także rolę wykonawcy w projekcie kierowanym przez prof. dr hab. Ryszarda Przewłockiego dotyczącego poszukiwania nowych ligandów receptorów opioidowych, będących tzw. stronniczymi agonistami (ang. *biased agonists*). Aktywacja określonych ścieżek sygnałowych przy jednoczesnym braku wpływu na inne, jest nowym kierunkiem badań w terapii bólu, dającej nadzieje na uniknięcie efektów ubocznych związanych ze stosowaniem klasycznych opioidów, takich jak efekt euforyczny czy depresja ośrodka oddechowego. Moją rolą była realizacja zadań modelowania molekularnego, w tym poszukiwanie determinantów molekularnych dla stronniczego agonizmu (B.3, B.14) oraz prowadzenie kampanii wirtualnego skringu do poszukiwania związków wywołujących określony efekt funkcjonalny na badane białka.

Z ramienia Instytutu Farmakologii im. Jerzego Maja PAN, prowadziłam także badania w ramach szeregu współprac krajowych, m.in. z Wydziałem Matematyki i Informatyki UJ oraz z Instytutem Biologii Medycznej PAN w Łodzi.

Współpraca z Wydziałem Matematyki i Informatyki UJ zaowocowała utworzeniem pierwszego w kraju nieformalnego ośrodka badań cheminformatycznych, złożonego z kilkunastu naukowców – osób z tytułem naukowym doktora, doktorantów i studentów



