

Pracownia Farmakoepidemiologii i Farmakoekonomiki  
Wydziału Farmaceutycznego  
Collegium Medicum  
Uniwersytetu Jagiellońskiego

**MODELOWANIE *IN SILICO* PROARYTMICZNEGO DZIAŁANIA  
POTENCJALNYCH SUBSTANCJI LECZNICZYCH**

Barbara Wiśniowska

Promotor  
Prof. dr hab. Jerzy Brandys

Kraków 2011

Składam serdeczne podziękowania  
Panu Prof. Jerzemu Brandysowi  
za pomoc i wskazówki udzielone podczas pisania tej pracy.

---

## Spis treści

<b>I. WSTĘP</b>	<b>4</b>
<b>1. PROCES BADAŃ I ROZWOJU PRODUKTU LECZNICZEGO</b>	<b>4</b>
1.1. Etapy rozwoju produktu leczniczego	4
1.1.1. Koszty procesu rozwoju produktu leczniczego	4
1.2. Znaczenie metod alternatywnych	5
<b>2. TOKSYKOLOGIA OBLICZENIOWA</b>	<b>9</b>
<b>3. KARDIOTOKSYCZNOŚĆ</b>	<b>10</b>
3.1. Działanie proarytmiczne	11
3.2. Przedłużenie czasu repolaryzacji i TdP	12
3.3. Inhibicja kanału hERG a polekowa arytmia TdP	14
3.4. Regulacje prawne	15
3.5. Metody wczesnej oceny potencjału proarytmicznego	16
3.5.1. Pomiar transportu jonów rubidu	16
3.5.2. Kompetycyjne wiązanie radioligandów	17
3.5.3. Pomiary fluorescencji	17
3.5.4. Pomiary elektrofizjologiczne	18
3.5.5. Metody <i>in silico</i>	21
<b>4. MODELE KOMPUTEROWE</b>	<b>23</b>
4.1. Modele mechanistyczne i empiryczne	23
4.2. Proces badań i rozwoju oparty na modelach	25
4.3. Modele efektów toksycznych	26
4.4. Jakość modelu	28
<b>5. TECHNIKI MODELOWANIA</b>	<b>29</b>
5.1. Sztuczne sieci neuronowe	29
5.1.1. Sztuczny neuron	29
5.1.2. Sieć neuronowa	30
5.1.3. Proces uczenia ANN	32
5.1.4. Systemy hybrydowe	35
5.1.5. Właściwości i zastosowania	37
5.2. Drzewa decyzyjne	38
5.2.1. Zstępujące konstruowanie drzewa	39
5.2.2. Kryterium wyboru testu	40
5.2.3. Testy kandydujące	41

---

5.2.4. Kryterium stopu	42
5.2.5. Nadmierne dopasowanie	43
5.2.6. Zespoły drzew decyzyjnych (lasy losowe i drzewa wzmacniane)	44
5.3. Metody probabilistyczne – algorytmy Bayesa	45
5.4. Klasyfikacja bayesowska	46
5.4.1. Naiwny klasyfikator bayesowski	46
5.4.2. Sieci bayesowskie (sieci przekonań)	47
5.4.3. Wnioskowanie w sieciach Bayesa	49
5.4.4. Uczenie sieci	49
<b>II. CEL PRACY</b>	<b>51</b>
<b>III. CZĘŚĆ DOŚWIADCZALNA</b>	<b>53</b>
<b>1. BUDOWA BAZY DANYCH</b>	<b>53</b>
1.1. Metodyka	54
1.2. Wyniki	55
1.2.1. Baza danych	55
1.2.2. Charakterystyka rekordów	56
<b>2. WSPÓŁCZYNNIKI EKSTRAPOLACJI</b>	<b>86</b>
2.1. Metodyka	86
2.2. Wyniki	92
2.2.1. Współczynniki ekstrapolacji	92
2.2.2. Walidacja wewnętrzna	93
2.2.3. Walidacja zewnętrzna	94
<b>3. MODELE</b>	<b>101</b>
3.1. Metodyka	101
3.1.1. Oprogramowanie i sprzęt komputerowy	101
3.1.2. Wektor wejściowy	104
3.1.3. Wartość wyjściowa	108
3.1.4. Budowa modeli – testowane parametry	108
3.1.5. Weryfikacja modeli	114
3.1.6. Ocena błędu	117
3.1.7. Analiza wrażliwościowa	119
3.1.8. Notacja	119
3.2. Wyniki	121
3.2.1. Analiza wrażliwościowa	121
3.2.2. Modelowanie	124

---

---

<b>IV. PODSUMOWANIE WYNIKÓW I DYSKUSJA</b>	<b>130</b>
<b>1. BAZA DANYCH</b>	<b>130</b>
<b>2. WSPÓŁCZYNNIKI EKSTRAPOLACJI</b>	<b>132</b>
<b>3. MODELE</b>	<b>133</b>
3.1. Analiza wrażliwościowa	133
3.2. Modelowanie	134
3.2.1. Odciski cząsteczek	134
3.2.2. Algorytmy	134
3.2.3. Komitety ekspertów	135
<b>V. WNIOSKI</b>	<b>147</b>
<b>VI. STRESZCZENIE</b>	<b>148</b>
<b>VII. OZNACZENIA I SYMBOLE</b>	<b>149</b>
<b>VIII. PUBLIKACJE</b>	<b>151</b>
<b>IX. PIŚMIENNICTWO</b>	<b>152</b>

## I. WSTĘP

### 1. PROCES BADAŃ I ROZWOJU PRODUKTU LECZNICZEGO

Rozwój leku począwszy od obiecującego odkrycia do końcowego produktu leczniczego jest procesem długim i kosztownym [361]. Ostatecznym celem tego procesu jest identyfikacja cząsteczki wywołującej w organizmie ludzkim pożądany efekt oraz ustalenie efektywności i bezpieczeństwa terapii prowadzonej z jej zastosowaniem [323].

#### 1.1. Etapy rozwoju produktu leczniczego

Współcześnie proces identyfikacji nowego leku i jego wprowadzenia na rynek obejmuje kilka głównych etapów. Pierwszym jest synteza ogromnej liczby cząsteczek, które mogą wchodzić w interakcje ze specyficznymi białkami czy wpływać na procesy enzymatyczne żywych komórek. Następnym etapem jest tzw. skринing czyli badanie przesiewowe mające na celu wyodrębnienie cząsteczek, które mają największą szansę na spełnienie wymagań stawianym potencjalnym lekom podczas badań klinicznych. W kolejnym etapie wybrane nowe substancje, będące potencjalnymi komponentami leku, poddawane są złożonemu cyklowi biologicznych, farmakologicznych i toksykologicznych testów w warunkach *in vitro*. W tym wczesnym stadium badań eliminowana jest największa część kandydatów na lek. Cząsteczki, które pozytywnie przejdą cykl rygorystycznych testów badane są w warunkach *in vivo* na zwierzętach. Wszystkie informacje gromadzone podczas testów przedklinicznych wykorzystywane są następnie do ukierunkowania badań klinicznych przeprowadzanych z udziałem ochotników i pacjentów.

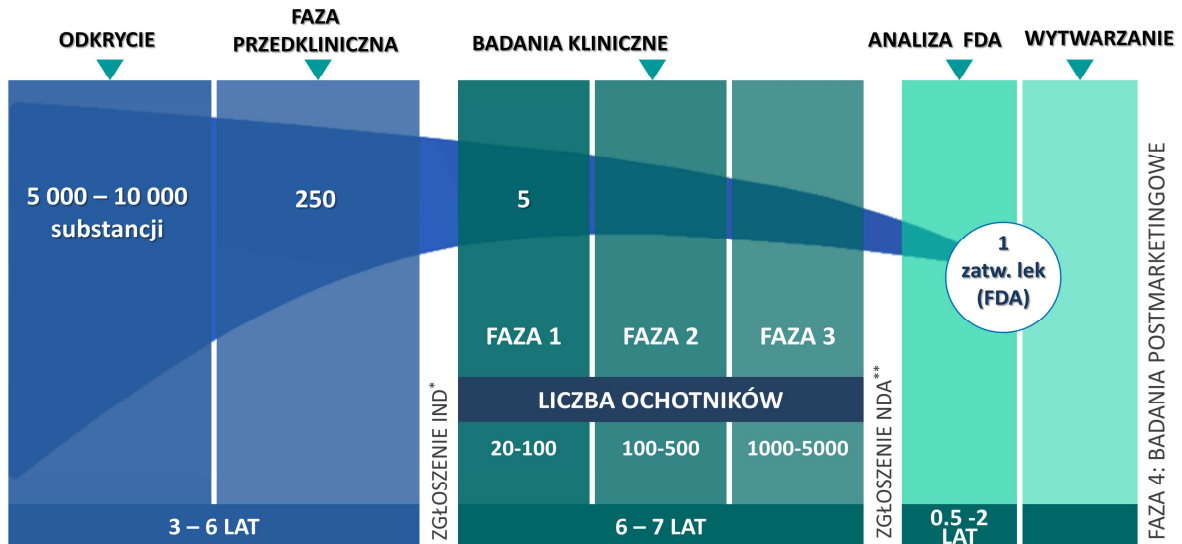
##### 1.1.1. Koszty procesu rozwoju produktu leczniczego

Według Amerykańskiej Agencji ds. Żywności i Leków (FDA) zanim cząsteczki eksperymentalne zostaną zaaprobowane do szerokiego stosowania badane są średnio 12 do 15 lat. Podczas procesu rozwoju nowego leku z 5 000 cząsteczek wybranych do badań przedklinicznych, jedynie 5 (średnio) dociera do fazy badań klinicznych z udziałem ludzi, z których jedna zostanie później zatwierdzona jako lek (Rycina 1).

Szacowany koszt wprowadzenia jednego nowego leku na rynek waha się w szerokim zakresie, od 800 milionów dolarów do prawie 2 miliardów [47,63,64]. Najbardziej aktualny raport Tufts Center for the Study of Drug Development ustala ten koszt w roku 2006 na 1.2 miliarda dolarów [309].

Wydatki te nie rozkładają się równomiernie na poszczególne etapy badań potencjalnej substancji leczniczej i wdrażania leku a rosną wykładniczo (Rycina 2), większość pochłania przygotowanie i przeprowadzenie całości badań klinicznych, zgodnie z obowiązującymi - coraz surowszymi zarówno pod kątem naukowym jak i etycznym -

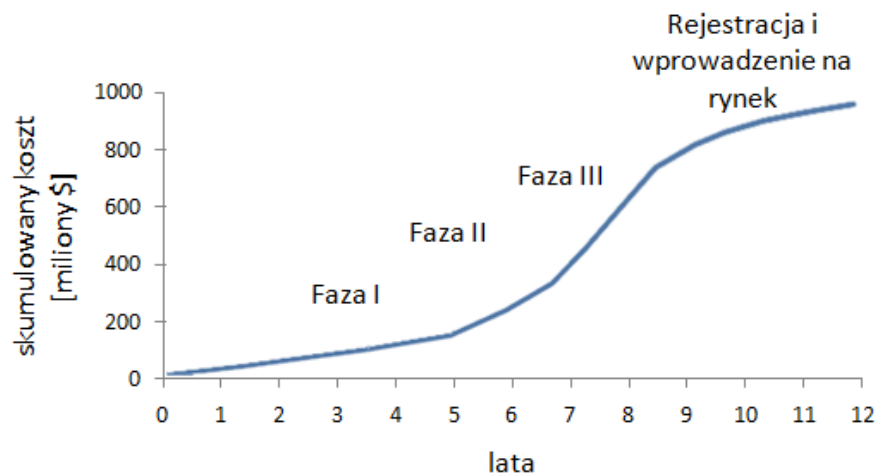
procedurami i zasadami. Niepowodzenie na tym etapie wiąże się ze znacznymi stratami zarówno czasu poświęconego na doprowadzenie cząsteczki do tej fazy badań jak i środków finansowych na to przeznaczonych.



\* NDA (ang. *New Drug Application*) – nowy lek po III fazie badań klinicznych

\*\* IND (ang. *Investigational New Drug*) – badany produkt leczniczy

Rycina 1. Ramy czasowe procesu rozwoju nowego leku [263].



Rycina 2. Skumulowany koszt poszczególnych faz badań nad lekiem [63].

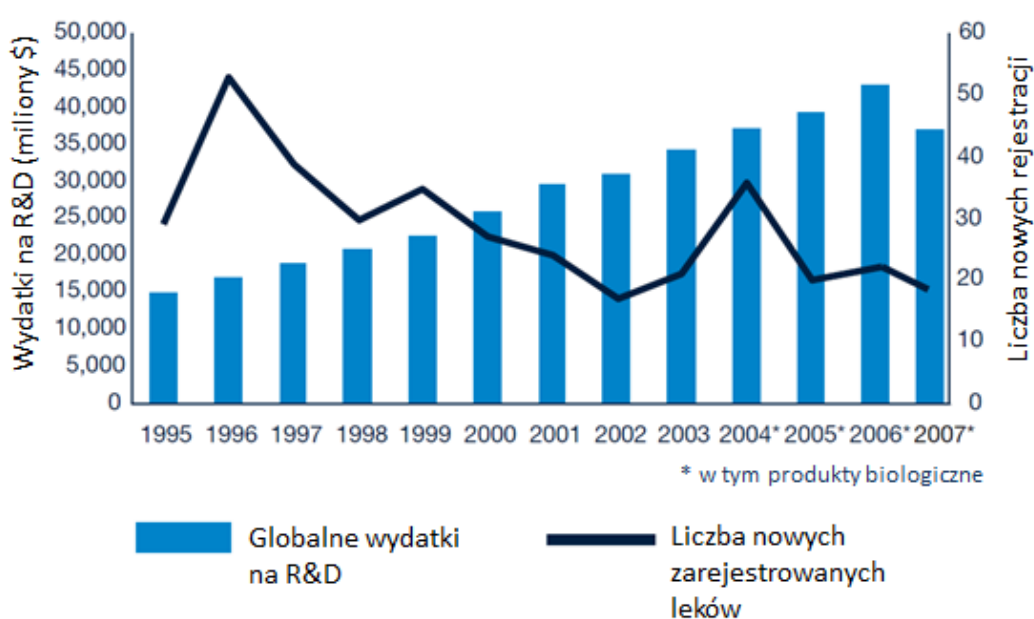
## 1.2. Znaczenie metod alternatywnych

Zwiększenie kosztów wprowadzenia nowego leku na rynek i zmniejszenie opłacalności całego procesu powodowane jest przede wszystkim znacznym wzrostem wymagań, dotyczących m.in. bezpieczeństwa, stawianych nowym lekom przez agencje rejestrujące (FDA, EMA) skutkującym zwiększeniem odsetka cząsteczek odrzucanych na kolejnych etapach rozwoju potencjalnych leków. To z kolei jest przyczyną wciąż malejącej

liczby rejestracji, rozumianych jako wprowadzenie na rynek nowych cząsteczek, a więc leków określanych mianem innowacyjnych (Rycina 3), pomimo że przemysł farmaceutyczny jest jedną z najintensywniej inwestujących w naukę oraz badania i rozwój (BiR, ang. *R&D – Research and Development*) gałęzi gospodarki [272 - raporty Wspólnotowego Centrum Badawczego Komisji Europejskiej - JRC].

Coraz wyraźniej podkreśla się także względy humanitarno-etyczne wskazujące na konieczność uzyskania maksymalnej możliwej redukcji doświadczeń prowadzonych z wykorzystaniem zwierząt, zastępowania ich innymi, równocennymi metodami badawczymi zgodnie z zasadą 3R (ang. *Refine, Reduce, Replace*) oraz optymalizacji niezbędnych badań prowadzonych z udziałem ludzi (zdrowych ochotników oraz osób chorych, w warunkach klinicznych). Ponadto wartość prognostyczna doświadczeń na zwierzętach, zarówno pod względem skuteczności jak i toksyczności, nie jest optymalna m.in. ze względu na niemożliwe do usunięcia różnice międzygatunkowe utrudniające, a niekiedy uniemożliwiające ekstrapolowanie wyników badań wykonanych na zwierzętach bezpośrednio na ludzi [252].

Sytuacja ta powoduje, że priorytetem programów rozwoju leków oraz oceny bezpieczeństwa przeprowadzanej przez organy nadzorujące i regulujące jest możliwie wczesna identyfikacja poważnych, związanych z toksycznością zagrożeń zanim poczynione zostaną znaczne inwestycje zasobów i czasu oraz wykorzystanie do tego celu metod alternatywnych. Efektem tego jest rosnące zainteresowanie technikami i narzędziami modelowania matematycznego [89,92,97,235,369,381]. Szacuje się, że obecnie około 10% środków przeznaczanych przez przemysł farmaceutyczny na badania i rozwój inwestowanych jest w obszarze modelowania komputerowego i symulacji, a do roku 2016 wydatki te wzrosną do 20% [369].



Rycina 3. Obniżenie produktywności procesu badań i rozwoju leków [97].



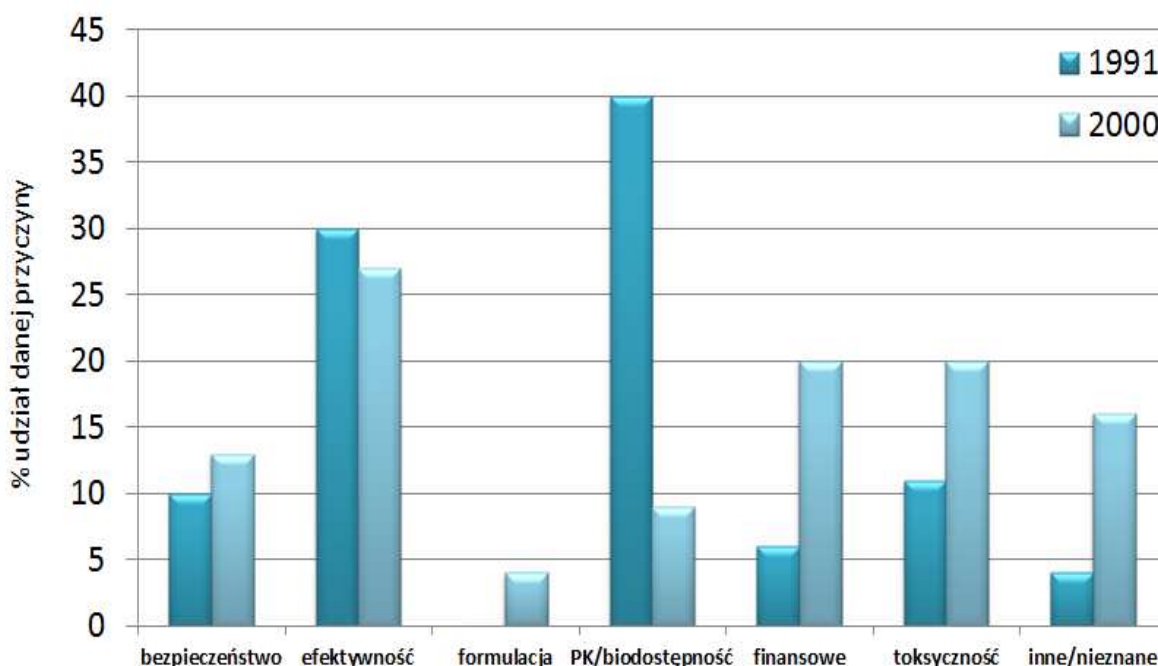
Wytyczne FDA dotyczące nowych idei badań naukowych „Critical Path Initiative” wskazują na konieczność opracowywania nowoczesnych, bardziej efektywnych kosztowo, metod oceny bezpieczeństwa leków, a także zanieczyszczeń obecnych w preparatach. Między innymi sugerowane jest zastosowanie wysokowydajnych metod, w tym technik drążenia danych (*ang. data mining*) i metod modelowania komputerowego QSAR, jako wstępnych narzędzi skryningowych pozwalających na prognozowanie potencjalnych działań niepożądanych projektowanego leku oraz lepsze planowanie dalszych testów niezbędnych na kolejnych etapach badań [147,194,367,368]. Również obowiązujące od roku 2007 w Unii Europejskiej rozporządzenie Parlamentu Europejskiego i Rady Europy REACH (*ang. Registration, Evaluation, Authorisation and Restriction of Chemicals*) wprowadzające uregulowania w zakresie rejestracji, oceny, udzielania zezwoleń oraz stosowanych ograniczeń w zakresie substancji chemicznych [94] kładzie duży nacisk na wykorzystywanie podczas procesu oceny i rejestracji związków chemicznych alternatywnych metod badań, takich jak predyktywne modele QSAR. Kolejnym obok REACH przykładem praktycznego wykorzystania komputerowych modeli toksyczności i ich predykcji jest Canadian Environmental Protection Act [35,228]. Zgodnie z tym dokumentem wyniki modelowania używane są jako równorzędne dla dostępnych danych z innych badań do kategoryzacji substancji chemicznych znajdujących się w kanadyjskim Wykazie Substancji Krajowych (DSL – *ang. Domestic Substances List*). Oceną potencjału toksykologii predykcyjnej zajmuje się także zespół ICSAS (*ang. Informatics and Computational Safety Analysis Staff*), będący częścią Centrum do spraw Oceny i Badań nad Lekami (CDER – *ang. Center for Drug Evaluation and Research*) Amerykańskiej Agencji ds. Leków i Żywności czy też laboratorium doświadczalne amerykańskiego Centrum Toksykologii Obliczeniowej (*ang. National Center for Computational Toxicology*), działającego w ramach rządowej organizacji badającej i chroniącej środowisko (EPA – *ang. Environmental Protection Agency*). Zespoły te we współpracy z firmami informatycznymi, farmaceutycznymi, chemicznymi opracowują toksykologiczne bazy danych, tworzą i oceniają techniki drążenia danych i metody QSAR, dążąc do stworzenia zestawu narzędzi do oceny efektów (mutagenności, kancerogenności i in.) oraz dawek toksycznych.

Powyższe przykłady jak i deklaracje przemysłu farmaceutycznego wskazują, że realizowane komputerowo modele (modele *in silico*) znajdują zastosowanie na wszystkich etapach rozwoju leku począwszy od poszukiwania i projektowania nowych aktywnych struktur, skryningu wirtualnych lub zsyntetyzowanych bibliotek związków, poprzez przewidywanie i projektowanie właściwości ADME, ocenę prawdopodobieństwa wiązania z docelową strukturą białkową, aż do oceny bezpieczeństwa i skuteczności potencjalnego leku. Modelowanie *in silico* wykorzystywane jest także na etapie decyzji o dopuszczeniu nowych produktów do obrotu, bieżącego oraz okresowego monitorowania procesu produkcji czy efektów stosowania leku w populacji (tzw. IV fazie badań klinicznych).

Zadaniem modeli komputerowych jest poprawa prawidłowości prognoz stawianych podczas kolejnych etapów badań w oparciu o istniejącą wiedzę [195], a w

konsekwencji zwiększenie efektywności przemysłu farmaceutycznego [40,195,226]. Metody *in silico*, poza względami ekonomicznymi, oferują także znacznie większą w porównaniu z testami *in vivo* i *in vitro*, wydajność skringu cząsteczek, odpowiadającą potrzebom wynikającym z m.in. rozwoju metod chemii kombinatorycznej. Doświadczenia *in silico* przeprowadzane zarówno na realnie istniejących, zsyntetyzowanych cząsteczkach jak i na cząsteczkach wirtualnych, pozwalają do etapu syntezy czy badań *in vitro* wybrać te najbardziej obiecujące. Dzięki temu liczba struktur, które poddawane będą skringowi w warunkach *in vitro* może być odpowiednia do zastosowania dokładniejszych, jednak cechujących się niską lub średnią wydajnością technik. Odpowiedni dobór struktur może także prowadzić do ograniczenia liczby koniecznych do wykonania w dalszych etapach badań na zwierzętach.

Obserwowane w ostatnim czasie znaczne zmniejszenie ilości niepowodzeń procesu BiR związanych z nieodpowiednimi właściwościami ADME, jest m.in. efektem rutynowego stosowania modelowania komputerowego do ich projektowania i oceny, podczas wstępnego etapu badań potencjalnych leków (Rycina 4). Podobnych efektów można spodziewać się jako wyniku skringu i modelowania właściwości toksycznych, które są jedną z podstawowych przyczyn niepowodzeń na etapie badań nad lekiem oraz wycofywania zarejestrowanych produktów leczniczych [190].



Rycina 4. Przyczyny eliminacji potencjalnych leków podczas procesu badań i rozwoju [190].

## 2. TOKSYKOLOGIA OBLICZENIOWA

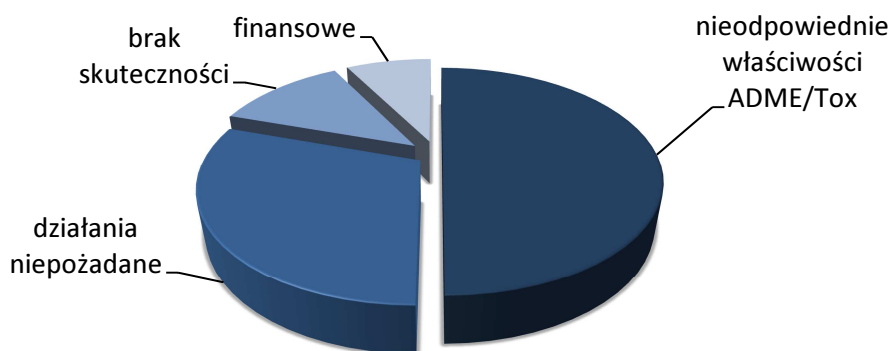
Krytyczne znaczenie zagadnienia bezpieczeństwa leku w procesie jego tworzenia i rejestracji oraz ciągły intensywny wzrost znaczenia komputerowych metod obliczeniowych wraz z rosnącym zainteresowaniem narzędziami modelowania jest przyczyną obserwowanego w ostatnim czasie dynamicznego rozwoju nowej dziedziny nauki jaką jest toksykologia obliczeniowa (*ang. computational toxicology*) [89,92,367,368]. Zgodnie z definicją Amerykańskiej Agencji Ochrony Środowiska (EPA) toksykologia obliczeniowa to zastosowanie modeli matematycznych i komputerowych w celu przewidywania efektów niepożądanych oraz lepszego zrozumienia prostych lub złożonych mechanizmów, poprzez które dana substancja chemiczna wywiera działanie szkodliwe. Modelowanie toksyczności od kilku lat znajduje się w centrum zainteresowania przede wszystkim firm farmaceutycznych. Toksykologiczne metody *in silico* na etapie projektowania nowych leków, wykorzystywane są przy wyborze struktur wiodących wykazujących najmniejszą toksyczność, a zatem tych, które następnie będą podlegały procesowi optymalizacji i dalszemu rozwojowi z największym prawdopodobieństwem sukcesu rozumianego jako rejestracja nowego leku [240,257]. Choć modelowanie i predykcja toksyczności w warunkach *in silico* nie jest jeszcze metodą rutynowo stosowaną podczas badań nad lekiem, konieczność rozwoju alternatywnych metod oceny toksyczności leków dostrzega nie tylko przemysł farmaceutyczny.

Jak wspomniano techniki modelowania komputerowego wspomagają obecnie toksykologiczne i farmakologiczne profilowanie substancji [240,257] oraz wspierają decyzje organów nadzorujących dotyczące bezpieczeństwa chemicznego i ryzyka działań niepożądanych [13,120]. Wyniki tego rodzaju doświadczeń są już wykorzystywane np. przez amerykańską Agencję Ochrony Środowiska (EPA), Narodowy Instytut Zdrowia (NIH) czy instytucje rejestrujące leki podczas procesu decyzyjnego, zwłaszcza gdy wyniki badań bezpieczeństwa leku nie są jednoznaczne, a także gdy decyzja musi być podjęta w sytuacji braku wszystkich pożądanых informacji (produkty degradacji, metabolity, zanieczyszczenia). Dane uzyskane w wyniku modelowania komputerowego dla strukturalnie podobnych związków efektywnie usprawniają także prowadzone, m.in. w Stanach Zjednoczonych reewaluacje substancji leczniczych [367].

### 3. KARDIOTOKSYCZNOŚĆ

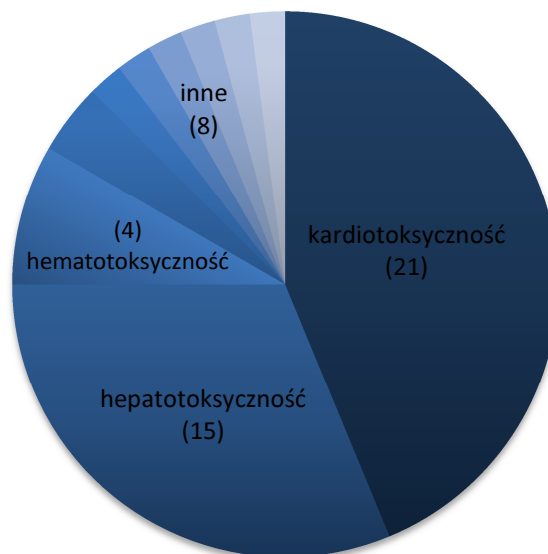
Regulacje dotyczące prowadzenia badań nowych substancji będących potencjalnymi komponentami leków muszą zbalansować dwa przeciwstawne cele jakimi są zapewnienie bezpieczeństwa leków oraz ułatwienie pacjentom dostępu do nowych terapii. Ocena leku przed jego wprowadzeniem na rynek obejmuje badania kliniczne bezpieczeństwa i skuteczności. Testy te są jednak stosunkowo krótkie i przeprowadzane są z udziałem małych, ściśle zdefiniowanych grup pacjentów. Dlatego po zaaprobowaniu leku, gdy jest on dostępny dla ogółu pacjentów, jest stosowany w nowych wskazaniach, a czas terapii czy przyjmowane dawki znacznie różnią się od warunków symulowanych w badaniach klinicznych, mogą pojawiać się nieobserwowane wcześniej działania niepożądane. Poważne i występujące ze znaczną częstotliwością czy wyraźnie zmieniające współczynnik ryzyka do korzyści konsekwencje zdrowotne tych efektów są przyczyną wprowadzania istotnych ograniczeń stosowania leku lub nawet jego wycofania z obrotu.

Właściwości toksyczne, obok braku efektywności i niewłaściwego profilu ADME są jedną z głównych przyczyn niepowodzeń podczas rozwoju nowych potencjalnych leków (Rycina 5), a także po wprowadzeniu leku na rynek.



Rycina 5. Główne problemy BiR.

Analiza dostępnych publikacji przeglądowych [110,202,206,251,308,381,394] oraz raportów Centrum do spraw Oceny i Badań nad Lekami (CDER) [34] pokazała, że spośród 48 przypadków wycofania leków w ubiegłych latach, 15 związanych było z ujawnieniem działań hepatotoksycznych i 21 z działaniami niepożądanymi w obrębie układu krążenia, w tym 11 z działaniem proarytmicznym (Rycina 6). Działanie hepato-, neuro- i kardiotoxyczne są również głównymi przyczynami niepowodzeń na etapie badań klinicznych [252,328]. Ocenia się, że największy wpływ na zmniejszenie ryzyka podczas badań klinicznych będzie miała redukcja tych trzech przyczyn oraz że redukcja liczby wycofań z powodu kardio- i hepatotoksyczności o połowę spowoduje zmniejszenie ogólnej liczby wycofań aż o 40% [328].

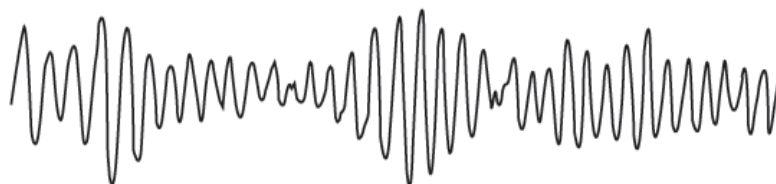


**Rycina 6.** Główne przyczyny wycofań leków z rynku.

Kardiotoxyczność może być wynikiem działania leków powodujących upośledzenie funkcji lub nekrozę kardiomiocytów, uszkodzenia zastawek, niedokrwienie mięśnia sercowego i/lub arytmie komorowe. Jej znaczenie dla pacjentów, organów regulujących i nadzorujących bezpieczeństwo stosowania leków oraz przemysłu farmaceutycznego jest ogromne [26,52,56,167,275]. Szczególną uwagę zwrócono w ostatnim czasie na działanie proarytmiczne będące istotną przyczyną wycofań produktów leczniczych ze sprzedaży, a także wprowadzenia znacznych ograniczeń stosowania niektórych preparatów.

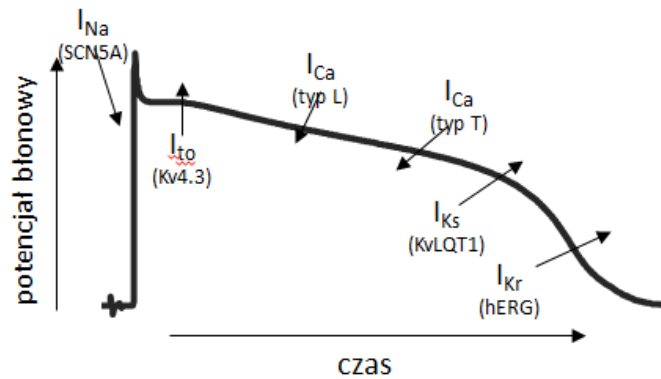
### 3.1. Działanie proarytmiczne

Ryzyko działania proarytmicznego oceniane jest jako możliwość wystąpienia częstoskurczu komorowego torsades de pointes (TdP). TdP to rzadka postać tachykardii komorowej charakteryzująca się zmiennym kierunkiem wychyleń i różnym kształtem zespołów QRS w zapisie EKG (Rycina 7). W większości przypadków arytmia TdP ulega samoograniczeniu, może jednak przejść w zagrażające życiu migotanie komór.



**Rycina 7.** Zapis EKG arytmii TdP.

Przyczyną tego rodzaju arytmii jest przede wszystkim widoczne w zapisie EKG wydłużenie odstępu QT, choć w bardzo rzadkich przypadkach może być efektem poszerzenia kompleksu QRS [5,370]. Na poziomie komórkowym zaburzenie depolaryzacji i wydłużenie odstępu QT wyraża się w postaci przedłużenia czasu trwania potencjału czynnościowego komórek serca (AP).



Rycina 8. Rola kanałów jonowych w powstawaniu potencjału czynnościowego.

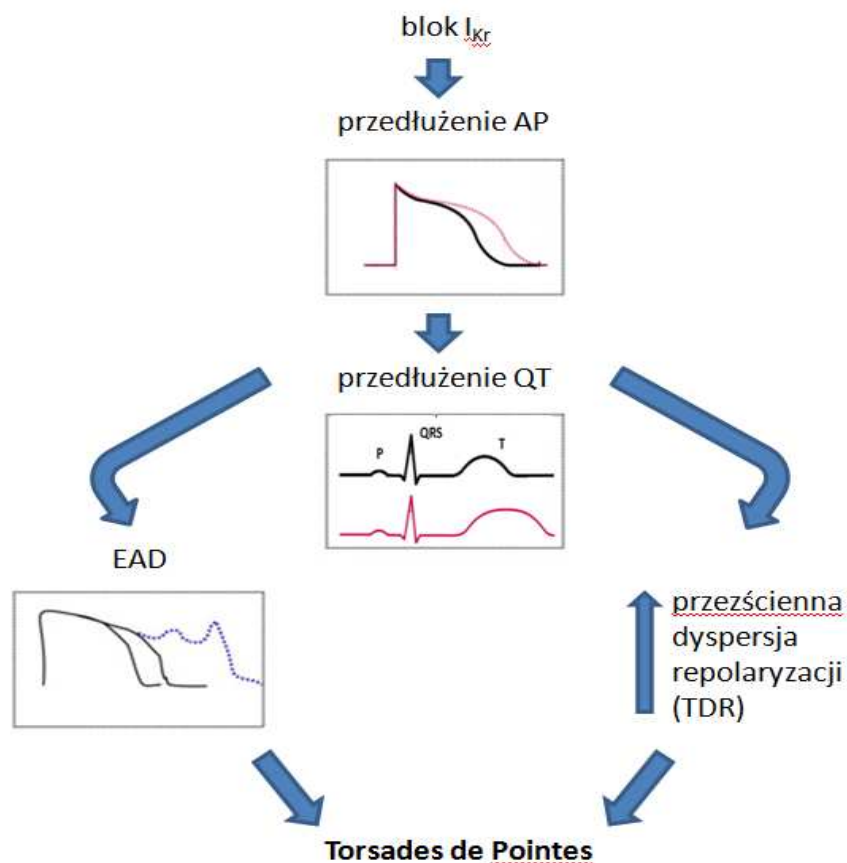
Na rycinie 8 przedstawiono główne prądy i kanały jonowe uczestniczące w generacji potencjału czynnościowego. Wydłużenie odstępu QT może być związane z mutacjami genów kodujących podjednostki kanałów jonowych biorących udział w powstawaniu potencjału czynnościowego (wrodzony zespół LQTS), bądź spowodowane przez czynniki egzogenne, w tym leki [59,284]. Przedłużenie czasu repolaryzacji może powstawać na drodze kilku mechanizmów. Stwierdzono jednak, że mechanizmem prowadzącym do wydłużenia odstępu QT dla wszystkich leków wykazujących działanie proarytmiczne jest hamowanie szybkiego odkomórkowego prądu potasowego, oznaczanego skrótem  $I_{Kr}$  [125,126,243], w późnej fazie repolaryzacji komórek mięśnia sercowego. To z kolei jest efektem inhibicji kanału błonowego, którego podjednostka  $\alpha$  kodowana jest przez gen hERG - *human ether-a-go-go-related gene* [274,293].

### 3.2. Przedłużenie czasu repolaryzacji i TdP

Przedłużenie czasu trwania repolaryzacji komórek mięśnia sercowego nie jest samo w sobie warunkiem wystarczającym powstania arytmii, a efekt taki jest w istocie podstawą mechanizmu działania leków antyarytmicznych klasy III [130,321]. Hamowanie  $I_{Kr}$  lub  $I_{hERG}$  nie stanowi bezpośrednio przyczyny TdP, ale raczej prowadzi do zmian elektrofizjologicznych na poziomie komórki i tkanki, które mogą ułatwić rozpoczęcie i podtrzymanie nieprawidłowego rytmu. Wydłużenie repolaryzacji jest łączone z arytmogenezą zwykle na dwa sposoby (Rycina 9).

Na poziomie komórkowym nadmierne wydłużenie AP powiązane jest z tzw. wczesnymi depolaryzacjami następczymi (*ang. early after depolarization* - EAD) czyli depolaryzacjami pojawiającymi się w późnej fazie repolaryzacji [372]. EAD nasilają się przy

wolnym rytmie serca, jednoczenie blokery prądu  $I_{Kr}$  wywołują przedłużenie AP odwrotnie proporcjonalne do częstości rytmu [66,130]. Uważa się, że EAD są następstwem głównie ponownego otwierania kanałów wapniowych typu L w fazie repolaryzacji, możliwe, że pewną rolę odgrywa również wymiennik sodowo-wapniowy [372]. Generacja wczesnych potencjałów następczych może być z kolei czynnikiem wyzwalającym arytmie typu torsades de pointes (oraz uczestniczyć w powstawaniu związanego z arytmia zjawiska „R-na-T”) [201]. Szczególną wrażliwość, na indukowane lekami przedłużenie AP i EAD, wykazują włókna Prukinięgo, charakteryzujące się ograniczoną rezerwą repolaryzacyjną [29,81]. Drugą drogą, na jakiej przedłużenie repolaryzacji może predestynować do wystąpienia arytmii TdP jest przezścienna dyspersja okresu repolaryzacji (*ang. transmural dispersion of repolarization, TDR*) [6,7,201,310]. TDR jest efektem heterogennej ekspresji kanałów w ścianie komór i wpływa na morfologię załamka T w zapisie EKG [6,7]. Osłabienie prądu  $I_{Kr}$  ma silniejszy wpływ na repolaryzację mid-miokardium niż innych regionów ścian komór (endo-, epikardium), co wynika z faktu późniejszej repolaryzacji i mniejszej liczby kanałów  $I_{Ks}$  w tych komórkach [29]. Właściwości te sprzyjają nasileniu TDR. Prowadzi to do zwiększenia niejednorodności refrakcyjności i tworzenia korzystnych warunków rozwoju arytmii re-entrant [6,7,201].