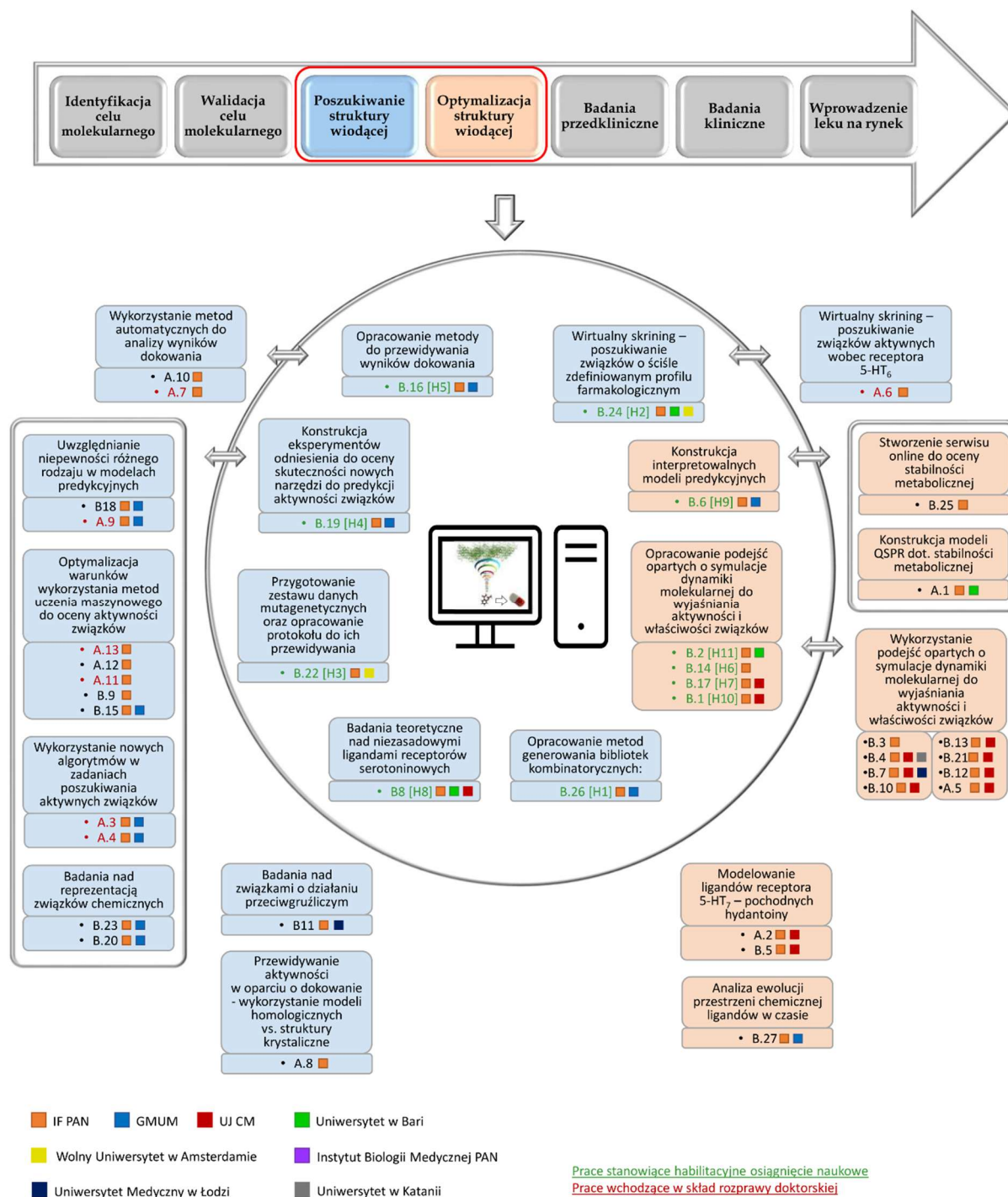
(w większości są to członkowie Grupy Uczenia Maszynowego, GMUM). Opracowane w obrębie grupy metody opisano w 12 artykułach naukowych (opublikowanych zarówno przed, jak i po uzyskaniu stopnia naukowego doktora: A.3, A.4, A.9, B.6, B.15, B.16, B.18, B.19, B.20, B.23, B.26, B.27), a ich zastosowanie obejmuje nie tylko ocenę aktywności związków, ale również możliwość generowania nowych potencjalnych ligandów (B.26), wspieranie procesu optymalizacji struktur związków, dzięki konstrukcji interpretowalnych modeli uczenia maszynowego (B.6 [H9]), symulację wyników dokowania z dwuwymiarowej struktury związku (B.16), analizę przestrzeni chemicznej ligandów (oraz zmian zachodzących w tej przestrzeni w czasie, B.27), analizę deskryptorów generowanych z trójwymiarowej struktury związku w zależności od jego konformacji (B.20), konstrukcję eksperymentów i zestawów danych stanowiących punkt odniesienia do oceny kolejnych narzędzi (B.19 [H4]) oraz stworzenie dwuwymiarowego fingerprintu podstrukturalnego, kodującego nie tylko obecność określonych podstruktur w związku, ale również informującego o występowaniu połączeń pomiędzy nimi (B.23). Ponadto, przeprowadziliśmy szczegółową ocenę różnych metryk stosowanych przy automatycznej ocenie podobieństwa związków chemicznych (B.15) i włączyliśmy niepewność modelu predykcyjnego do informacji zwracanych przez model, co pozwala na ocenę przydatności dostarczanych przez niego odpowiedzi (B.18).

Współpraca z Instytutem Biologii Medycznej PAN w Łodzi dotyczy badań nad optymalizacją terapii przeciwbakteryjnej gruźlicy opartej o wzorce ewolucyjne obecne w DNA bakterii. W projekcie tym, prowadzę badania *in silico* w odniesieniu do trzech białek będących celami molekularnymi w stosowanych obecnie terapiach przeciwprątkowych (B.11). Po uzyskaniu stopnia naukowego doktora kontynuowałam także prace (wspólnie z badaczami z Wydziału Farmaceutycznego UJ CM) nad poszukiwaniem związków pozwalających na leczenie zakażeń bakteriami wykazującymi oporność wielolekową: B.10, B.21 (w tym wspólne prace z dr hab. n. farm. prof. nadzw. Iwoną Głowacką z Uniwersytetu Medycznego w Łodzi: B.7) oraz poszukiwaniem nowych ligandów receptora 5-HT<sub>7</sub> opartych o rdzeń hydantoiny (B.5, B.13).

Po uzyskaniu stopnia naukowego doktora poszerzyłam również zestaw modelowanych celów biologicznych, a opracowane metodologie obliczeniowe stosowałam również coraz szerzej we wspomnianych już wcześniej badaniach nad receptorami opioidowymi, lecz także w badaniach nad ligandami receptorów histaminowych i sigma (B.17 [H7], B.12, B.4; Rysunek 16).