Rycina 9. Zależność inhbicja prądu  $I_{Kr}$  – TdP.

### 3.3. Inhibicja kanału hERG a polekowa arytmia TdP

Charakterystyczną cechą polekowej arytmii TdP jest różnorodność (zarówno grup terapeutycznych jak i pod względem struktury) substancji kojarzonych z jej występowaniem w praktyce klinicznej. Wydłużanie czasu repolaryzacji i wywoływanie zaburzeń rytmu serca, arytmii typu torsades de pointes, jest znanym działaniem niepożądanym stosowania wielu leków antyarytmicznych [3,57,98,267]. Począwszy od wczesnych lat 90-tych ubiegłego wieku obserwowano rosnącą liczbę doniesień postmarketingowych dotyczących występowania tego specyficznego działania kardi toksycznego podczas stosowania leków pozbawionych aktywności przeciwoarytmicznej.

Jednymi z pierwszych leków, które zwróciły uwagę instytucji nadzorujących bezpieczeństwo stosowania farmaceutyków na problem polekowych zaburzeń rytmu związanych z wydłużaniem odstępu QT, arytmia typu torsades de pointes i przypadkami nagłej śmierci sercowej były dwa leki przeciwhistaminowe - terfenadyna i astemizol, [388,389,400]. Działania niepożądane terfenadyny, zależne od stężenia, występowały po przekroczeniu maksymalnej dawki terapeutycznej, ale także już w zakresie dawek terapeutycznych, u pacjentów stosujących równocześnie leki hamujące enzymy cytochromu P-450 czy z wrodzonym zespołem wydłużonego odstępu QT (zespół LQTS) [388]. Nowsze leki antyhistaminowe (ebastyna, loratadyna, ceteryzyna, feksofendyna i in.) wymagają dalszych badań, które pozwolą określić ich bezpieczeństwo sercowe, jednak niektóre z nich blokują szybki prąd potasowy i potencjalnie mogą powodować arytmie. Częstość występowania działań kardi toksycznych leków przeciwhistaminowych jest niska [216], jednak ryzyko arytmii musi być rzetelnie ocenione ze względu na powszechność przepisywania i stosowania tego rodzaju leków w terapii chorób niezagrażających życiu i samoograniczających się. Wydłużenie odstępu QT i torsades de pointes mogą być wywołane również przez niektóre antybiotyki (makrolidy i fluorochinolony), leki antymalaryczne czy imidazolowe leki przeciwgrzybicze [57,136,231,248]. Arytmie torsades de pointes są jednak rzadkie, a ich wystąpienie związane jest ze stosowaniem tylko niektórych leków przeciwbakteryjnych wydłużających odstęp QT. Podawana dożylnie erytromycyna przedłuża czas trwania potencjału czynnościowego, powoduje dyspersję repolaryzacji komórek oraz czasami arytmie typu TdP [273]. W grupie chinolonów przedłużenie odstępu QT udowodniono dla sparfloksacyny oraz grepafloksacyny, obecnie wycofanych z rynku większości krajów, natomiast lewofloksacyna i ofloksacyna uznawane są za bezpieczne. Chinina [131], podobnie jak halofantryna, wydłuża odstęp QT już w dawkach zwykle stosowanych, zwłaszcza w połączeniu z meflochiną [248]. Ketokonazol wykazuje bezpośrednie działanie proarytmiczne, a także przedłużając zależny od cytochromu P-450 metabolizm innych leków wydłużających odstęp QT [136]. Trójcykliczne leki antydepresyjne, amitryptylina, doksepina, dezipramina, imipramina i klomipramina łączone są z wydłużaniem odstępu QT oraz przypadkami nagłej śmierci sercowej (dezipramina, imipramina, klomipramina) [17]. Z takim efektem wiąże się



również stosowanie dawek terapeutycznych i toksycznych neuroleptyków, takich jak chloropromazyna, flufenazyna, haloperidol, prochloroperazyna, tiorydazyna czy trifluoperazyna [128]. Brytyjski Komitet ds. Bezpieczeństwa Leków (*ang. Committee on Safety of Medicines – CSM*) w okresie od 1971-1995 odnotował 40 raportów (w tym 16 zgonów) poważnych działań niepożądanych w obrębie układu krążenia, głównie arytmii, dotyczących kolejnego leku antypsychotycznego – pimozydu [48]. Sertindol został dobrowolnie wycofany przez producenta i skierowany do ponownej oceny ze względu na sprzeczne wyniki badań pre- i postmarketingowych [18]. Sporo uwagi, w związku z pojawieniem się raportów dotyczących wydłużania odstępu QT i TdP, poświęcono także cyzaprydowi [393].

Pomimo różnorodności grup terapeutycznych i struktur substancji wszystkie one posiadają wspólny efekt działania – redukcję prądu potasowego hERG poprzez bezpośrednią blokadę farmakologiczną lub zaburzenie dojrzewania i transportu podjednostek kanału do błony komórkowej. Jak do tej pory tak silny związek z przedłużaniem odstępu QT i TdP nie został udowodniony dla żadnego innego kanału jonowego, poza kanałem hERG [27].

#### 3.4. Regulacje prawne

Do roku 2009 wydłużanie odstępu QT w zapisie EKG było przyczyną wycofania ze sprzedaży 11 substancji, a dla wielu innych wprowadzenia znacznych ograniczeń stosowania.

Wynikiem rosnącego zainteresowania tym problemem było sformułowanie przez Europejską Agencję Leków (*ang. European Medicines Agency - EMA*) dokumentu „punkty do rozważenia” (*ang. „points to consider”*) [93], w którym zaproponowano kilka przedklinicznych testów do określania ryzyka arytmii serca projektowanych leków [19,93], a następnie opracowane wspólnie przez FDA i Health Canada i wdrożone w UE w 2005 wytyczne ICH (*ang. International Conference of Harmonization*) regulujące zasady badań i rejestracji nowych leków [14,46,96]. Metody przedklinicznej oceny potencjału proarytmicznego cząsteczek specyfikują wytyczne ICH S7B [144]. Zalecane badania mogą być wykonywane na trzech poziomach: 1) pomiary parametrów zapisu EKG, wraz z oceną ciśnienia tętniczego krwi i częstości akcji serca, u zwierząt przytomnych lub pod narkozą, 2) ocena charakterystyki potencjału czynnościowego z wykorzystaniem izolowanych tkanek lub zwierząt pod narkozą oraz 3) badania przesiewowe w kierunku właściwości blokowania kanału hERG z wykorzystaniem kultur komórek sercowych, linii komórkowych wykazujących ekspresję ludzkich kanałów jonowych serca, perfundowanych preparatów sercowych (metoda *Langendorffa*) czy preparatów izolowanych tkanek (włókna Purkiniego, mięśnie brodawkowe miokardium) pochodzących od zwierząt laboratoryjnych (kroliki, świnki morskie, psy i świnie).

### 3.5. Metody wczesnej oceny potencjału proarytmicznego

Skrining cząsteczek w kierunku blokowania kanału hERG jest jedną z metod oceny właściwości proarytmicznych zalecanych przez wytyczne S7B ICH. Siła blokowania kanału hERG jest podstawowym wyznacznikiem potencjału kardi toksycznego danej substancji. Ocena wartości  $IC_{50}$  (stężenia powodującego zmniejszenie prądu  $I_{Kr}$  o połowę) jest wstępnym testem bezpieczeństwa kardiologicznego badanej cząsteczki i stanowi podstawę decyzji o kontynuacji bądź przerwaniu dalszych badań we wczesnej fazie procesu rozwoju leku [93,116,294]. Zgodnie z wytycznymi ICH do ilościowej oceny interakcji lek-kanał hERG w warunkach *in vitro* wykorzystuje się badania przepływu rubidu, wiązania radioaktywnych ligandów, pomiary fluorescencji oraz elektrofizjologiczne [20,67,105,245]. „Złoty standard” oznaczeń stanowią obecnie pomiary elektrofizjologiczne techniką patch clamp umożliwiającą badanie przepływu jonów przez pojedyncze kanały przy jednoczesnej kontroli wartości potencjału błonowego komórki.

#### 3.5.1. Pomiar transportu jonów rubidu

Kanały potasowe wykazują wysoką przewodność dla jonów rubidu ze względu na ich zbliżone do jonów potasowych rozmiary, dlatego też ocena wpływu badanej substancji na transport rubidu podczas repolaryzacji błony komórkowej może być zastosowana jako miara potencjału blokowania kanału potasowego hERG tej substancji. Podczas inkubacji systemu testowego wewnątrzkomórkowe jony potasu zastępowane są jonami  $Rb^+$ . W następnym etapie komórki inkubowane są z roztworem substancji badanej i depolaryzowane za pomocą wysokich stężeń jonów potasowych  $K^+$ . W obecności substancji blokującej hERG przepływ rubidu przez kanał potasowy jest całkowicie lub częściowo zahamowany. W przypadku gdy badana substancja nie posiada właściwości blokowania hERG transport rubidu pozostaje niezakłócony. W doświadczeniu mierzona jest dystrybucja jonów rubidu pomiędzy środowiskiem wewnętrznym komórki a supernatantem [105,341].

Metody ilościowej oceny transportu jonów rubidu obejmują zarówno techniki wykorzystujące rubid radioaktywny ( $^{86}Rb^+$ ) wraz z metodami analizy znaczników oraz nieradioaktywne jony rubidu i metody atomowej spektroskopii absorpcyjnej (AAS) [304].

Techniki pomiaru transportu jonów rubidu zapewniają większą wydajność w porównaniu ze standardową techniką patch-clamp, są także znacznie tańsze. Otrzymywane wyniki są jednak wysoce zależne od warunków przeprowadzenia eksperymentu, jak czas oznaczenia czy kinetyka interakcji substancji z kanałem hERG, czego rezultatem jest mniejsza spójność i zgodność z wynikami otrzymywanymi techniką patch-clamp. Nie bez znaczenia jest również problem radioaktywnych odpadów. Pomiary z wykorzystaniem ASS unikają radioaktywności, jednak ich wydajność oceniana jest jako umiarkowana [246,304].

### 3.5.2. Kompetycyjne wiązanie radioligandów

Istotą tego rodzaju oznaczeń jest wypieranie znakowanych radioaktywnie silnych blokerów kanału hERG jak np. [<sup>3</sup>H] dofetylid, [<sup>3</sup>H] astemizol, [<sup>35</sup>S] MK-499, [<sup>125</sup>I] BeKm-1, przez cząsteczki substancji badanej [104,378].

Początkowo kanały hERG wybranego systemu testowego są całkowicie wysyczone znakowaną radioaktywnie substancją i tworzone są kompleksy znacznik-hERG. Następnie tak przygotowany system inkubowany jest z substancją badaną, która wypiera znacznik z kompleksów proporcjonalnie do siły jej powinowactwa do miejsca wiązania [245].

Oznaczenia z wykorzystaniem znaczników radioaktywnych pozwalają na przeprowadzenie skriningu zdolności wiązania testowanych substancji do kanału hERG. Nie dostarczają jednak żadnej informacji na temat elektrofizjologicznych efektów takiego połączenia. Także rozróżnienie pomiędzy agonistami i antagonistami nie jest możliwe. Użyteczność metody ograniczona jest także możliwością wystąpienia wyników fałszywie negatywnych, w przypadku gdy testowana substancja nie współzawodniczy ze znacznikiem o miejsce wiązania [100,245,246].

Możliwość wykrycia wszystkich typów blokerów kanału hERG daje zastosowanie kilku znaczników równocześnie. Odpowiednia w tym celu jest np. kombinacja znakowanego blokera wewnątrzkomórkowego miejsca wiązania [<sup>3</sup>H]-dofetylidu, blokera zewnątrzkomórkowego [<sup>125</sup>I] BeKm-1 oraz blokera kanału w stanie zamkniętym – ergotoksyny [104,105,234].

Oznaczenia z wykorzystaniem kompetycyjnego wiązania radioligandów nie dostarczają tak szczegółowych informacji jak pomiary elektrofizjologiczne, jednak istotną ich zaletą jest prostota, wysoka wydajność i stosunkowo niski koszt.

### 3.5.3. Pomiary fluorescencji

Metoda ta bazuje na zmianach potencjału błonowego wynikających ze zmian natężenia prądu potasowego hERG oraz barwnikach o specyficznych właściwościach. Barwnik może ulegać przemieszczeniu od zewnątrz do wnętrza komórki lub odwrotnie powodując zmianę intensywności fluorescencji lub powstanie zjawiska transferu energii rezonansu fluorescencyjnego FRET (*ang. Förster Resonance Energy Transfer lub Fluorescence Resonance Energy Transfer*) będącego wynikiem relokacji barwnika wewnątrz dwuwarstwy lipidowej błony komórkowej [20,67].

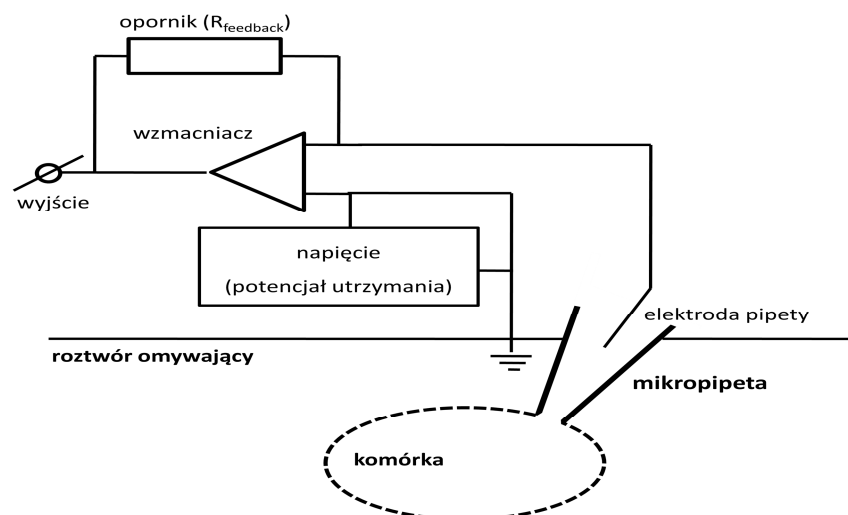
Znaczącym ograniczeniem tej metody jest wolna kinetyka redystrybucji stosowanych barwników oraz konieczność wieloetapowych przygotowań eksperymentu, co istotnie zmniejsza wydajność [20]. Nowsza metoda FLIPR (*ang. Fluorescent Imaging Plate Reader*) również wykorzystująca zjawisko fluorescencji zapewnia wyższą wydajność, może jednak dawać znaczny odsetek wyników fałszywie pozytywnych [245,341].

Stosowane barwniki potencjometryczne pozwalają obserwować w czasie rzeczywistym kinetykę zmian potencjału błonowego. Dokładność tej techniki jest jednak mniejsza niż metod patch clamp, zwłaszcza w przypadku antagonistów hERG zależnych od napięcia lub stosowania (*ang. voltage- lub usedependent*). Charakteryzuje się ona także dużą wrażliwością na artefakty doświadczalne. Pomimo wspomnianych wad, stosunkowo niski koszt oraz wysoka wydajność technik wykorzystujących pomiary fluorescencji sprawia, że są one stosowane na wczesnych etapach rozwoju leku w celu skringingu właściwości blokowania kanału potasowego hERG badanych substancji [67,245].

#### 3.5.4. Pomiary elektrofizjologiczne

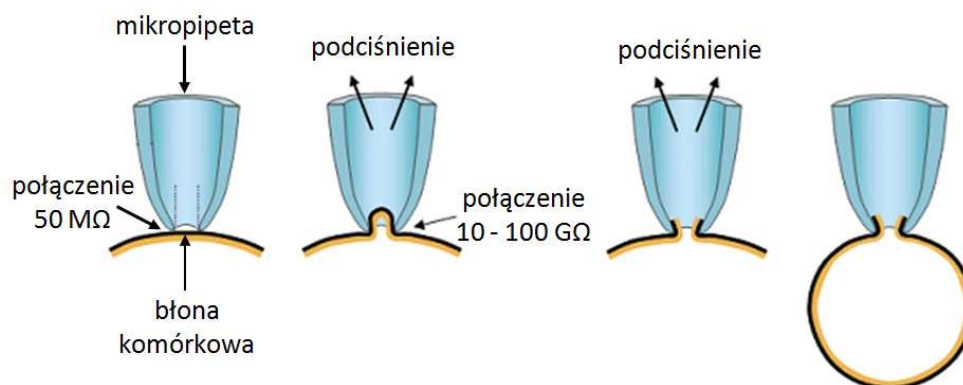
Pomiary elektrofizjologiczne (technika łątkowa, *ang. Patch Clamp*) są ogólnie przyjętym standardem badań funkcjonalnych bramkowanych potencjałów kanałów potasowych hERG i oceny wpływu potencjalnych leków i innych związków chemicznych na ten kanał. Opracowana przez Nehera i Sakmanna technika patch-clamp [244], dostarczająca szczegółowych, wysokiej jakości danych stanowi obecnie złoty standard pomiarów [387].

Metoda ta umożliwia szczegółowe prześledzenie zmian natężenia prądu potasowego w czasie z jednoczesną ścisłą kontrolą wartości napięcia (potencjału błonowego komórki). Pierwszym etapem oznaczenia jest wytworzenie ścisłego mechanicznego i elektrycznego połączenia (o oporności od  $\sim 1$  do  $\sim 10$  gigaomów) pomiędzy pipetą (z elektrodą chlorosrebrową połączoną z ujemnym wejściem odpowiedniego wzmacniacza) a błoną komórkową i utworzenie obwodu przetwornika prądowo napięciowego (Rycina 10).



**Rycina 10.** Schemat obwodu zastępczego pomiarów patch clamp w konfiguracji whole-cell. Protokół napięć ustalony przez badacza jest stale porównywany z mierzonym potencjałem, różnica jest na bieżąco kompensowana odpowiednim natężeniem prądu (current injection) [237].

Pomiary patch clamp mogą być przeprowadzane w różnych konfiguracjach. Ścisłe połączenie pipeta-błona komórkowa, będące pierwszym krokiem otrzymywania wszystkich konfiguracji, wytwarza się za pomocą delikatnego podciśnienia. Połączenie takie zapewnia izolację elektryczną wewnętrznego środowiska pipety od roztworu omywającego, co jest niezwykle istotne ze względu na minimalizację prądu upływu oraz stosunku sygnału do zakłóceń. Pozwala to na zwiększenie dokładności pomiarów oraz mierzenie prądów o natężeniu rzędu pA. Najczęściej stosowana jest konfiguracja whole-cell (Rycina 11), w której mierzona aktywność elektryczna jest sumą aktywności wszystkich kanałów znajdujących się w tzw. łacie (*ang. patch*). W odpowiednich warunkach eksperymentalnych konfiguracja ta pozwala na pomiar prądu przepływającego przez jeden rodzaj kanałów otwartych za pomocą właściwego protokołu aktywacji lub dzięki farmakologicznej blokadzie innych kanałów. W celu otrzymania konfiguracji whole-cell, po utworzeniu ścisłego połączenia zwiększa się podciśnienie tak, aby przerwana została ciągłość błony komórki.



Rycina 11. Procedura tworzenia konfiguracji whole-cell.

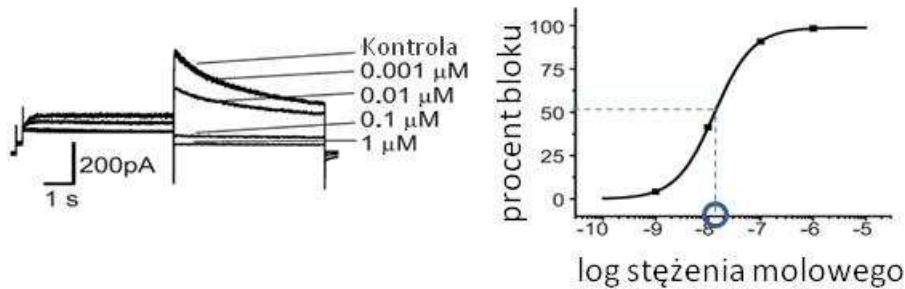
Prąd potasowy  $I_{Kr}$  wyzwalany jest za pomocą odpowiedniego protokołu napięć przykładanych do błony komórkowej. Podczas eksperymentu komórki znajdujące się w potencjale utrzymania (*ang. holding potential*) wynoszącym około  $-80$  mV, otrzymują bodziec depolaryzujący ( $0$  mV lub powyżej) co prowadzi do aktywacji kanałów i następczej szybkiej ich inaktywacji (brak przepływu prądu przez kanał). W czasie gdy błona komórkowa ulega repolaryzacji - potencjał błonowy zmniejsza się, stopniowo ustępuje inaktywacja kanałów potasowych, które następnie przez pewien czas przed zamknięciem pozostają w stanie otwartym [177,385].

Pomiary amplitudy opóźnionego dokomórkowego prądu potasowego ( $I_{Kr}$ ) wykonywane są w obecności różnych stężeń badanego związku w roztworze omywającym, kontrolę stanowi czysty roztwór. Poprzez dopasowanie współczynników równania Hilla (Równanie 1) do uzyskanych danych wyznaczana jest krzywa stężenie-odpowiedź i odczytywane jest stężenie  $IC_{50}$  (Rycina 12).

$$f = \frac{1}{1 + \left( \frac{IC_{50}}{[D]} \right)^{n_H}}$$

$f$  – procent zablokowania kanału  
 $[D]$  – stężenie leku  
 $IC_{50}$  – stężenie leku powodujące  
połowiczny blok kanału  
 $n_H$  – współczynnik Hill'a

Równanie 1.

Rycina 12. Wyznaczanie wartości  $IC_{50}$ .

Badania funkcjonalne kanałów jonowych wykonywane są z wykorzystaniem komórek wykazujących naturalną ekspresję hERG, jak np. izolowane mioocyty różnych gatunków zwierząt [32,68,117,289], mysie komórki mięśniówki przedsionków AT-1 [30], komórki neuroblastomy [8], lub z zastosowaniem heterologicznych systemów ekspresji (np. HEK, XO, CHO), do których wprowadzane jest ludzkie DNA lub mRNA kodujące białka pożądanego kanału.

Zaletą badań z wykorzystaniem izolowanych kardiomiocytów jest możliwość oceny wpływu badanej substancji na kanał potasowy w jego naturalnym środowisku (m.in. warunki fizjologiczne, biochemiczne, transdukcja sygnałów) z uwzględnieniem potencjalnego wpływu niezidentyfikowanych do tej pory podjednostek pomocniczych czy czynników wewnątrzkomórkowych. Z drugiej jednak strony badania te wiążą się z pewnymi problemami technicznymi, jak np. wstępowanie prądów zakłócających, których wpływ na wynik pomiaru należy wyeliminować [98]. Nie bez znaczenia jest również mała dostępność tkanek pochodzenia ludzkiego. Dlatego też obecnie chętniej stosowane są heterologiczne systemy ekspresji pozwalające na analizę funkcji kanałów przy braku innych protein. Do najczęściej wykorzystywanych systemów ekspresji dla kanału hERG należą embrionalne komórki nerki ludzkiej HEK 293 (*ang. human embryonic kidney*), komórki jajnika chomika chińskiego CHO (*ang. Chinese Hamster Ovary*) oraz oocyty żaby rodzaju *Xenopus* (*ang. Xenopus laevis oocytes*).

#### a. komórki XO

Dobrze znaną i udokumentowaną metodą heterologicznej ekspresji kanałów potasowych hERG są wstrzyknięcia RNA do cytoplazmy oocytów żaby *Xenopus laevis* (XO). Zastosowanie w doświadczeniach komórek XO pozwala na lepszą niż w przypadku komórek ssaków, kontrolę ekspresji kanałów, która jest proporcjonalna do ilości

wstrzykniętego mRNA, oraz dostarcza jednorodnych danych przez długi czas. Jednak wysoka lipofilność środowiska tych komórek może prowadzić do absorpcji cząsteczek badanego związku i zmniejszenia jego efektywnego stężenia w miejscu działania, a w konsekwencji do przeszacowania potencjału blokującego [293], co więcej otrzymywana wartość  $IC_{50}$  jest w znacznym stopniu uzależniona od właściwego przygotowania oocytów. Ponadto, żaby *Xenopus* są zwierzętami zmiennocieplnymi i nie możliwe jest badanie oocytów w temperaturze 37°C, co może być dodatkową przyczyną błędów w pomiarze potencjału blokowania kanału hERG przez badaną substancję.

#### b. komórki ssaków (CHO i HEK)

Tego rodzaju ograniczenia nie dotyczą komórek ssaków. Ponadto uważa się, że są one lepszym odwzorowaniem komórek ludzkich, m.in. ze względu na fakt, że proces dojrzewania białek w oocytach *Xenopus* różni się od tego w komórkach ssaków, stąd powstające ostatecznie białko może być inne od natywnego występującego u ssaków. W przeciwieństwie do komórek XO, komórki ssaków umożliwiają prowadzenie eksperymentów w temperaturze fizjologicznej, bardziej odpowiedniej w badaniach profilu bezpieczeństwa leków [410], możliwa jest także kontrola ich środowiska wewnętrznego. Kultury komórkowe mogą być transfekowane genem kodującym  $\alpha$ -podjednostkę kanału potasowego hERG stabilnie lub przejściowo. Komórki HEK wykazują bardziej stabilny poziom ekspresji białka niż komórki CHO. Z kolei, w przypadku linii komórkowych HEK poziom ekspresji może różnić się pomiędzy poszczególnymi seriami komórek, a dodatkowo w komórkach tych występuje endogenny przemijający prąd potasowy -  $I_{to}$  [410]. Prąd  $I_{to}$  zwykle uważa się za zaniedbywalny, niemniej jednak, niekiedy może on mieć istotny wpływ na wyniki pomiarów, ze względu na kluczową rolę stosunku prądu potasowego hERG do  $I_{to}$  dla farmakologii kanału hERG [324]. Z drugiej jednak strony komórki HEK, ze względu na ludzkie pochodzenie, mogą być lepszą reprezentacją warunków fizjologicznych w jakich potencjalny lek będzie stosowany. Podsumowując, zarówno linie CHO jak i HEK posiadają swoje charakterystyczne zalety i ograniczenia, dodatkowo brak ścisłych regulacji i określonych standardowych warunków pomiaru powodują, że decyzja o użyciu danej linii komórkowej uzależniona jest wyłącznie od preferencji badacza przeprowadzającego eksperyment [27].

#### 3.5.5. Metody *in silico*

Poza krótko scharakteryzowanymi powyżej badaniami *in vitro*, dobrze ugruntowane podstawy posiadają również metody łączące technologie wykorzystujące hodowle komórkowe w połączeniu z dostępnymi metodami modelowania matematycznego. Oparte na pomiarach fizjologicznych komputerowe modele kanałów jonowych, komórek, tkanek czy organów, po pozytywnej walidacji mogą wspomagać i ukierunkowywać eksperymenty z wykorzystaniem zwierząt oraz badania kliniczne, a nawet stanowić dla nich cenną i wiarygodną alternatywę. Na wczesnych etapach rozwoju

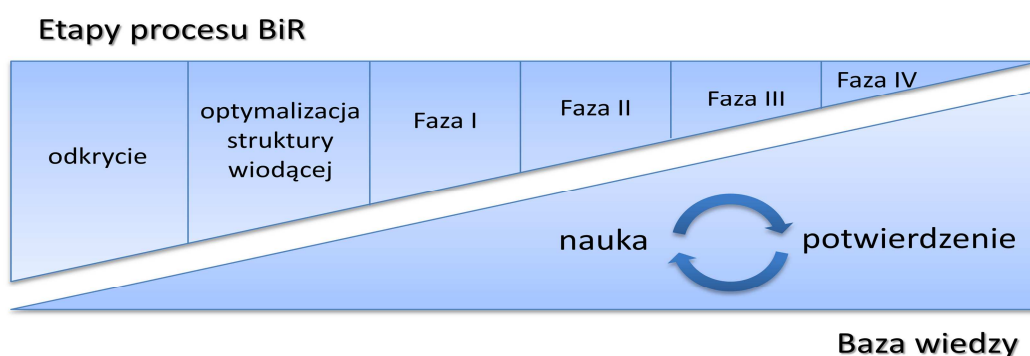
leku metody eksperymentalne *in vitro* często nie wykazują pożądanej wydajności, a ich koszty są zbyt wysokie [133]. Techniki modelowania komputerowego nie wymagają specjalnego wyposażenia, odczynników czy syntezy badanych substancji, dają możliwość testowania znacznej liczby substancji czy szerokiego zakresu stężeń w stosunkowo krótkim czasie, dlatego też metody *in silico* są atrakcyjnym i cenionym narzędziem wysokosprawnego skriningu, zwłaszcza we wczesnych etapach badań nad lekiem [295].

Współcześnie wykorzystywane metody *in silico* obejmują modelowanie homologiczne, standardowe i holograficzne techniki 2D-QSAR i 3D-QSAR [m.in. 12,91,127,172,305] oraz metody klasyfikacyjne (Tabela 26).



#### 4. MODELE KOMPUTEROWE

Każdy etap procesu rozwoju leku ma na celu dostarczenie niezbędnych informacji pozwalających ocenić prawdopodobieństwo sukcesu, czyli opracowania leku bezpiecznego i skutecznego oraz dostarczenie go pacjentom. Niezwykle istotne znaczenie dla pomyślnego zakończenia procesu BiR ma ciągłe i konsekwentne wzbogacanie bazy wiedzy dotyczącej właściwości badanej substancji. W roku 1997 Sheiner [312] przedstawił założenia projektu „Ciągle zdobywać wiedzę i potwierdzać hipotezy” (*ang. Learn and Confirm*) jako nowoczesne, racjonalne i skuteczne podejście do procesu rozwoju leku (Rycina 13). Skuteczne wdrożenie tego projektu wymaga zastosowania narzędzi modelowania i symulacji.



**Rycina 13.** Schemat podstaw projektu Learn and Confirm.

##### 4.1. Modele mechanistyczne i empiryczne

Model definiowany jest jako system aksjomatów, pojęć i ich wzajemnych zależności opisujący w sposób przybliżony i syntetyczny pewien wybrany aspekt rzeczywistości za pomocą równań matematycznych. Model stanowi więc uproszczoną reprezentację rzeczywistości ukierunkowaną na zobrazowanie wybranych, najważniejszych cech modelowanego zjawiska.

Dwie główne metody tworzenia modeli to modelowanie empiryczne (oparte na danych, statystyczne) i mechanistyczne.

Model mechanistyczny został zdefiniowany przez Sheiner i Steimer [311] jako system, oparty na rozpatrywaniu poszczególnych zjawisk i procesów, odzwierciedlający obserwowane lub zakładane mechanizmy biologiczne lub fizyczne, którego parametry znajdują interpretację w świecie rzeczywistym. W modelu empirycznym parametry wyrażonych matematycznie zależności przyczyna-efekt dopasowywane są do danych i mogą, ale nie muszą mieć odniesienia do rzeczywistości.

Modelowanie mechanistyczne wymaga elementarnej wiedzy na temat czynników i mechanizmów regulujących proces, zjawisko czy system. Model taki powinien wykazywać

zgodność z obserwowanym zachowaniem systemu - retrodykcja, a także umożliwiać przewidywanie jego zachowania w przyszłości czy pod wpływem zaburzeń – predykcja.

Budowa modelu mechanistycznego zwykle obejmuje następujące kroki:

1. zastosowanie podstawowej wiedzy na temat interakcji pomiędzy zmiennymi procesu w celu określenia struktury modelu,
2. przeprowadzenie eksperymentów w celu ustalenia parametrów modelu,
3. gromadzenie danych o procesie w celu walidacji modelu,
4. jeśli model nie wykazuje pożądanej sprawności powrót do kroku 1 i zweryfikowanie wiedzy o procesie.

Proces ten nie wymaga dużej ilości informacji, a model nie jest wrażliwy na możliwe idiosynkrazje w danych, jednak tworzenie tego rodzaju modelu może być niezwykle czasochłonne.

W przypadku gdy modelowany system jest złożony i wiedza o jego strukturze i funkcjonowaniu jest ograniczona, model musi zostać zbudowany w oparciu o pewne hipotezy dotyczące zależności przyczyna-efekt.

Budowa modelu empirycznego jest wynikiem nieco odmiennej, niż w przypadku modelu mechanistycznego, procedury:

1. zgromadzenie danych o modelowanym procesie, zjawisku, systemie,
2. określenie struktury korelacji pomiędzy zmiennymi,
3. zastosowanie odpowiedniej techniki numerycznej w celu znalezienia parametrów o maksymalnej korelacji,
4. zwalidowanie modelu z wykorzystaniem nowych danych,
5. jeśli model nie wykazuje pożądanej sprawności przejście do kroku 2.

Procedura taka powoduje, że model empiryczny powstaje w wyniku prób i błędów oraz jest w znacznym stopniu zależny od dostępności, jakości i reprezentatywności danych użytych do jego budowy i walidacji. Numerycznie modele te, w przeciwieństwie do wielu modeli mechanistycznych, są jednak znacznie prostsze w użyciu. Prawidłowo skonstruowany model może działać retroaktywnie wyjaśniając na podstawie danych przyczyny badanego zjawiska. Możliwa jest także predykcja, zarówno lokalna (interpolacja) jak i globalna, poza zakres danych wykorzystywanych do jego budowy (ekstrapolacja).

Definicja modelu i modelowania nie precyzuje technik matematycznych jakie można wykorzystać do budowy modeli. Wybór odpowiedniego narzędzia uwarunkowany jest rodzajem modelowanego problemu oraz charakterystyką posiadanej bazy danych. Od zastosowanej techniki uzależniony jest z kolei rodzaj generowanego modelu. Drzewa decyzyjne, które mogą być wyrażone w postaci zestawu reguł, metoda k-najbliższych sąsiadów (k-NN) czy regresja logistyczna, gdzie współczynniki określają relatywną

istotność danego parametru dla ostatecznej klasyfikacji pozwalają na tworzenie tzw. białych skrzynek (*ang. white-box models*), czyli modeli umożliwiających fizyczną interpretację ich parametrów. Drugą grupą modeli są tzw. czarne skrzyńki (*ang. black-box models*) tworzone z wykorzystaniem technik inteligencji obliczeniowej takich jak sztuczne sieci neuronowe (*ang. Artificial Neural Networks, ANN*) czy metody wektorów nośnych (*ang. Support Vector Machines, SVM*), które nie pozwalają na interpretację parametrów. Niemniej jednak ich efektywność jest często większa niż białych skrzynek, co może tłumaczyć ich znaczną popularność, zwłaszcza w domenach gdzie trafność predykcji jest bardziej istotna niż interpretowalność modelu.

Modele można również podzielić na opisowe i predyktywne. Modele opisowe z założenia zastosować można do ograniczonego zbioru warunków. Do ich budowy najczęściej wykorzystywane są narzędzia inteligencji obliczeniowej implementujące techniki uczenia nienadzorowanego m.in. klasteryzacja, asocjacja, ekstrakcja cech. Zadaniem modeli predyktywnych jest przewidywanie przyszłego zachowania systemu lub wyników. Najczęściej wykorzystywane do budowy tego rodzaju modeli narzędzia inteligencji obliczeniowej implementują techniki uczenia nadzorowanego m.in. klasyfikacja, regresja, selekcja cech.

#### 4.2. Proces badań i rozwoju oparty na modelach

Postęp technologiczny w zakresie cyfrowego generowania, gromadzenia, przetwarzania i analizy informacji oraz tworzenie coraz dokładniejszych modeli zjawisk i systemów dały możliwość zastosowania modelowania i symulacji w procesie badań i rozwoju leku.

Modele matematyczne charakteryzujące zależności pomiędzy właściwościami, strukturą substancji i in. a efektem ich działania, wspomagające podejmowanie decyzji stanowią podstawę opartego o modele procesu BiR (*ang. model-based drug development*). Implementacja tego procesu obejmuje gromadzenie wiedzy, przygotowanie danych i budowę bazy danych, konstrukcję i walidację modelu oraz symulację efektu z wykorzystaniem różnego rodzaju szeroko rozumianych metod obliczeniowych.

Tworzenie modeli poszczególnych elementów procesu badań i rozwoju leku ma na celu ograniczenie i ukierunkowanie eksperymentów kosztownych czy niebezpiecznych dla ludzi lub zwierząt, a także zwiększenie przepustowości testów przesiewowych czy przyspieszenie badań. Zastosowanie modeli pozwala również na łączenie informacji pochodzącej z różnych źródeł i uogólnianie wyników.

#### 4.3. Modele efektów toksycznych

Wykorzystywane m.in. przez przemysł farmaceutyczny techniki komputerowej predykcji toksyczności można ogólnie podzielić na trzy kategorie: modelowanie molekularne, systemy ekspertowe (modele mechanistyczne) oraz systemy oparte na danych (modele empiryczne).

- a. *modelowanie molekularne* – metody modelowania i przewidywania zdarzeń biochemicznych istotnych z punktu widzenia toksyczności. Komponentami modelowania molekularnego są m.in. grafika molekularna, narzędzia chemii obliczeniowej, obszerne bazy danych czy określanie właściwości fizykochemicznych na podstawie struktury (*ang. QSAR - Quantitative Structure-Activity Relationship*). Metody te mogą być wykorzystane do predykcji toksyczności, jeżeli w mechanizmie danego efektu pośredniczy receptor (np. cytochrom P450, receptor estrogenowy) i znana jest struktura białka receptorowego. Techniki te jednak nie są odpowiednie do modelowania i predykcji toksyczności uwarunkowanej złożonymi i częściowo nieznanymi mechanizmami.
- b. *systemy ekspertowe* – oparte na zasadach, ogólnych regułach wiążących aktywność biologiczną ze strukturą cząsteczki. Ich podstawą jest usystematyzowana wiedza ekspertów z dziedziny toksykologii oraz interpretacje danych toksykologicznych. Tworzenie bazy wiedzy dla systemu ekspertowego wymaga przeprowadzenia rozległego i dobrze zaplanowanego przeglądu literatury oraz „przetłumaczenia” specyficznych przypadków na generalne zasady. Dzięki temu predykcja jest możliwa nawet wtedy, gdy dostępnych jest niewiele danych eksperymentalnych (np. biotransformacja czy prognozowanie punktów końcowych związanych ze zdrowiem człowieka). Ponadto, dzięki specyficznej elastyczności działania ludzkiego mózgu, w tego rodzaju systemach możliwa jest integracja różnorodnej informacji chemicznej i biologicznej (pomiarów *in vivo* i *in vitro*, właściwości cząsteczek – doświadczalne i obliczone, fragmenty struktury i in.). Jednym z intensywniej rozwijanych działów tego obszaru wiedzy, jest konstrukcja automatycznego lub półautomatycznego systemu wnioskowania, a więc w bardzo dużym uproszczeniu – systemu naśladowującego funkcjonalnie ludzki mózg, który pozostaje wciąż niedościgłym wzorem. Celem jest skrócenie czasu oczekiwania na decyzję systemu i wzrost jego skuteczności, a testowane metody oparte są o najbardziej zaawansowane techniki modelowania matematycznego.

Niektóre z dostępnych programów wykazują stosunkowo wysoką specyficzność (nawet do 80%), jednak wszystkie tego typu programy cechuje niska czułość, zazwyczaj około 50%. Systemy ekspertowe stosowane są w badaniach nad potencjalnym lekiem i wykorzystywane do „oznakowania” substancji, co w efekcie prowadzi do zwrócenia na te cząsteczki baczniejszej uwagi podczas dalszych

etapów, nie przesądzając jednak czy przewidziane działanie w istocie wystąpi. Niewątpliwą zaletą systemów eksperckich jest transparentność predykcji. Ponieważ system taki zwykle zbudowany jest z zestawu reguł JEŻELI – TO, stosunkowo łatwo jest zidentyfikować przyczyny i uzasadnienie danej odpowiedzi systemu, co ułatwia zrozumienie i dalszą ocenę przewidzianego efektu. Taka konstrukcja pozwala na łatwe przeszukiwanie bazy wiedzy i uzyskanie potrzebnych informacji. Prostsza w porównaniu z systemami bazującymi na danych budowa modelu pozwala łatwo dodać kolejne reguły, zaktualizować wiedzę czy usunąć jej część bez konieczności przebudowy modelu. Przykładowo firma Novartis używająca komercyjnego programu DEREK, aby zwiększyć możliwość jego praktycznego zastosowania w interesujących ją obszarach, dołączyła do jego bazy danych własne informacje z obszaru, który nie był dobrze reprezentowany. Budowa systemów ekspertowych umożliwia również dzielenie się danymi, bez konieczności ujawniania i udostępniania danych poufnych. Jedną z największych wad systemów ekspertowych może być nieadekwatna, nieprecyzyjna baza wiedzy, gdzie zasady ogólne konstruowane są na podstawie kilku tylko przykładów. Dla tego rodzaju systemów niezwykle istotne jest regularne aktualizowanie wiedzy i dołączanie do niej nowych dowodów naukowych. Trudne jest także formalne zdefiniowanie tzw. domeny zastosowalności i prawidłowej walidacji systemów ekspertowych, ponieważ bardzo często nie jest jasne jakie cząsteczki posłużyły do budowy bazy wiedzy i w jakich warunkach przeprowadzono doświadczenie, którego wynik został zapisany i wykorzystany.