Jak przedstawiono na Rysunku 16, w okresie podoktorskim prowadziłam znacznie więcej badań w obszarze optymalizacji struktur związków (tylko 3 spośród 18 artykułów





**Rysunek 16.** Analiza tematów badawczych rozpatrywanych podczas mojej dotychczasowej pracy naukowej (wraz z wynikającymi z realizacji badań publikacjami), ze wskazaniem partnerów współpracy oraz wyróżnieniem badań stanowiących habilitacyjne osiągnięcie naukowe (dla porównania wskazano również publikacje wchodzące w skład rozprawy doktorskiej).

opublikowanych w tematyce optymalizacji struktury wiodącej ukazały się przed uzyskaniem stopnia naukowego doktora). W szczególności, po obronie pracy doktorskiej zajęłam się bardziej szczegółowo metodami structure-based, a zwłaszcza symulacjami dynamiki

molekularnej oraz opracowywaniem metod do automatycznej analizy ich rezultatów ([H6], [H7], [H10], [H11]). Poza artykułami wchodzącymi w skład osiągnięcia habilitacyjnego, jestem autorem 8 innych publikacji, w których wykorzystałam opracowane metody do analizy wyników dynamiki molekularnej do wyjaśniania obserwowanych zależności pomiędzy zachowaniem kompleksów ligand-białko podczas symulacji, a aktywnością (lub innymi właściwościami) ligandów.

Ponadto, w ramach rozwoju narzędzi pozwalających na optymalizację struktury wiodącej, przygotowałam narzędzie online do oceny związków pod kątem ich stabilności metabolicznej z wykorzystaniem interpretowalnych modeli uczenia maszynowego ([H9]). Dzięki dostarczaniu informacji o podstrukturach mających największy wkład w daną odpowiedź modelu wraz z predykcjami dla najbardziej podobnych strukturalnie związków możliwe jest przeprowadzenie optymalizacji struktury związku wejściowego pod kątem ocenianego parametru.

Jako rozszerzenie badań związanych z opracowaniem modeli predykcyjnych do oceny stabilności metabolicznej prowadzonych podczas studiów doktoranckich (A.1), przygotowałam serwis online do analizy tej własności związków, skupiający 108 modeli predykcyjnych, zarówno klasyfikacyjnych, jak i regresyjnych (B.25).

Zupełnie nowym obszarem badań, poruszonym dopiero po uzyskaniu stopnia naukowego doktora było również opracowanie podejścia do generowania nowych związków [H1], gdyż wcześniej (przed doktoratem) nie zajmowałam się podejściami generatywnymi, a jedynie oceną związków już istniejących.

Część mojej pracy doktorskiej stanowiło natomiast przygotowanie protokołu do automatycznej oceny wyników dokowania. Obejmował on reprezentację kompleksów ligand-receptor przy pomocy deskryptorów oddziaływań strukturalnych oraz deskryptorów typu Spectrophores, ocenę przez metody uczenia maszynowego, oraz wieloetapową analizę wyników wykorzystującą specjalnie przygotowany schemat wagowania (uwzględnił on skuteczność klasyfikacji poszczególnych metod uczenia maszynowego, wartość funkcji oceniającej programu dokującego i jakość modelu homologicznego białka; A.7). Z kolei w okresie podoktorskim, skonstruowałam metodę, która pozwala na przewidywanie wyników dokowania w postaci charakterystyki oddziaływań ligand-cel biologiczny, z rozróżnieniem różnych typów oddziaływań ([H5]).

Zupełnie nową tematyką badań, nieporuszaną w rozprawie doktorskiej, były również badania z wykorzystaniem danych o zmianie aktywności ligandów na skutek punktowych mutacji receptorów (wraz z protokołem do przewidywania wpływu takich podmian

aminokwasowych na powinowactwo związków; **H3**) oraz dogłębne badania teoretyczne nad determinantami aktywności dla niezasadowych ligandów receptorów serotoninowych (**[H8]**).