- c. *systemy oparte na danych (statystyczne, empiryczne)* – do budowy modeli predykcyjnych konieczne są dane doświadczalne oraz obliczone ze struktury cząsteczki. Na podstawie zestawu różnego rodzaju parametrów, za pomocą różnorodnych metod statystycznych systemy te uczą się i odkrywają matematyczne zależności pomiędzy niejednorodnymi elementami zbioru uczącego. Typowym przykładem takiej procedury są zależności SAR/QSAR. Klasyczna analiza QSAR stosuje techniki regresji pozwalające opracować równanie na podstawie danych, które może być wykorzystane do przewidywań dla cząsteczek o podobnej strukturze (np. kongenerów) i mechanizmie działania. Szczególnie istotny dla poprawnego działania systemów statystycznych jest odpowiedni dobór właściwości chemicznych istotnych z punktu widzenia badanego efektu toksycznego. Wskazanie tych parametrów a priori oraz zagwarantowanie, że żadne istotne cechy nie zostały pominięte jest niezwykle trudne. Modele statystyczne są znacznie częściej niż systemy ekspertowe wykorzystywane do predykcji właściwości cząsteczek. Ponieważ doświadczenia badające toksyczność są zwykle kosztowne i czasochłonne, techniki drążenia danych zyskały sobie w ostatnich latach znaczną popularność, należą do nich m.in.: metoda cząstkowych najmniejszych kwadratów (PLS), regresja wieloczynnikowa,

metoda wektorów nośnych (SVM), sztuczne sieci neuronowe, metody rekursywnego podziału (drzewa klasyfikacyjne i regresyjne), metoda k-najbliższych sąsiadów, analiza dyskryminacyjna, techniki oparte na logice bayesowskiej i inne. Oprócz prawidłowej predykcji określonych efektów istotnym zadaniem systemów statystycznych jest identyfikacja cech chemicznych odpowiedzialnych za wystąpienie obserwowanego działania, co może umożliwić chemikom optymalizację cząsteczki. Należy zaznaczyć, że modele matematyczne przewidują jedynie korelację pomiędzy strukturą czy właściwościami cząsteczki, a określonym efektem biologicznym, jednak odkryta korelacja nie jest jednoznaczna ze związkiem przyczynowo-skutkowym.

#### 4.4. Jakość modelu

Każdy model powinien spełniać określone kryteria merytoryczne i jakościowe. Opracowano wiele wytycznych dotyczących dobrej praktyki modelowania w medycynie i farmacji [264,320,402]. Jednak nie ma ustalonego standardu pozwalającego określić co stanowi "dobry model" i w jaki sposób powinna przebiegać formalna ocena jego jakości. Wśród niektórych czynników wymienianych jako istotne dla jakości modelu znajduje się m.in. jakość danych wykorzystywanych do jego budowy, zgodność struktury modelu z rozwiązywanym problemem czy zgodność wewnętrzną i zewnętrzną, odpowiednio przeprowadzona weryfikacja. Z punktu widzenia predykcyjnych modeli QSAR najistotniejsza wydaje się być efektywność, skuteczność. Jednak i w tym przypadku brak jest określonych standardowych miar dopasowania, schematów testowania czy innych kryteriów oceny jakości modelu, a dobór odpowiednich wskaźników pozostaje w gestii osoby tworzącej model.

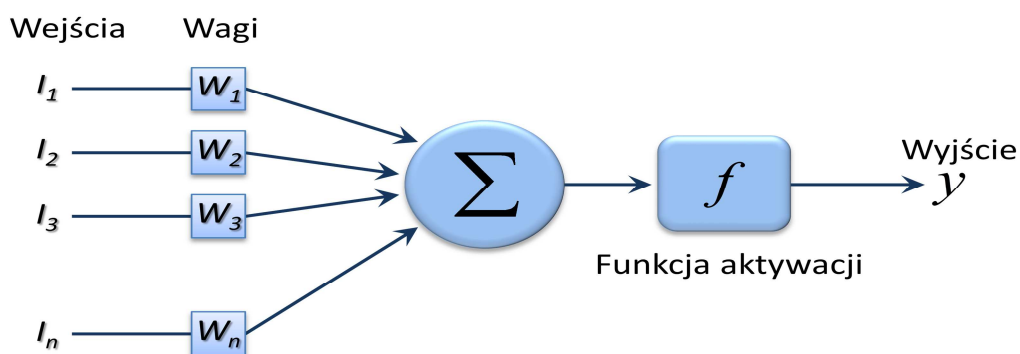
## 5. TECHNIKI MODELOWANIA

### 5.1. Sztuczne sieci neuronowe

Sztuczne sieci neuronowe (*ang. Artificial Neural Network, ANN*) to matematyczne modele i algorytmy, programy komputerowe zaprojektowane by symulować sposób, w jaki ludzki mózg przetwarza informacje [2]. Jest to oczywiście model bardzo uproszczony. Sieć zbudowana jest z wielu, zorganizowanych w warstwy podjednostek, tzw. sztucznych neuronów lub elementów przetwarzających połączonych ze sobą wagami. Sieć gromadzi wiedzę poprzez wykrywanie wzorców oraz związków pomiędzy danymi i uczy się przez doświadczenie, a nie programowanie.

#### 5.1.1. Sztuczny neuron

Pierwszą próbę opracowania matematycznego modelu neuronu podjęli w 1943 roku McCulloch i Pitts [28,335]. Zbudowany przez nich neuron (Rycina 14, Równanie 2), podobnie jak biologiczny posiadał wiele wejść oraz jedno wyjście. W sieciach wielowarstwowych wartości docierających do neuronu sygnałów wejściowych mogą być zarówno danymi pierwotnymi podawanymi przez badacza jak i sygnałami wyjściowymi innej, wcześniejszej komórki sieci. Sygnał wejściowy po przemnożeniu przez tzw. współczynniki wagowe, zsumowaniu i przetworzeniu przez funkcję aktywacji generuje sygnał wyjściowy. W przypadku neuronu McCullocha i Pittsa funkcja aktywacji jest funkcją progową, co oznacza, że działa on zgodnie z zasadą „wszystko albo nic” - na wyjściu wartość 0 (brak pobudzenia) lub 1 (pobudzenie) w zależności czy suma jest mniejsza czy większa od wartości progowej.



Rycina 14. Model neuronu McCulloch'a i Pitts'a.

$$y = f(x_1 \cdot w_1 + x_2 \cdot w_2 + \dots + x_n \cdot w_n) = \sum_{j=1}^n x_j w_j \quad \text{Równanie 2.}$$

$$y = \begin{cases} 1 & \text{dla } \sum_{i=1}^n x_i w_i \geq f \\ 0 & \text{dla } \sum_{i=1}^n x_i w_i < f \end{cases} \quad \begin{matrix} x \in \{0,1\} \\ w_i \in \{-1,1\} \\ i = 1,2,3,\dots,n \end{matrix}$$

Funkcja progowa jest aktualnie rzadko stosowana ze względu na trudności pojawiające się podczas uczenia tego typu neuronów [77]. Obecnie funkcje aktywacji używane w sieciach neuronowych mają charakter ciągły, a sygnał wyjściowy zmienia się proporcjonalnie do wejścia. Wśród nich wyróżnić można dwa typy neuronów - liniowe i nieliniowe.

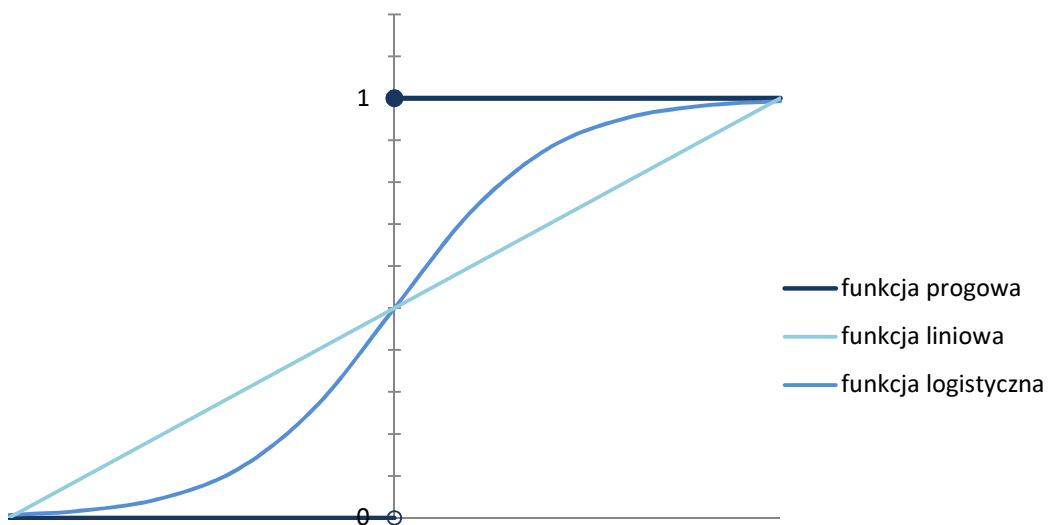
Działanie neuronów nieliniowych jest bliższe funkcjonowaniu jednostek biologicznych. Najbardziej zbliżony do naturalnego model można uzyskać przez zastosowanie funkcji sigmoidalnej (logistycznej). Przykładem najczęściej wykorzystywanych funkcji logistycznych jest sigmoida unipolarna (Rycina 15, Równanie 3) oraz tangens hiperboliczny (Rycina 15, Równanie 4)

$$y = \frac{1}{1 + e^{-\lambda x}}; y \in \{0,1\}; \lambda > 0$$

Równanie 3.

$$y = \tanh\left(\frac{\lambda x}{2}\right) = \frac{e^x - 1}{e^x + 1} = \frac{1 - e^{-\lambda x}}{1 + e^{-\lambda x}}; \lambda > 0$$

Równanie 4.



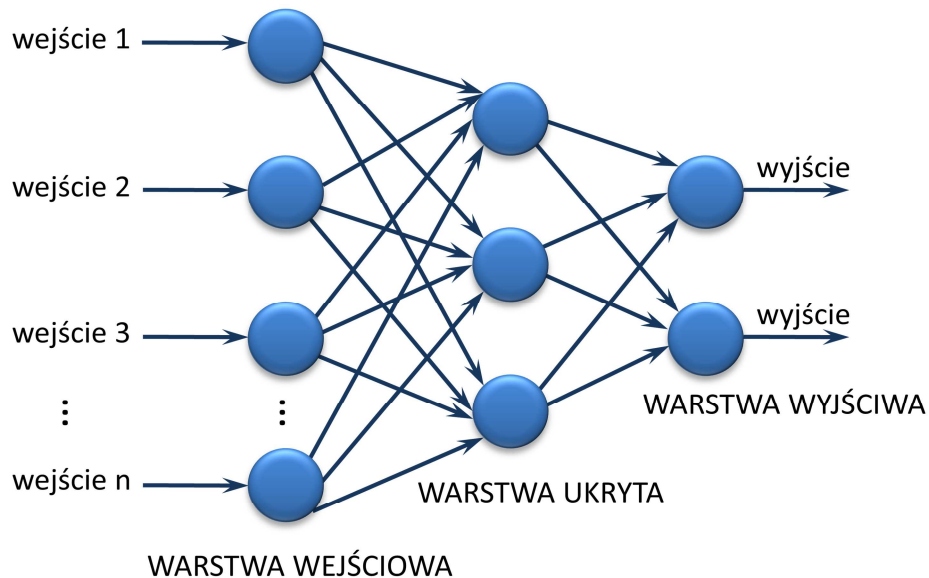
Rycina 15. Funkcje aktywacji.

### 5.1.2. Sieć neuronowa

Sztuczny neuron podobnie jak jego biologiczny pierwowzór posiada zdolność uczenia się, jednak jako pojedynczy element ma znacznie ograniczone możliwości. Dlatego też tworzone są bardziej złożone struktury zbudowane z większej liczby komórek odpowiednio ze sobą połączonych, co znacznie zwiększa możliwości obliczeniowe. Sieci neuronowe strukturalnie zbudowane są z warstw powstających przez połączenie wyjść jednych neuronów z wejściami innych (Rycina 16). System taki zdolny jest do równoległego przetwarzania informacji – co stanowi jedną z najcenniejszych właściwości



sieci neuronowej. Warstwa pierwsza (wejściowa) nie uczestniczy w przetwarzaniu sygnału, jest ona reprezentacją danych wejściowych (wektora wejściowego). Ostatnia warstwa obrazuje stan układu (wektor wyjściowy). W przypadku sieci wielowarstwowej pomiędzy warstwą wejściową a wyjściową znajdują się tzw. warstwy ukryte. Zazwyczaj neurony sieci połączone są na zasadzie każdy z każdym.



Rycina 16. Schemat architektury dwuwarstwowej.

Architektura sieci neuronowej charakteryzowana jest przez:

- liczbę wejść i wyjść,
- liczbę warstw,
- liczbę neuronów w poszczególnych warstwach,
- liczbę wag w każdym neuronie,
- sposób połączenia pomiędzy lub wewnątrz warstw,
- określenie, który neuron otrzymuje sygnał korygujący,
- sposób modyfikacji wag (algorytm uczenia),
- funkcję aktywacji.

Sieci, w których tak jak na przedstawionym schemacie (Rycina 16), sygnał propagowany jest od wejścia do wyjścia nazywane są jednokierunkowymi (*ang. feedforward*). Konstruuje się także sieci, w których informacja może być przekazywana w dwóch kierunkach, a w warstwach ukrytych występują zamknięte pętle (sprężenia zwrotne), w których sygnały mogą krążyć i ulegać zmianom do momentu uzyskania przez sieć określonego stanu. Są to tzw. sieci rekurencyjne (*ang. feedback*), najczęściej wykorzystywane jako pamięć asocjacyjna, m.in. BAM (*ang. Bidirectional Associative Memory*), maszyna Boltzmanna czy sieci Hopfielda [77].

### 5.1.3. Proces uczenia ANN

Uczenie sztucznej sieci neuronowej to proces adaptacyjnych zmian parametrów sieci odpowiednio do specyfiki rozwiązywanego problemu. Metody uczenia sieci podzielić można na dwa podstawowe typy [28,253,335]:

- uczenie nadzorowane (z nauczycielem, *ang. supervised*)
- uczenie bez nadzoru (*ang. unsupervised*)

W niniejszej pracy wykorzystano wyłącznie algorytmy uczenia nadzorowanego sieci neuronowych, dlatego też tylko one zostaną dokładniej omówione.

Zasadniczym celem nadzorowanych metod trenowania jest nauczenie sieci przewidywania wartości wyjścia lub wyjść na podstawie wartości wejściowych. Uczenie z nauczycielem polega na wielokrotnym, iteracyjnym przedstawianiu sieci danych w postaci par wektor wejściowy - oczekiwane wyjście. W procesie tym zmieniane są współczynniki wagowe połączeń neuronów aż do momentu, w którym wartość błędu osiągnie pewną założoną wartość. Podstawowym i najczęściej stosowanym algorytmem korekty wag jest metoda wstecznej propagacji błędu (BP – *ang. Back Propagation*) i reguła delta [77,192,335].

W trakcie uczenia sieci z zastosowaniem reguły delta uaktualnianie wartości wag może następować przyrostowo (po prezentacji każdej pary uczącej) lub w sposób skumulowany (po prezentacji całego zbioru treningowego).

Przy prezentacji par danych błąd ( $\delta$ ) jest definiowany jako różnica pomiędzy odpowiedzią oczekiwaną a bieżącą odpowiedzią sieci na dany sygnał wejściowy.

$$\delta^{(j)} = z^{(j)} - y^{(j)} \quad \text{Równanie 5.}$$

gdzie:

$z^{(j)}$  - oczekiwana odpowiedź sieci na sygnał w iteracji j

$y^{(j)}$  - rzeczywista (bieżąca) odpowiedź sieci na sygnał w iteracji j

Do określenia wartości całkowitego błędu popełnianego przez sieć najczęściej wykorzystuje się funkcję kwadratową (Równanie 6) [242].

$$Q = \frac{1}{2} \sum_{j=1}^j (z^{(j)} - y^{(j)})^2 \quad \text{Równanie 6.}$$

gdzie:

j – liczba par uczących

$z^{(j)}$  - oczekiwana odpowiedź sieci na sygnał w iteracji j

$y^{(j)}$  - rzeczywista (bieżąca) odpowiedź sieci na sygnał w iteracji j

Minimalizacja różnic pomiędzy  $z^{(j)}$  a  $y^{(j)}$  odpowiada minimalizacji funkcji błędu Q (celu).

Proces iteracyjnego uczenia sieci metodą delty, polega więc na powtarzaniu schematu:

1. przedstawienie ciągu uczącego,
2. obliczenie błędu  $\delta$ ,
3. korekta wag neuronów proporcjonalnie do iloczynu wektora błędu i sygnału wejściowego:

$$w_i^{j+1} = w_i^j + \Delta w_i^j \quad \text{Równanie 7.}$$

$$\Delta w_i^j = \eta \delta^j x_i^j \quad \text{Równanie 8.}$$

gdzie:

$w_i^j, w_i^{j+1}$  - waga i-tego neuronu w iteracji j i j+1 (odpowiednio)

$\Delta w_i^j$  - zmiana wagi i-tego neuronu w iteracji j

$\eta$  - współczynnik decydujący o szybkości uczenia

$\delta^j$  - wartość błędu w iteracji j

$x_i^j$  - sygnał wejściowy wagi i-tego neuronu w iteracji j

aż do osiągnięcia zakładanego minimum błędu lub maksymalnej dopuszczalnej przez badacza liczby iteracji.

Realizacja algorytmu wstecznej propagacji błędu polega na „rzutowaniu” błędu  $\delta$  danego neuronu na wszystkie neurony, które stanowiły jego wejście. Błąd przekazywany jest przez neurony w kierunku odwrotnym do propagacji sygnałów uczących – od wyjścia poprzez warstwy ukryte do wejścia [242,335].

Rezultatem zastosowania algorytmu wstecznej propagacji błędu jest duża efektywność procesu uczenia i zmniejsza złożoność obliczeniowa, a ponadto pojawia się możliwość przetwarzania równoległego czy rozproszenia procesu uczenia sieci. Z drugiej jednak strony, iteracyjny charakter algorytmu wiąże się z czasochłonnością uczenia. Ponadto ze względu na wielomodalność funkcji błędu możliwe jest, że proces uczenia zakończy się w jednym z wielu jej minimów lokalnych, dla których wartość sygnału wyjściowego różni się znacznie od wartości sygnału wzorcowego, co stanowi jedną z głównych wad algorytmu wstecznej propagacji błędu. Powyższy problem może być związany z nieprawidłowym wyborem początkowych wartości wag. Zalecany jest dobór losowy współczynników wagowych, a także kilkakrotne powtarzanie procesu uczenia rozpoczynające się od różnych ich wartości. Korzystne jest również podawanie wektorów wzorcowych w losowej kolejności. Zabiegi te tylko w pewnym stopniu zapobiegają problemom pojawiającym się w trakcie uczenia sieci.

Kolejną istotną niedogodnością związaną z algorytmem wstecznej propagacji błędu jest brak możliwości matematycznego wyrażenia zbudowanego modelu, stąd też określenie tego typu modeli jako „czarnych skrzynek” (*ang. black-box*). W zastosowaniach praktycznych sztucznych sieci neuronowych szczególnie istotną rolę odgrywa szybkość uczenia, dlatego też opracowanych zostało wiele modyfikacji algorytmu wstecznej propagacji błędu (m.in. metoda zmiennego współczynnika  $\eta$ , metoda momentu) przyspieszających zbieżność procesu uczenia oraz pomagających uniknąć minimów lokalnych [242].

a. metoda zmiennego współczynnika  $\eta$

Występujący w równaniu 8 współczynnik  $\eta$  decyduje o wielkości zmian wag warunkowanych określoną wartością błędu sieci i wartością sygnałów wejściowych, a więc może w sposób istotny wpływać na szybkość uczenia się sieci. Za mała jego wartość powoduje, że spadek gradientu funkcji błędu jest zbyt wolny, natomiast wartość zbyt duża - gwałtowne zmiany parametrów sieci i brak stabilności w trakcie procesu uczenia [77,335].

Opracowano różne strategie zmian współczynnika  $\eta$  uzależniających jego wartość od wartości funkcji błędu, przy czym zmiany te mogą być wprowadzane przy każdej iteracji lub dopiero po prezentacji wszystkich zestawów uczących w epoce [242]. Przyspieszenie zbieżności procesu uczenia uzyskać można przez zastosowanie reguły delta-delta Jacobsa lub jego zmodyfikowanej wersji – algorytmu delta-bar-delta, w którym współczynnik uczenia uzależniony jest nie od aktualnej wartości gradientu, ale od pewnej trwalszej tendencji w jego zmianach. Modyfikacja ta zmniejsza wrażliwość algorytmu na nagłe zmiany gradientu funkcji błędu, przyspiesza proces uczenia, a także zwiększa prawdopodobieństwo osiągnięcia minimum globalnego.

b. metoda momentu  $\alpha$

Inną techniką pozwalającą na przyspieszenie procesu uczenia i zapobieganie fluktuacjom jest metoda momentu [335]. Polega ona na wprowadzeniu do wzoru na zmianę wag dodatkowego członu uwzględniającego tzw. „bezwładność” procesu tych zmian (Równanie 9).

$$w_i^{j+1} = w_i^j + \Delta w_i^j + \alpha M^j \quad \text{Równanie 9.}$$

$$M^j = w_i^j - w_i^{j-1}$$

gdzie:

$\alpha$  – współczynnik momentu

$w_i^{j-1}, w_i^j, w_i^{j+1}$  - waga i-tego neuronu w iteracji j-1, j oraz j+1 (odpowiednio)

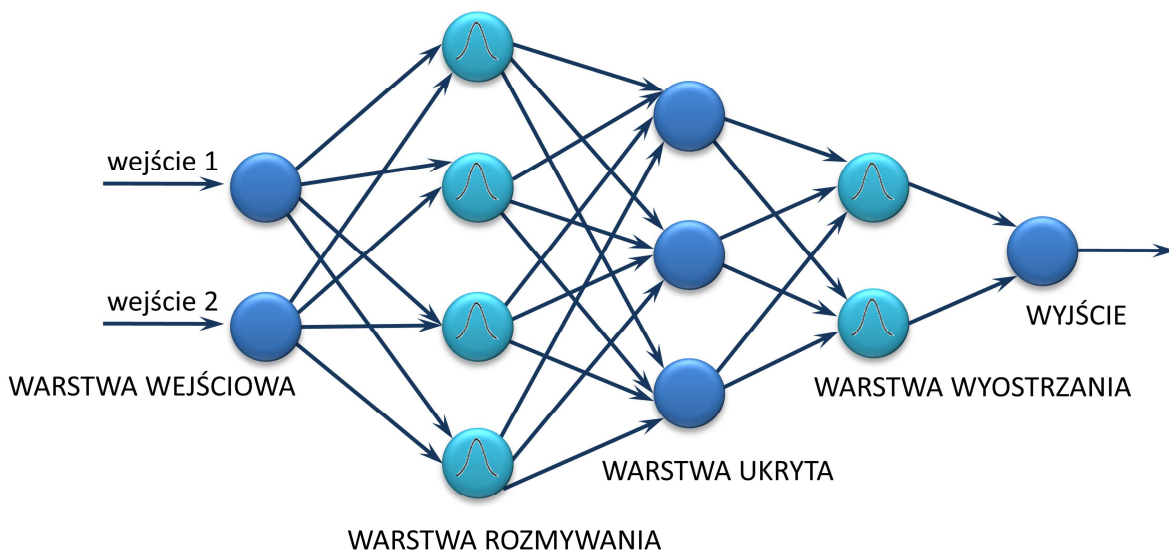
$\Delta w_i^j$  - zmiana wagi i-tego neuronu w iteracji j

Modyfikacja ta pozwala na uzależnienie korekty współczynników wagowych nie tylko od wartości bieżącego błędu sieci, ale także od wcześniejszego przebiegu procesu uczenia oraz zwiększa prawdopodobieństwo opuszczenia minimum lokalnego przez funkcję błędu. Wartość współczynnika  $\alpha$ , podobnie jak w przypadku  $\eta$ , może być zmieniana w trakcie uczenia sieci.

Inne przykłady algorytmów ze zmiennym współczynnikiem uczenia i współczynnikiem momentu to m.in. algorytm Parka, Yuna i Kima, Lee i Parka, Quickprop czy RPROP.

#### 5.1.4. Systemy hybrydowe

Systemy hybrydowe charakteryzują się łączeniem różnych metod sztucznej inteligencji w celu tworzenia nowej jakości oraz eliminacji słabych stron technik stosowanych w formie klasycznej. Przykładem takiego systemu są m.in. sieci neuronowo-rozmyte (*ang. Neuro-fuzzy network*), które pozwalają uniknąć dwóch najczęściej zarzucanych klasycznym sieciom neuronowym ograniczeń, a mianowicie niejawności postaci modelu oraz niemożności operowania danymi symbolicznymi. Połączenie klasycznej sieci neuronowej i technik bazujących na logice rozmytej daje w efekcie system przetwarzający zarówno dane numeryczne jak i symboliczne, posiadający bazę wiedzy w postaci reguł rozmytych łatwych do interpretacji przez człowieka oraz przejętą od ANN zdolność uczenia się (Rycina 17).

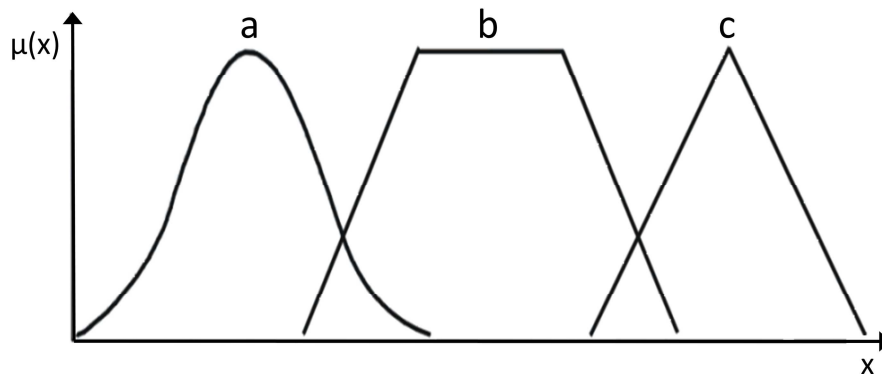


Rycina 17. Architektura sieci neuronowo-rozmytej.

Logika rozmyta (wielowartościowa), będąca rozwinięciem klasycznej teorii zbiorów Kanta, umożliwia matematyczny opis pojęć i zjawisk o charakterze niejednoznaczny. W teorii zbiorów rozmytych dany element może należeć do zbioru częściowo, a stopień jego przynależności do konkretnego zbioru rozmytego określa funkcja przynależności ( $\mu$ ), przypisująca mu wartość z przedziału  $(0;1)$ .

- $\mu_A(x) = 1$  - pełna przynależność elementu  $x$  do zbioru rozmytego  $A$   
 $\mu_A(x) = 0$  - brak przynależności  
 $0 < \mu_A(x) < 1$  - przynależność częściowa

Funkcja przynależności opisująca zbiór rozmyty najczęściej przyjmuje kształt trójkąta, trapezu lub dzwonowaty (krzywa Gaussa) (Rycina 18).



Rycina 18. Kształty funkcji przynależności: a) krzywa Gaussa, b) funkcja trapezoidalna, c) funkcja trójkątna.

Wartości wejściowe podawane sieci mogą przyjmować postać zmiennych lingwistycznych (terminów podstawowych), terminów podstawowych z modyfikatorami (natężenia, kontrastu, ograniczenia) lub formę numeryczną. W ostatnim przypadku dane poddawane są procesowi rozmywania (fuzyfikacji). Metoda Singleton jest najczęściej stosowaną techniką rozmywania. W sieci rozmytej w postaci liczb rozmytych mogą być również wagi sieci lub funkcje aktywacji. Uczeniu podlegają zarówno wagi, wejścia jak i parametry funkcji przynależności (centrum, szerokość). Sposób połączenia informacji lingwistycznej i numerycznej w sieci może być różny, najczęściej jednak zmienne w postaci symbolicznej stosowane są na etapie wstępnego doboru parametrów, a do właściwego procesu uczenia stosuje się zmienne numeryczne. Wynik działania sieci rozmytej może być przedstawiony w postaci zmiennych lingwistycznych wraz z przypisaną do nich wartością funkcji przynależności. Jeśli konieczny jest bardziej precyzyjny wynik, w postaci pojedynczej wartości, to rozmytą odpowiedź systemu poddaje się procesowi defuzyfikacji (wyostrzania), w którym defuzyfikator dokonuje przekształceń zbioru rozmytego w ściśle określony punkt rozwiązania [242,253]. Stosowane metody wyostrzania obejmują CA (*ang. Center-Average*), CS (*ang. Center-of-Sums*), COA (*ang. Center-of-Area*), SLIDE (*ang. Semi-Linear Defuzyfikation Method*) czy M-SLIDE (*ang. Modified Semi-Linear Defuzyfikation Method*) [286].

#### 5.1.5. Właściwości i zastosowania

Jedną z najcenniejszych właściwości sztucznych sieci neuronowych jest zdolność równoległego oraz rozproszonego przetwarzania danych w przeciwieństwie do szeregowego działania tradycyjnego procesora. Dużym atutem sieci jest także zastąpienie klasycznego procesu programowania przez uczenie, zdobywanie wiedzy poprzez doświadczenie. Do podstawowych cech ANN mających odzwierciedlenie w biologicznym układzie nerwowym należą również:

- zdolność generalizacji,
- predykcja i interpolacja,
- mała wrażliwość na błędy w danych,
- zdolność do efektywnej pracy nawet po częściowym uszkodzeniu sieci (usunięciu neuronów lub połączeń).

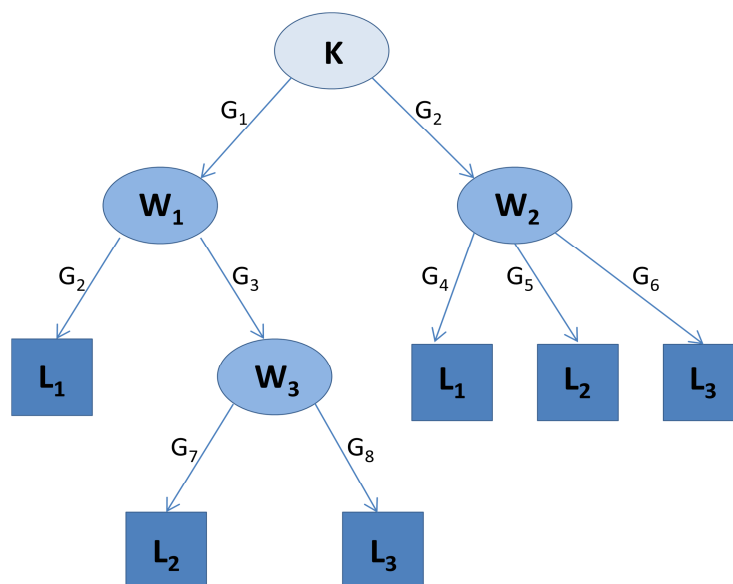
Dzięki tym właściwościom sztuczne sieci neuronowe mogą być zastosowane wszędzie tam, gdzie wykorzystanie tradycyjnych metod napotyka duże trudności, rozwiązania analityczne są niemożliwe lub bardzo trudne do realizacji, w zagadnieniach wymagających kojarzenia i przetwarzania niepełnej lub niedokładnej informacji. Adaptacyjna natura sieci powoduje, że takie systemy obliczeniowe znajdują szczególne zastosowanie w obszarach nauki, gdzie dostępne są dane na temat pewnego zjawiska, ale jego przebieg jest nieznany lub słabo zrozumiany. Są one nowoczesnym i atrakcyjnym narzędziem przetwarzania danych, które w swych założeniach odbiega od klasycznych systemów przetwarzania informacji. ANN stanowią użyteczne narzędzie w rozwiązywaniu takich problemów jak klasyfikacja obiektów i sygnałów, rozpoznawanie i wykrywanie cech, optymalizacja, asocjacja czy aproksymacja. Sieci neuronowe są wygodnym i efektywnym narzędziem, przydatnym w realizacji bardzo wielu różnych praktycznych zadań. Są z powodzeniem stosowane w niezwykle szerokim zakresie problemów, takich jak ekonomia, medycyna, farmacja, inżynieria, geologia czy fizyka [242,335].

## 5.2. Drzewa decyzyjne

Drzewa decyzyjne są jedną z form analizy wieloczynnikowej, graficzną metodą wspomaganą procesu decyzyjnego. Jest to jedna z częściej stosowanych technik analizy danych, która w uczeniu maszynowym wykorzystywana jest do wyodrębniania wiedzy ze zbioru przykładów. Oferują unikalną możliwość uzupełnienia, zastąpienia czy wspomaganie tradycyjnych form analizy statystycznej, narzędzi drążenia danych (*ang. data mining*) czy opracowanych w ostatnim czasie wielowymiarowych form sprawozdawczości i analiz funkcjonujących w obszarze tzw. inteligencji, analizy biznesowej (*ang. business intelligence*). Drzewa decyzyjne w eksploracji danych mogą być wykorzystane do zadań takich jak:

- opis – zmniejszenie wymiarowości, uproszczenie informacji zawartej w danych z zachowaniem jej podstawowych cech,
- klasyfikacja – odkrywanie w danych dobrze separowalnych klas obiektów (tzw. klas abstrakcji),
- generalizacja – mapowanie zmiennych niezależnych na zależne, umożliwiające przewidywanie wartości zmiennej zależnej w przyszłości.

Drzewa działają w oparciu o zasadę „dziel i rządź”. Polega ona na podziale złożonego problemu na prostsze, a następnie rekursywnym zastosowaniu tej strategii do stworzonych podzbiorów. Każda reguła (test) przypisuje rekordy, obserwacje zbioru danych do odpowiedniej gałęzi na podstawie ich wartości. Efektem kolejno stosowanych reguł podziału jest utworzenie hierarchicznej struktury o charakterystycznym kształcie odwróconego drzewa (Rycina 19).



**Rycina 19.** Schemat drzewa decyzyjnego. K – korzeń (parametr  $n_1$ ),  $W_1$ - $W_3$  – węzły (testy przeprowadzane na kolejnych parametrach  $n_2$ - $n_i$ ),  $L_1$ - $L_3$  – liście (etykieta kategorii/decyzja),  $G_1$ - $G_8$  – gałęzie (wartość parametru-możliwe wyniki testu).



Formalnie drzewo decyzyjne to spójny skierowany acykliczny graf. Krawędzie takiego grafu są nazywane gałęziami, wierzchołki, z których wychodzi co najmniej jedna krawędź - węzłami, a pozostałe wierzchołki – liśćmi. Korzeń drzewa decyzyjnego stanowi wybrany parametr (atrybut), natomiast gałęzie reprezentują wartości tego parametru. Korzeń połączony jest z liśćmi drzewa poprzez szereg węzłów odpowiadających kolejno sprawdzanym warunkom i gałęzi odpowiadających wynikom tych testów. Drzewo decyzyjne reprezentuje więc sekwencję warunków wpływających na ostateczną decyzję. Reguły przydziału we wszystkich węzłach są wyczerpujące i wzajemnie wykluczające się, dlatego każdy przykład ze zbioru danych może znaleźć się w dokładnie jednym liściu. Po ustaleniu ostatecznej struktury drzewa i określeniu wszystkich reguł podziału, możliwe jest jego wykorzystanie do przewidywania efektu (liści) dla nowych, nieznanych przykładów.

### 5.2.1. Zstępujące konstruowanie drzewa

Konstruowanie drzew decyzyjnych następuje poprzez rekurencyjny podział zbioru rekordów do momentu, gdy powstający po podziale podzbiór osiągnie założony stopień jednorodności. Pierwszym krokiem konstrukcji drzewa jest wybranie atrybutu, który będzie korzeniem drzewa, a następnie decyzja czy rozpatrywany węzeł ma zostać etykietą – liściem drzewa zgodnie z kryterium stopu czy węzłem rozgałęziającym. W ostatnim przypadku z puli dostępnych atrybutów wybierany jest ten, którego wartość będzie podstawą rozgałęzienia i tworzenia kolejnych węzłów. W kolejnych węzłach opisany cykl powtarzany jest rekurencyjnie, z tym że w dalszych węzłach brane pod uwagę są wyłącznie przykłady, które spełniają warunki, aby dotrzeć do określonego węzła drzewa a zbiór dostępnych atrybutów jest pomniejszony o uprzednio wykorzystane [44].

Z określonego zbioru danych skonstruować można wiele drzew decyzyjnych, ich liczba rośnie wykładniczo w zależności od liczebności zbioru uczącego. Każde drzewo charakteryzuje określona dokładność, precyzja czy trafność przewidywania. Znajdzenie optymalnej konstrukcji ze względu na rozmiar przestrzeni wyszukiwania jest trudne obliczeniowo (w praktyce niewykonalne dla większych zbiorów). Niemniej jednak opracowane zostały wydajne algorytmy indukcji wystarczająco dobrych, choć nie optymalnych, drzew decyzyjnych generujących rozwiązanie w rozsądnym czasie. Algorytmy te zwykle wykorzystują strategię zachłanną budowy drzewa, polegającą na podejmowaniu serii lokalnie optymalnych decyzji o wyborze atrybutu, według którego należy podzielić zbiór. W roku 1966 Hunt [142] przedstawił pierwszy algorytm indukcji drzewa decyzyjnego - algorytm CLS (*ang. Concept Learning System*). Przy założeniu, że  $D_t = \{(\bar{x}, y) : \bar{x} \in X, y \in Y\}$  jest zbiorem rekordów uczących związanych z węzłem  $t$ , opisanych wektorem atrybutów  $\bar{x} = (a_1, a_2, \dots, a_m)$  i etykietą klasy  $y$ , rekurencyjna definicja algorytmu Hunta wygląda następująco:

1. *jeżeli* wszystkie rekordy z  $D_t$  należą do jednej klasy  $y_t$ , to  $t$  jest węzłem terminalnym (liściem) o etykiecie  $y_t$ ; należy zakończyć pracę
2. *jeżeli* rekordy z  $D_t$  należą do więcej niż jednej klasy, należy wybrać jeden atrybut  $a_i$ , utworzyć węzeł decyzyjny i podzielić zbiór uczący  $D_t$  na podzbiory  $D_{t1}, D_{t2}, \dots, D_{tm}$  odpowiadające wartościom atrybutu  $a_i$
3. zastosować rekurencyjnie kroki 1 – 3 do utworzonych podzbiorów  $D_{t1}, D_{t2}, \dots, D_{tm}$

Algorytm Hunta [142] jest podstawą wielu innych stosowanych algorytmów indukcji drzew decyzyjnych, w tym ID3, C4.5 czy CART.

### 5.2.2. Kryterium wyboru testu

Zasadnicze znaczenie dla wydajności drzewa, w znaczeniu dokładności, wiarygodności, rozmiarów i możliwości interpretacji ma wybór atrybutów do korzenia i kolejnych węzłów. Atrybuty powinny być wybierane według malejącej ich istotności, przy czym test otrzymuje tym wyższą ocenę im mniej zróżnicowane przykłady znajdują się w podziorach powstających po jego zastosowaniu. Opracowano wiele metod służących do oceny użyteczności atrybutów. Zgodnie z klasyfikacją zaproponowaną przez Ben-Bassat [21] metody te można podzielić na:

- *bazujące na teorii informacji* - mierzące różnicę między zbiorem przykładów w węźle  $t$  a zbiorami przykładów w węzłach potomnych ze względu na rozkład częstości klas,
- *bazujące na miarach odległości* - mierzące różnice między poszczególnymi zbiorami przykładów w węzłach potomnych ze względu na rozkład częstości klas,
- *bazujące na ocenie statystycznej niezależności* - mierzące statystyczną niezależność między rozkładem klas a podziałem zbioru przykładów na podzbiory.