Zarówno w okresie przed, jak i po uzyskaniu stopnia naukowego doktora, dokonałam wirtualnej oceny związków, będących częścią komercyjnych bibliotek struktur (A.6, **[H2]**); w przypadku publikacji A.1 był to receptor 5-HT<sub>6</sub>, a w publikacji **[H2]** – receptor 5-HT<sub>7</sub>, jednak w przypadku skriningu prowadzonego w ramach badań podoktorskich, analizowano również oddziaływania z innymi podtypami receptorów serotoninowych (5-HT<sub>1A</sub>, 5-HT<sub>1B</sub>, 5-HT<sub>2A</sub>, 5-HT<sub>2B</sub>, 5-HT<sub>6</sub>).

Z uwagi na fakt, że moją rozprawę doktorską stanowiły w znakomitej części publikacje teoretyczne, rozważające z jednej strony wpływ różnych czynników na skuteczność predykcijną metod uczenia maszynowego (A.11, A.13), a z drugiej strony, wykorzystałam również nowe algorytmy optymalizacji maszyny wektorów nośnych oraz nowe algorytmy uczące oparte na procedurze randomizacji do oceny związków chemicznych (A.3, A.4), a także opracowałam protokół do uwzględniania niepewności wyników badań biologicznych w modelach predykcyjnych (A.9), po uzyskaniu stopnia naukowego doktora postanowiłam uporządkować kwestię wpływu różnych czynników na efektywność procesu oceny aktywności związków. Zaowocowało to przygotowaniem eksperymentów odniesienia i zestawem danych dla 100 celów biologicznych, pozwalającym na porównywanie skuteczności przy tworzeniu nowych narzędzi do ewaluacji potencjału leczniczego związków chemicznych **[H4]**.

## **6. Informacja o osiągnięciach dydaktycznych, organizacyjnych oraz popularyzujących naukę lub sztukę**

### **6.1. Działalność dydaktyczna**

- rok akademicki 2021/2022 – „Uczenie maszynowe w projektowaniu leków”, Uniwersytet Jagielloński Wydział Matematyki i Informatyki, prowadzenie kursu dla studiów II stopnia na kierunku Informatyka oraz dla studiów II stopnia na kierunku Matematyka (specjalność: matematyka komputerowa), forma: wykład (45 h) + ćwiczenia (15 h)
- rok akademicki 2018/2019 – „Wstęp do cheminformatyki”, Uniwersytet Jagielloński Wydział Matematyki i Informatyki, prowadzenie kursu dla studiów doktoranckich na kierunku Informatyka, forma: wykład (30 h)
- promotor pomocniczy w przewodzie doktorskim mgr Agnieszki Wojtuch (data otwarcia przewodu doktorskiego: 27.09.2018, Uniwersytet Jagielloński Wydział Matematyki i Informatyki, promotor: prof. dr hab. Jacek Tabor, dyscyplina: informatyka)

## 6.2. Działalność popularyzacyjna

W ramach działalności popularyzacyjnej, w latach 2011–2015 prowadziłam współpracę z Fundacją Uniwersytet Dzieci, opracowując i prowadząc warsztaty matematyczne dla młodych adeptów nauki w wieku 8–9 lat, zarówno podczas zajęć weekendowych, jak i w szkołach podstawowych. Tematy przygotowanych przeze mnie warsztatów były następujące: „Ile stóp ma łokieć?”, „Czy kąty są w modzie?” oraz „Co można wykopać z danych?”.

W latach 2010–2012, brałam również udział w Festiwalu Nauki w Krakowie oraz w Małopolskiej Nocy Naukowców, gdzie wraz z innymi członkami Naukowego Koła Chemików wykonywałam pokazy chemiczne na Rynku Głównym w Krakowie oraz na Wydziale Chemii UJ.

## 6.3. Działalność organizacyjna

W okresie studiów doktoranckich byłam członkiem Komitetu Organizacyjnego Giełdy Prac Dyplomowych 2013 – konferencji studenckiej, która odbyła się w dniach 6–7.06.2013 na Wydziale Chemii UJ.