Metody oceny użyteczności atrybutów oparte o pojęcia teorii informacji wywodzą się z koncepcji entropii Shannona [44]. Entropia rozkładu prawdopodobieństwa decyzji wynikających z przyjęcia danego atrybutu jako węzła rozgałęziającego wyrażana jest w postaci zależności:

$$E(D_{tm}) = -\sum_{i=1}^n p_i \log_2(p_i) \quad \text{Równanie 10.}$$

gdzie:

$D_{tm}$  - podzbiór będący wynikiem zastosowania danego atrybutu  $a_k$  jako testu

$p_i$  - prawdopodobieństwo wystąpienia elementu z  $i$ -tej klasy (w przypadku nieznanomości prawdopodobieństw mogą być zastąpione liczebnościami zdarzeń) część rekordów należących do klasy  $i$

$n$  - ilość klas

Lokalną minimalizację entropii można również rozumieć jako maksymalizację przyrostu informacji (kryterium korzyści) [44].

$$Gain(D_{tm}, a_k) \equiv E(D_{tm}) - E_k(D_{tm}) \quad \text{Równanie 11.}$$

gdzie:

$E(D_{tm})$  – entropia podzbioru uczącego węzła  $t_m$

$E_k(D_{tm})$  – entropia podzbioru uczącego powstałego w wyniku podziału zbioru  $D_{tm}$  wg atrybutu  $a_k$

$a_k$  – atrybut wybrany jako test

Jako kolejny węzeł danego drzewa wybierany jest atrybut, dla którego przyrost informacji jest największy. Oznacza to, że preferowane są atrybuty, które dzielą zbiór  $D$  na podzbiory zawierające przykłady o najmniejszym zróżnicowaniu klas.

Druga grupa wskaźników użyteczności atrybutów mierzy przede wszystkim rozdzielność, dywergencję lub dyskryminację pomiędzy klasami. Najczęściej stosowaną miarą jest indeks Giniego:

$$Gini(D_{tm}) = 1 - \sum_{i=1}^n p_i^2 \quad \text{Równanie 12.}$$

gdzie:

$p_i$  – prawdopodobieństwo  $i$ -tej klasy w węźle  $D_{tm}$

$n$  – ilość klas

Miary statystycznej niezależności mogą być również przypisane do jednej z dwóch powyższych grup. Najbardziej popularnym testem w tej kategorii jest statystyka  $\chi^2$ , mierząca różnicę pomiędzy faktycznym rozkładem poszczególnych par wartości atrybutów  $y_1$  i  $y_2$  a ich rozkładem oczekiwanym przy założeniu niezależności tych cech.

### 5.2.3. Testy kandydujące

Zaprezentowane powyżej miary jakości testów wykorzystywane są do dokonania wyboru najlepszego testu ze zbioru testów kandydujących. Zbiór testów kandydujących odpowiada początkowej zawartości zbioru  $S$  dostępnych testów, będącego argumentem rekurencyjnej funkcji konstruowania drzewa.

$a_k : X \mapsto S_{a_k}$  jest funkcją określoną na zbiorze elementów  $X$ , której przeciwdziedzina jest zbiór wartości atrybutu  $a_k$ .

Test jest w ogólnym przypadku dowolną funkcją określoną na zbiorze elementów  $X$ :

$$t : X \mapsto R_t$$

gdzie:  $D_t = (x_i: x_i \in X, y_i)$

$D_t$  - zbiór rekordów uczących

$x_i$  - parametry opisujące rekord

$y_i$  - etykieta klasy

W praktyce rozważane są wyłącznie testy funkcyjnie zależne od atrybutów, przyjmuje się także ograniczenie, że wynik testu może zależeć od wartości dokładnie jednego atrybutu. Użycie większej liczby atrybutów w jednym teście może uprościć drzewo, jednak dobór tego rodzaju testów jest procesem znacznie trudniejszym i kosztownym obliczeniowo.

Test reprezentuje podział zbioru przykładów rozważanego w węźle drzewa decyzyjnego na podzbiory odpowiadające wynikom testu.

Rodzaje testów:

- a. testy tożsamościowe – test utożsamiany jest z wartością atrybutu; testy te stosowane są dla atrybutów nominalnych lub porządkowych o skończonej liczbie wartości.

$$t(x) = a_k(x) \quad \forall x \in X \quad \text{Równanie 13.}$$

- b. testy równościowe

$$t(x) = \begin{cases} 1 & \text{dla } a_k(x) = v \\ 0 & \text{dla } a_k(x) \neq v \end{cases} \quad \text{gdzie } v \in S_{a_k} \quad \text{Równanie 14.}$$

- c. testy przynależnościowe – sprawdzające przynależność do zbioru dla wartości atrybutu; mogą prowadzić do prostszych drzew niż te, które uzyskiwane są za pomocą testów tożsamościowych czy równościowych; dla każdego atrybutu o skończonej przeciwdziedzinie  $A$  można określić  $2^{|A|}$  różnych testów przynależnościowych, dlatego są one trudne do praktycznego zastosowania.

$$t(x) = \begin{cases} 1 & \text{dla } a_k(x) \in V \\ 0 & \text{dla } a_k(x) \notin V \end{cases} \quad \text{gdzie } V \subset S_{a_k} \quad \text{Równanie 15.}$$

$V$  – właściwy podzbiór zbioru wartości atrybutu

- d. testy nierównościowe - sprawdzające nierówność dla wartości atrybutu;

$$t(x) = \begin{cases} 1 & \text{dla } a_k(x) \leq v \\ 0 & \text{dla } a_k(x) > v \end{cases} \quad \text{gdzie } v \in S_{a_k} \quad \text{Równanie 16.}$$

$v$  – wartość progowa atrybutu  $a_k$

#### 5.2.4. Kryterium stopu

Ostateczny kształt drzewa, poza kryterium wyboru testu, determinowany jest również przez takie parametry jak kryterium stopu czy stopień scalania drzewa.

Kryterium stopu określa kiedy proces rozrostu drzewa ma się zakończyć. Przy każdym rekurencyjnym wywołaniu procedury generowania drzewa liczebność przekazywanego jej zbioru rekordów uczących ulega zmniejszeniu, ponieważ pozostają w nim tylko te rekordy, dla których wynik testu odpowiada wychodzącej gałęzi drzewa. Wzrost drzewa zatrzymuje się jeśli przekazywany zbiór przykładów jest pusty bądź zawiera przykłady wyłącznie jednej kategorii. W przypadku gdy każdy test może być zastosowany tylko raz, wyczerpaniu może ulec także zbiór możliwych do wykorzystania testów  $S$ . Kryterium zatrzymania rozbudowy drzewa może być również osiągnięcie założonej liczby gałęzi lub dokładności podziałów. W tych przypadkach tworzony jest liść z etykietą kategorii większościowej. W sytuacji gdy zbiór uczący nie jest poprawny lub zbiór atrybutów nie opisuje przykładów w dostatecznym stopniu, nie jest reprezentatywny dla pojęcia docelowego algorytm zatrzymuje się zgłaszając niemożność nauczenia się pojęcia docelowego lub tworzone są liście z etykietą większościową.

#### 5.2.5. Nadmierne dopasowanie

Procedura indukcyjnej budowy drzewa może prowadzić do nadmiernego dopasowania do zbioru uczącego, co manifestuje się bardzo małym błędem na zbiorze uczącym i jednocześnie zbyt dużym błędem rzeczywistym. Drzewo takie może mieć nadmiernie złożoną strukturę i odzwierciedlać przypadkowe zależności występujące w zbiorze uczącym. Problemowi temu, zgodnie z zasadą brzytwy Ockhama, przeciwdziała tzw. przycinanie, mające na celu ograniczenie złożoności drzewa i zmniejszenie błędu rzeczywistego przy zachowaniu jak największej spójności ze zbiorem uczącym. Przycinanie polega na zastąpieniu niektórych poddrzew liśćmi z etykietą kategorii rekordów najczęściej występujących. Zabieg taki skutkuje pogorszeniem klasyfikacji zbioru uczącego, ale może poprawić wyniki dla zbioru zewnętrznego. Uproszczenie struktury drzewa daje także większą czytelność dla człowieka oraz oszczędność pamięci i lepszą wydajność obliczeniową. Algorytmy przycinania mogą realizować swoje zadanie w różny sposób:

- przycinanie proste (*ang. post-pruning*) – opisane wcześniej; najczęściej stosowane,
- przycinanie od środka (*ang. post-pruning*) – przycięcie węzła nie oznacza zastąpienia go liściem, ale jednym z jego węzłów potomnych; wymaga reorganizacji drzewa, ze względu na związane z nią komplikacje jest rzadko stosowany,
- przycinanie w trakcie wzrostu (*ang. pre-pruning*) – jest to właściwie modyfikacja kryterium stopu zapobiegająca nadmiernemu rozrostowi drzewa w trakcie jego konstrukcji; znalezienie odpowiedniego kryterium stopu jest trudne, a mechanizm taki mało precyzyjny; zwykle błąd rzeczywisty jest znaczny.

Kluczowe znaczenie ma sposób szacowania błędu rzeczywistego, na podstawie którego podejmowana jest decyzja o przycinaniu. Wyróżnić można dwie metody:

- przycinanie z oddzielnym zbiorem przycinania – stosowane w przypadkach odpowiednio licznych zbiorów uczących; rekordy dzielone są na zbiór treningowy i zbiór przycinania, jeśli po zastąpieniu poddrzewa liściem nie zwiększy się błąd dla zbioru przycinania, sprawdzane poddrzewo jest przycinane,
- przycinanie na podstawie rekordów uczących – stosowane jeśli zbiór uczący nie pozwala na wydzielenie części rekordów, decyzja podejmowana jest na podstawie różnego rodzaju heurystyk bazujących na całym zbiorze uczącym.

#### 5.2.6. Zespoły drzew decyzyjnych (lasy losowe i drzewa wzmacniane)

Pojedyncze drzewa decyzyjne mogą być bardzo efektywnymi klasyfikatorami, niekiedy jednak możliwa jest poprawa dokładności poprzez łączenie wyników zbioru drzew. Lasy losowe (*ang. random forest*) i modele wzmacniane (*ang. boosting*) to dwie metody łączenia drzew.

Losowy las realizuje algorytm Random Forest opracowany przez Breimana [25]. Ze zbioru uczącego pobierane są próby losowe (losowanie ze zwracaniem), które tworzą nowe zbiory treningowe dla poszczególnych drzew. Następnie odpowiedni algorytm generuje drzewa. Zestaw testów możliwych do wykorzystania w węzłach, w przeciwieństwie do prostych drzew, obejmuje jedynie losowo wybrane atrybuty, a nie pełny ich zbiór. W metodzie tej wynik klasyfikacji całego modelu wyznaczany jest poprzez uśrednienie lub głosowanie na podstawie wartości zmiennej zależnej przewidywanej przez każde z drzew składowych modelu. Pojedyncze drzewa charakteryzują się zwykle niskim błędem klasyfikacji i dużą wariancją, stąd wrażliwe są na szum w danych. Technika Random Forest pozwala na osiągnięcie niskiej wartości błędu oraz zmniejszenie wariancji poprzez agregację słabo skorelowanych ze sobą drzew o niskim błędzie i wysokiej wariancji. Niska wartość błędu lasu wynika z generowania drzew bez przycinania, natomiast niska wartość wariancji związana jest z łączeniem drzew oraz losowym wyborem atrybutów.

Wzmacnianie (*ang. boosting*) jest techniką polegającą na zastosowaniu sekwencji prostych modeli, przy czym każdy kolejny przykład większą wagę do tych rekordów, które zostały błędnie sklasyfikowane przez poprzednie modele. Wykazano, że dla niektórych problemów estymacji i predykcji przewidywania generowane przez takie modele są bliższe rzeczywistym wartościom niż rozwiązania podawane przez pojedyncze drzewa.

### 5.3. Metody probabilistyczne – algorytmy Bayesa

Podstawą probabilistycznych metod uczenia się jest twierdzenie podane przez Thomasa Bayesa, XVIII-wiecznego matematyka angielskiego. Teoria Bayesa mówi o szacowaniu prawdopodobieństwa zdarzenia z uwzględnieniem zarówno prawdopodobieństwa *a priori* (przed obserwacją) zdarzenia jak i posiadanych danych.

Do wyrażania zależności między zdarzeniami służą prawdopodobieństwa warunkowe. Prawdopodobieństwo zajścia zdarzenia  $A$  wtedy, gdy zaszło zdarzenie  $B$ , definiuje się dla dowolnych zdarzeń  $A, B \subseteq \Sigma$  (gdzie  $\Sigma$  - zbiór zdarzeń elementarnych), przy założeniu, że  $P(B) > 0$ , jako prawdopodobieństwo warunkowe

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{Równanie 17.}$$

gdzie  $P(A \cap B)$  oznacza prawdopodobieństwo łącznego zajścia zdarzeń  $A$  i  $B$ .

Z prawdopodobieństwami warunkowymi wiąże się wzór Bayesa:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad \text{Równanie 18.}$$

oraz wzór określający prawdopodobieństwo całkowite (zupelne):

$$P(B) = \sum_{i=1}^n P(A_i) P(B|A_i) \quad \text{Równanie 19.}$$

który zachodzi dla każdego zdarzenia  $A \subseteq \Sigma$  oraz zdarzeń  $B_1, B_2, \dots, B_n \subseteq \Sigma$  spełniających warunki:

1.  $P(B_i) > 0$  dla każdego  $i=1, 2, \dots, n$ ,
2.  $P(B_i \cap B_j) = 0$  dla  $i, j=1, 2, \dots, n; i \neq j$ ,
3.  $\bigcup_{i=1}^n B_i = \Sigma$ .

W kontekście metod uczenia się i odkrywania wiedzy wzór będący treścią twierdzenia Bayesa (Równanie 20) podaje się zazwyczaj w postaci:

$$P(C_i|X) = \frac{P(X|C_i) \cdot P(C_i)}{P(X)} \quad \text{Równanie 20.}$$

gdzie:

$X = (x_1, x_2, \dots, x_n)$  jest  $n$ -wymiarowym wektorem wartości atrybutów opisującym dany przykład ze zbioru uczącego (rekord danych),

$C_i$  oznacza pewną klasę, a  $X$  dane (obserwacje) mogące wpływać na ocenę prawdopodobieństwa tej klasy,  
 $P(C_i)$  jest prawdopodobieństwem *a priori* klasy  $C_i$ , określonym bez uwzględniania jakichkolwiek danych,  
 $P(X)$  jest prawdopodobieństwem danych,  
 $P(X|C_i)$  jest prawdopodobieństwem danych  $X$  w klasie  $C_i$ .

Celem wyznaczania prawdopodobieństw *a posteriori* jest zwykle wybór klasy najbardziej prawdopodobnej w świetle zaobserwowanych danych. W takim przypadku można pominąć prawdopodobieństwo danych  $P(X)$ , gdyż nie zależy ono od klas i nie wpływa na wynik porównania prawdopodobieństw poszczególnych klas  $C_i$ . Uwzględnia się wyłącznie iloczyn występujący w liczniku równania Bayesa:

$$P(C_i|X) = P(X|C_i) \cdot P(C_i) \quad \text{Równanie 21.}$$

Przy nieznanym prawdopodobieństwie klas  $C_i$  zakłada się, że są one sobie równe lub szacuje jego wartość na podstawie ich proporcji w próbie pobranej ze zbioru uczącego.

#### 5.4. Klasyfikacja bayesowska

Interpretując wzór Bayesa w kontekście uczenia się pojęć, klasy rozważane są jako funkcje klasyfikujące elementy dziedziny (przykłady, obserwacje, rekordy) na podstawie wartości ich atrybutów. Rolę danych pełni zbiór uczący ( $T$ ) z danymi etykietami pojęcia docelowego. Zadanie uczenia się polega na wyborze z przestrzeni klas  $C$  klasy najlepszej na podstawie zbioru uczącego  $T$ . Wyboru tego można dokonać kilkoma metodami, m.in. wykorzystując algorytm ML (*ang. Maximum Likelihood* - metoda maksymalnej zgodności), MAP (*ang. Maximum a Posteriori* - metoda maksymalnego prawdopodobieństwa *a posteriori*) czy BOC (*ang. Bayes Optimal Classifier* - optymalny klasyfikator bayesowski). Ze względu na pewne ograniczenia algorytmy te nie mają dużego znaczenia praktycznego.

##### 5.4.1. Naiwny klasyfikator bayesowski

Naiwny klasyfikator Bayesa jest jedną z metod uczenia maszynowego, stosowaną do rozwiązywania problemów sortowania i klasyfikacji. Zadaniem klasyfikatora Bayesa jest przyporządkowanie nowego przypadku, obserwacji do jednej z klas decyzyjnych ze zdefiniowanego *a priori* i skończonego zbioru. Podstawowym założeniem w tego rodzaju klasyfikatorach jest wzajemna niezależność statystyczna rozkładów predyktorów (zmiennych niezależnych) w klasach. Powyższe założenie znacznie upraszcza klasyfikację poprzez redukcję zadania wielowymiarowego do szeregu jednowymiarowych, warunkowe gęstości prawdopodobieństwa dla klas  $P(X|C_i)$  mogą być szacowane osobno dla każdej zmiennej.

Założenie to pozwala na zapis naiwnego klasyfikatora bayesowskiego w postaci:



$$P(X|C_i) = \prod_{k=1}^n P(x_k|C_i) \quad \text{Równanie 22.}$$

Prawdopodobieństwa  $P(x_1|C_i), \dots, P(x_n|C_i)$  są szacowane na podstawie zbioru uczącego. Dla zmiennych kategoriycznych stosowane są proporcje, natomiast dla zmiennych ciągłych zakłada się najczęściej rozkład normalny wartości parametru o średniej  $\mu$  i odchyleniu standardowym  $\sigma$  (Równanie 23), rzadziej rozkłady lognormalny, gamma lub Poissona.

$$P(x_k|C_i) = \left\{ \frac{1}{\sigma_{kj} \sqrt{2\pi}} \exp\left( -\frac{(x - \mu_{kj})^2}{2\sigma_{kj}^2} \right) \right\} \quad \text{Równanie 23.}$$

gdzie:

$\mu_{kj}$  – średnia i  $\sigma$  odchylenie standardowe wartość atrybutu  $x_k$  w zbiorze przykładów klasy  $C_i$

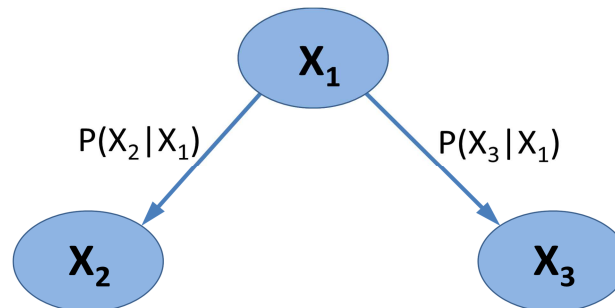
W praktyce, zapewnienie niezależności atrybutów jest trudne lub niemożliwe dla większości rzeczywistych problemów – stąd określenie klasyfikator „naiwny”. Jednak wyniki uzyskiwane przy zastosowaniu tego klasyfikatora, często trafniejsze niż innych bardziej złożonych metod, w połączeniu z dużą efektywnością obliczeniową sprawiają, że jest on jednym z częściej wykorzystywanych. Podejście takie jest szczególnie odpowiednie dla problemów o dużej liczbie wymiarów przestrzeni zmiennych wejściowych.

#### 5.4.2. Sieci bayesowskie (sieci przekonań)

Jak wspomniano wyżej w rzeczywistych problemach wystąpienie jakiegoś zdarzenia jest ściśle uzależnione od innego zdarzenia, stąd założenie o niezależności atrybutów jest zbyt dużym ograniczeniem.