Ponadto, w 2015 roku przygotowałam warsztaty cheminformatyczne (wykład + zajęcia praktyczne) dla uczestników konferencji studenckiej Liczby-Komputery-Życie, łączącej tematykę matematyczną, informatyczną i biologiczną (zajęcia odbyły się 15.05.2015 r.).

## 7. Pozostałe informacje

### 7.1. Udział w projektach badawczych

#### 7.1.1. Kierowanie projektami badawczymi

- 26.11.2019– „Opracowanie kompleksowego zestawu narzędzi obliczeniowych do poszukiwania ligandów określonego receptora na przykładzie receptora serotoninowego 5-HT<sub>7</sub>” – Narodowe Centrum Nauki: OPUS Nr 2018/31/B/NZ2/00165, wysokość dofinansowania: 645 200 PLN
- 15.10.2018–14.10.2021 „Rozwój grupy niezasadowych ligandów receptora serotoninowego 5-HT<sub>7</sub>, pochodnych 2-(6-okso-3-fenylo-1,6-dihydropirydazyn-1-ylo)-N-fenyloacetamidu, o zwiększonej stabilności metabolicznej” – Narodowe Centrum Nauki: SONATINA Nr 2018/28/C/NZ7/00145, wysokość dofinansowania: 790 877 PLN

- 01.10.2015–30.09.2016 „Wykorzystanie metod uczenia maszynowego w zadaniach komputerowo wspomaganego projektowania leków” – Narodowe Centrum Nauki: ETIUDA Nr 2015/16/T/NZ2/00058, wysokość dofinansowania: 76 476 PLN
- 12.04.2016–11.05.2018 „Poszukiwanie nowych struktur wiodących dla ligandów receptora serotoninowego 5-HT<sub>7</sub> o zwiększonej stabilności metabolicznej” – Narodowe Centrum Nauki: HARMONIA Nr 2015/18/M/NZ7/00377, wysokość dofinansowania: 243 120 PLN
- 13.02.2014–12.02.2016 „Opracowanie narzędzia do automatycznej oceny kompleksów ligandów z receptorami serotoninowymi opartego o wykorzystanie metod uczenia maszynowego” – Narodowe Centrum Nauki: PRELUDIUM Nr 2013/09/N/NZ2/01917, wysokość dofinansowania: 97 500 PLN

#### 7.1.2. Uczestnictwo w projektach badawczych w charakterze wykonawcy

- Od 24.07.2019 – “Racjonalne projektowanie i ocena funkcjonalnie stronniczych agonistów receptorów opioidowych”, Narodowe Centrum Nauki: Nr 2018/31/B/NZ7/03954 (OPUS), kierownik projektu: prof. dr hab. Ryszard Przewłocki
- Od 01.10.2020 – “Wykorzystanie wzorców ewolucyjnych w optymalizacji chemoterapii gruźlicy”, Narodowe Centrum Nauki: Nr 2019/34/E/NZ6/00221 (SONATA BIS), kierownik projektu: dr hab. Alina Minias (Instytut Biologii Medycznej PAN, Łódź)
- 01.01.2019–30.06.2021 “Polifarmakologiczna platforma skringowa in silico”, Narodowe Centrum Badań i Rozwoju: Nr LIDER/37/0137/L-9/17/NCBR/2018, kierownik projektu: dr hab. Rafał Kurczab (LIDER)
- 27.01.2015–26.08.2017 „Paradygmat minimalizacji pamięci w klastrowaniu”, Narodowe Centrum Nauki, Nr 2014/13/B/ST6/01792 (OPUS), kierownik projektu: dr hab. Jacek Tabor (Wydział Matematyki i Informatyki Uniwersytetu Jagiellońskiego)
- 01.11.2013–31.12.2015 “EXtention of academia-based PLATFORM to antidepressant hits discovery (PLATFORMex),” Narodowe Centrum Badań

i Rozwoju: Nr Pol-Nor/198887/73/2013, kierownik projektu: prof. dr hab. Andrzej Bojarski (Polsko-Norweska Współpraca Badawcza)

- 01.07.2013–31.12.2015 Allosterix “Innovative therapies for neurodegenerative and neurodevelopmental diseases based on mGlu receptor allosteric modulators”, Narodowe Centrum Badań i Rozwoju: Nr UDA-POIG.01.03.01-12-100/08-0, kierownik projektu: prof. dr hab. Andrzej Pilc
- 17.05.2012–16.05.2014 „Zastosowanie profili SIFt w wirtualnym skryningu”, Ministerstwo Nauki i Szkolnictwa Wyższego: Nr 0046/DIA/2012/41, kierownik projektu: Stefan Mordalski (Diamentowy Grant)
- 01.07.2011–31.12.2013 DeMeTer „Depresja-mechanizmy-terapia”, Nr POIG.01.01.02-12-004/09-00, kierownik projektu: prof. dr hab. Krzysztof Wędzony
- 08.2012–08.2013 – „Wpływ gęstości fingerprintów na efektywność rozpoznawania chemicznych wzorców strukturalnych”, Narodowe Centrum Nauki: Nr 2011/03/N/NZ2/02478, kierownik projektu: mgr Rafał Kurczab (PRELUDIUM)
- 01.07.2011–30.06.2013 – ProKog „Antagoniści receptora 5-HT6: Nowoczesne leki przeciwpsychotyczne o dodatkowym działaniu prokognitywnym”, Nr UDA-POIG.01.03.01-12-063/09-02, kierownik projektu: prof. dr hab. Piotr Popik
- 01.07.2011–31.03.2013 – ModAll „Modulacja allosteryczna – nowa strategia w farmakoterapii. Identyfikacja własności psychotropowych ligandów receptorów glutaminianergicznych III grupy”, Nr UDA-POIG.01.03.01-12-100/08-0, kierownik projektu: prof. dr hab. Andrzej Pilc
- 01.10.2010–31.06.2011 – „Stworzenie platformy do poszukiwania związków działających na układy serotonergiczny lub glutamatergiczny jako potencjalnych leków przeciwdepresyjnych lub przeciwlękowych”, Nr PNRF-103-AI-1/07, kierownik projektu: prof. dr hab. Andrzej Pilc (Polsko-Norweski Fundusz Badań Naukowych)

## 7.2. Nagrody i wyróżnienia

- rok akademicki 2008/2009, 2009/2010 – stypendium za wyniki w nauce, rok akademicki 2011/2012 – stypendium Rektora dla najlepszych studentów



- rok akademicki 2010/2011, 2011/2012 – stypendium Ministra Nauki i Szkolnictwa Wyższego za wybitne osiągnięcia dla studentów
- rok akademicki 2015/2016 – stypendium Ministra Nauki i Szkolnictwa Wyższego za wybitne osiągnięcia dla doktorantów
- 06.2010 – laureatka konkursu na płatne staże i praktyki w ramach projektu „Wiedza i kompetencje z fizyki, chemii i informatyki na potrzeby gospodarki – WIKING”
- laureatka VIII edycji konkursu na dofinansowanie staży zagranicznych (finansowanie w ramach projektu Interdyscyplinarne Studia Doktoranckie "Społeczeństwo - Technologie - Środowisko")
- laureatka IX edycji konkursu na dofinansowanie staży zagranicznych (finansowanie w ramach projektu Interdyscyplinarne Studia Doktoranckie "Społeczeństwo - Technologie - Środowisko")
- wyróżnienie w Konkursie na Najlepszego Studenta RP „Studencki Nobel 2012”
- II nagroda w konkursie na najlepszą prezentację z badań własnych podczas XXXIII Ogólnopolskiej Szkoły Chemii za wystąpienie pt. „Machine learning methods as virtual screening tools in computer-aided drug design.”, 10–14.11.2010, Jastrzębia Góra
- 30.05.2012 – nagrody za najlepszy komunikat ustny podczas konferencji „Gotowi do Pracy – Studenckie Prezentacje Badawcze” przyznane przez Sekcję Studencką Polskiego Towarzystwa Chemicznego oraz firmę Amara
- 06.04.2016 – zaproszenie do prowadzenia jednej z sesji wykładowych podczas konferencji GLISTEN Erlangen 2016 Conference