Sieci Bayesa służą do przedstawiania niepewności pozwalając na intuicyjną graficzną wizualizację wiedzy zawierającą wzajemne oddziaływania pomiędzy różnymi źródłami niepewności (np. niepewności eksperta czy dostępnych danych). Dużą zaletą sieci bayesowskich jest zdolność radzenia sobie z niekompletnymi zbiorami danych. Ponadto w przypadku tzw. sieci o poprawnej strukturze, pozwalają na łatwą reprezentację rzeczywistych połączeń, dzięki czemu wnioskowanie o relacjach jest łatwe do przeprowadzenia. Należy jednak pamiętać, że w przypadku sieci innych niż te o poprawnej strukturze, sama sieć w żadnym stopniu nie wpływa na występowanie lub brak zależności pomiędzy atrybutami, a jest ona jedynie wyrazem istniejących przekonań, założeń na ten temat. Niezwykle istotną cechą sieci bayesowskich w połączeniu z bayesowskimi technikami statystycznymi jest również możliwość połączenia wiedzy eksperta z danymi, co pozwala na tworzenie systemów opartych na apriorycznej wiedzy o danym problemie wzmocnionej rzeczywistymi danymi.

Sieć bayesowska to acykliczny graf skierowany, w którym węzły reprezentują określone na dziedzinie atrybuty, a łączące je krawędzie kierunki zależności pomiędzy nimi (Rycina 20). Ilościowo zależności te są określone przez związaną z węzłem tablicę prawdopodobieństw warunkowych ( $P(X|pa(x))$ , gdzie  $pa(X)$  oznacza zbiór wszystkich poprzedników węzła  $X$ ) określających wpływ wywierany na ten węzeł przez jego poprzedników w grafie (łączny rozkład prawdopodobieństwa).



**Rycina 20.** Struktura sieci bayesowskiej dla trzech zdarzeń  $X_1, X_2, X_3$ , gdzie istnieje relacja  $X_1 \rightarrow X_2$ , oraz  $X_1 \rightarrow X_3$ , brak relacji pomiędzy  $X_2$  i  $X_3$  (warunkowa niezależność).

Z definicji węzeł  $a_j$  jest następnikiem węzła  $a_i$  (lub równoważnie  $a_i$  jest poprzednikiem węzła  $a_j$ ) w sieci bayesowskiej, co oznacza się przez  $a_i \rightarrow a_j$ , jeśli spełniony jest jeden z warunków:

- w sieci istnieje krawędź skierowana od węzła  $a_i$  do  $a_j$ , czyli  $a_i \rightarrow a_j$ ,
- w sieci istnieje krawędź skierowana od węzła  $a_i$  do pewnego węzła  $a_k$  i węzeł  $a_j$  jest następnikiem węzła  $a_k$ , czyli  $a_i \rightarrow a_k$  i  $a_k \rightarrow a_j$ .

Podstawą formalizmu sieci Bayesa jest warunkowa niezależność. Zdarzenie  $A$  jest definiowane jako warunkowo niezależne od zdarzenia  $C$  jeżeli zachodzi zdarzenie  $B$ :

$$P(A|B \cap C) = P(A|B) \quad \text{Równanie 24.}$$

z powyższej definicji można również zapisać:

$$P(A \cap B|C) = P(A|C)P(B|C) \quad \text{Równanie 25.}$$

Z koncepcji tej bezpośrednio wynika przyczynowo-skutkowa struktura sieci oraz ich zdolność do zmniejszenia złożoności reprezentacji. Ogólnie, przykład (rekord) charakteryzowany przez  $n$  zmiennych wymaga obliczenia  $2^n - 1$  rozkładów prawdopodobieństwa dla wyznaczenia  $P(x_1, x_2, \dots, x_n)$ . Korzystając z założenia warunkowej niezależności i reguły łańcucha sieć Bayesa formalnie zapisać można:

$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | pa(x_i)) \quad \text{Równanie 26.}$$

gdzie  $pa_i$  – bezpośrednie poprzedniki węzła  $x_i$ .

#### 5.4.3. Wnioskowanie w sieciach Bayesa

Podstawowym zadaniem sieci bayesowskiej jest wnioskowanie o rozkładzie prawdopodobieństwa nieznanymi wartościami atrybutów na podstawie zbioru uczącego (znanych wartości atrybutów). Innymi słowami, zadanie wnioskowania jest często definiowane jako obliczenie wszystkich prawdopodobieństw *a posteriori* na podstawie danych.

W ogólnym przypadku wnioskowanie w dowolnych sieciach bayesowskich jest problemem NP-trudnym. Problem ten ulega uproszczeniu w szczególnym rodzaju sieci, nazywanych sieciami z pojedynczymi połączeniami (*ang. single connected networks*). W tego typu sieciach dowolne dwa węzły są połączone co najwyżej jedną ścieżką, złożoną z dowolnie skierowanych krawędzi.

#### 5.4.4. Uczenie sieci

Wersja zadania uczenia sieci jest determinowana przez dwa kryteria – znajomość (lub jej brak) struktury sieci oraz stopień obserwowalności danych (pełna lub częściowa) [44]. Stąd cztery warianty indukcji sieci obejmują przypadki:

- znana struktura sieci i pełna obserwowalność – wariant najprostszy zakłada, że określona została poprawna struktura sieci (wrodzona wiedza ucznia), oraz dostępny jest zbiór trenujący, w którym znane są wartości atrybutów dla wszystkich przykładów,
- nieznaną strukturę sieci i pełną obserwowalność - dostępny jest zbiór trenujący ze znanymi wartościami wszystkich atrybutów, ale struktura zależności, najbardziej zgodna z danymi trenującymi, musi być określona przez ucznia,
- znana struktura sieci i częściowa obserwowalność - zakłada, że określona została poprawna struktura sieci, natomiast nie są znane wszystkie wartości atrybutów, najczęściej wykorzystuje się gradientowe metody uczenia się prawdopodobieństw,
- nieznaną strukturę sieci i częściową obserwowalność – przypadek z oczywistych względów najtrudniejszy. Poszukiwanie skutecznych i ogólnych algorytmów dla tego problemu jest przedmiotem aktualnych badań.

##### a. zadanie poszukiwania struktury

Istnieją dwa podejścia do tworzenia struktury sieci bayesowskich. Pierwsze, stosowane zwłaszcza we wczesnym okresie badań nad tego rodzaju sieciami zakłada, że mogą one być projektowane na podstawie wiedzy, doświadczenia i obserwacji przez ekspertów w danej dziedzinie. Drugie podejście zakłada automatyczną lub pół-automatyczną konstrukcję dzięki zastosowaniu algorytmów uczących, które określają

relacje zależności i niezależności pomiędzy zmiennymi (węzłami). Ze względu na sposób optymalizacji metody tworzenia sieci dzielą się na:

- dokładne, gdzie algorytm przeszukuje wszystkie możliwe sieci w celu wybrania optymalnej,
- przybliżone, które nie gwarantują optymalnego rozwiązania, ale przeważnie prowadzą do rozwiązań bliskich optymalnemu.

Problem poszukiwania struktury sieci, czyli wytypowania grafu najlepiej modelującego zależności w danych, jest NP-trudny. Przy dużej ilości atrybutów (zmiennych) charakteryzujących przykłady przestrzeni poszukiwań jest zbyt duża dla jakichkolwiek automatycznych metod. Redukcja przestrzeni wyszukiwania jest możliwa poprzez określanie kolejności węzłów, która dyktuje hierarchię zależności w sieci, wykorzystanie apriorycznej wiedzy na temat istniejących zależności do budowy prototypu sieci a następnie jej optymalizację czy zastosowanie metod Monte Carlo i zbadanie ograniczonego zbioru możliwości.

b. zadanie szacowania parametrów

Przy znanej strukturze sieci kolejnym krokiem budowy jest określenie parametrów modelu kodujących siłę zależności pomiędzy węzłami. Szacowanie parametrów polega na określeniu warunkowych prawdopodobieństw na podstawie danych i przy znanej strukturze sieci. Podobnie jak w przypadku struktury, parametry mogą być wyznaczone przez eksperta i/lub w wyniku procesu uczenia na podstawie danych. Przy pełnej obserwowalności danych, estymacja parametrów jest trywialna i polega na policzeniu odpowiednich prawdopodobieństw przykładów. W przypadku brakujących wartości atrybutów do estymacji parametrów najczęściej stosowane są algorytmy maksymalizacji wartości oczekiwanej (ang. *expectation-maximization*, *EM*), metody Monte Carlo, aproksymacja Gaussa, aproksymacje ML i MAP [44].

## II. CEL PRACY

Znane, tradycyjne metody oceny efektów toksycznych potencjalnych leków *in vitro* posiadają pewne ograniczenia, które limitują ich zastosowanie we wczesnych etapach badań potencjalnego leku. Jednocześnie intensywny rozwój technik informatycznych i wzrost znaczenia komputerowych metod obliczeniowych oferuje możliwość stworzenia narzędzia wspomagającego i uzupełniającego tradycyjne eksperymenty oraz pozwalającego ograniczyć użycie zwierząt laboratoryjnych, zwiększyć bezpieczeństwo badań klinicznych, a także obniżyć koszty procesu rozwoju leku.

Celem pracy było wykorzystanie realizowanych komputerowo technik eksploracji danych (*ang. data mining*) do opracowania narzędzia wspomagającego i uzupełniającego tradycyjne badania potencjału arytmogennego substancji chemicznych i umożliwiającego skringing cząsteczek we wczesnej fazie badań przedklinicznych potencjalnych leków.

Realizacja celu pracy obejmowała:

- **systematyczny przegląd literatury i budowę bazy danych**

W przypadku każdego modelu kluczowe znaczenie dla jego sprawności i zdolności generalizacji mają dane zastosowane do jego budowy. Dostępność toksykologicznych baz danych zawierających informacje z doświadczeń *in vitro* jest niewielka, stąd pierwszym etapem pracy było zgromadzenie reprezentatywnych danych literaturowych, odpowiednich do opracowania modeli predykcyjnych.

- **opracowanie metodyki standaryzacji danych**

Otrzymywane w doświadczeniach *in vitro* wartości stężenia  $IC_{50}$  wykazują znaczne zróżnicowanie w zależności od zastosowanego modelu i temperatury utrzymywanej podczas eksperymentów. Z tego powodu konieczne jest opracowanie metod umożliwiających bezpośrednio porównanie danych o różnej charakterystyce i ekstrapolację wyników umożliwiającą wnioskowanie o efektach *in vivo*.

- **budowę modeli predykcyjnych**

Zadaniem tworzonych modeli - algorytmów matematycznych realizowanych w warunkach *in silico* - będzie klasyfikacja substancji na potencjalnie arytmogenne i potencjalnie bezpieczne dla wybranego modelu komórkowego przy zdefiniowanych warunkach oznaczenia. Pierwszym etapem budowy algorytmów będzie wyznaczenie i ocena empiryczna różnego rodzaju możliwie prostych deskryptorów opisujących strukturę i właściwości fizyko-chemiczne cząsteczek. Spośród wszystkich wyznaczonych deskryptorów w procesie analizy wrażliwości wybrane zostaną te, które posłużą do budowy modeli. Do

tworzenia systemu predykcji zastosowane zostaną różne metody inteligencji obliczeniowej m.in. sztuczne sieci neuronowe, metody podziałów rekursywnych (drzewa decyzyjne i ich zbiory) oraz algorytmy Bayesowskie i systemy oparte o logikę nieostrą.

- **weryfikację modeli**

Weryfikacja opracowanych modeli zostanie przeprowadzona za pomocą niezależnego zbioru walidacyjnego, w którym znajdą się rekordy spoza bazy danych wykorzystanej do budowy modelu.

Przewidywanym efektem pracy jest stworzenie systemu wspomagającego proces badań i rozwoju potencjalnych leków oraz ułatwiającego ocenę bezpieczeństwa leków i innych związków chemicznych obecnych na rynku. System ten oparty będzie o najlepsze uzyskane modele, a więc te, które będą wykazywały największą zdolność generalizacji. Jego wykorzystanie podczas rozwoju potencjalnych leków pozwoli ukierunkować badania, a w konsekwencji ograniczyć liczbę koniecznych doświadczeń *in vitro*.

### III. CZĘŚĆ DOŚWIADCZALNA

#### 1. BUDOWA BAZY DANYCH

Toksykologia obliczeniowa jak wspomniano wcześniej znalazła się w ostatnim czasie się w obszarze zainteresowań przemysłu chemicznego i farmaceutycznego oraz agencji zajmujących się rejestracją i nadzorowaniem bezpieczeństwa leków a także agencji ochrony środowiska. Jednym z głównych wyzwań pojawiających się podczas tworzenia nowych modeli QSAR czy oceny już istniejących jest brak toksykologicznych baz danych zawierających odpowiednie informacje. Rozwój technik informatycznych, niskie koszty dużych przestrzeni dyskowych, gromadzenia i przechowywania danych sprawiają, że coraz więcej danych z różnych źródeł (w tym laboratoriów naukowych czy badawczych) jest udostępnianych publicznie. Dostępność wysokiej jakości danych z doświadczeń *in vitro* uzyskanych w standardowych warunkach czy danych dotyczących toksyczności u ludzi jest jednak niewielka. Większość danych znajduje się w posiadaniu firm farmaceutycznych i nie jest udostępniana publicznie. Choć wraz ze wzrostem znaczenia metod alternatywnych w badaniu bezpieczeństwa leków podejmowane są różne inicjatywy, w których jednym z celów jest opracowanie i publikacja toksykologicznych baz danych [96]. Większość informacji wykorzystywanych do budowy toksykologicznych modeli predykcyjnych jest obecnie pozyskiwana z publicznie dostępnych baz danych jak np. Toxnet, ToxML czy DSSTox – zbiorów danych eksperymentalnych, których źródłem są organy regulujące i nadzorujące lub przemysł. Część dostępnych baz danych nie jest jednak postrzegana jako wiarygodne źródło informacji ze względu na pewne ograniczenia m.in. brak pewności co do jakości badań, których wyniki włączono do bazy, obecność danych uzyskanych w różnych (często nieokreślonych dokładnie w bazie danych) warunkach eksperymentalnych, niekompletność identyfikacji źródła, brak standaryzacji wymagań co do jednostek pomiarowych czy zawartość informacji wyłącznie na temat pozytywnych wyników testów toksyczności. Brakuje niezbędnych w modelowaniu QSAR informacji szczegółowych dotyczących np. podanej dawki/stężenia, czasu od podania substancji do oceny badanego parametru. Ponadto w przypadku, gdy dostępnych jest kilka różnych wyników jednego rodzaju badań dla danego związku chemicznego w bazie danych może znajdować się tylko jeden, który jej autorzy uznali za najbardziej wiarygodny. Publikowane bazy danych zawierają informacje dla wybranych efektów toksykologicznych, nieliczne obejmują kardiotoxycywność. W przypadku każdego modelu kluczowe znaczenie dla jego sprawności i zdolności generalizacji mają informacje zastosowane do jego konstrukcji, ich jakość, spójność oraz reprezentatywność, dlatego też pierwszym etapem pracy było zgromadzenie, usystematyzowanie i ocena dostępnych w literaturze naukowej danych dotyczących interakcji związków chemicznych z kanałem potasowym kodowanym przez gen hERG oraz stworzenie bazy danych, którą można wykorzystać do budowy modeli predykcyjnych.

### 1.1. Metodyka

Pierwszym etapem budowy bazy danych było sporządzenie listy leków o udowodnionym działaniu kardiotoksycznym (odnotowane TdP lub wydłużenie odstępu QT) oraz leków, które badane były w kierunku takiego działania. Wstępny spis został przygotowany w oparciu o bazę danych Roberta Fenichela [99], Międzynarodowy Rejestr Leków Powodujących Arytmie - Arizona CERT [10] (*ang. Center for Education and Research on Therapeutics*) oraz publikację Oliviera Roche [283]. Do kolejnych etapów pracy wybrano leki blokujące kanał hERG lub szybki opóźniony prąd prostowniczy ( $I_{Kr}$ ), wykluczono natomiast leki wywołujące zmianę potencjału czynnościowego komórek mięśnia sercowego w inny sposób. W celu uzyskania danych o wartości  $IC_{50}$  dla wybranych leków gromadzono publikacje dotyczące tego rodzaju eksperymentów, przeszukując bibliograficzne bazy danych Scopus, Medline oraz Google Scholar według schematu przedstawionego w tabeli 1. W przypadku braku wyników wyszukiwania dla nazwy leku jako słowa kluczowego używano nazwy klasy leku.

**Tabela 1.** Schemat kwerend stosowanych podczas wyszukiwania informacji.

baza danych	limit czasowy	słowa kluczowe
Medline	... - 2010	Nazwa leku/klasa leku +
		$IC_{50}$
Scopus		hERG
		Human Ether-A-Go-Go
Google Scholar		potassium channel
		potassium current

Szczegółowej analizie podlegały wszystkie dostępne anglojęzyczne publikacje (ponad 1000 pozycji). W specjalnie przygotowanym arkuszu kalkulacyjnym MS Excel zapisywano wyłącznie oryginalne wyznaczone przez autorów danej publikacji eksperymentalnie wartości  $IC_{50}$ . Poza wartością stężenia powodującą 50% blok prądu potasowego w zgromadzonych publikacjach poszukiwano informacji dodatkowych charakteryzujących doświadczenie:

- model wykorzystany do badań,
- technika pomiaru,
- rodzaj transfekcji,
- warunki eksperymentalne (temperatura oraz stężenie jonów  $K^+$  w płynie omywającym komórki),
- typ protokołu napięć: step, ramp, step-ramp, potencjał czynnościowy; potencjał utrzymania, potencjał depolaryzujący, potencjał pomiaru oraz czas trwania pulsu depolaryzującego.



Pełny zakres informacji gromadzono wyłącznie dla eksperymentów z wykorzystaniem pomiarów elektrofizjologicznych oraz jednego z trzech najczęściej stosowanych systemów ekspresji: XO, HEK lub CHO

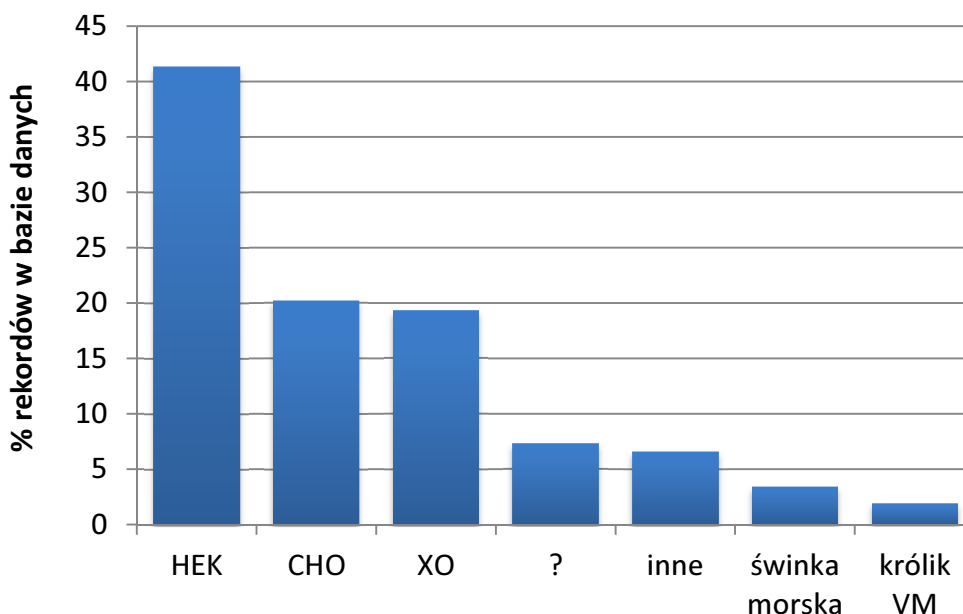
## 1.2. Wyniki

### 1.2.1. Baza danych

Ostateczna lista publikacji spełniających zadane kryteria włączenia zawiera ponad 300 pozycji. Dotyczą one 325 różnych substancji chemicznych, z których dla 277 wyznaczono przynajmniej jedną wartość  $IC_{50}$ . Ogółem zgromadzono 865 wartości  $IC_{50}$ . Ze zbioru wykluczono te wartości, dla których nie zdefiniowano modelu, nie znaleziono publikacji źródłowej bądź nie była ona dostępna. Pozostałych 774 wartości dla 256 substancji uzyskano podczas eksperymentów przeprowadzanych z wykorzystaniem 17 różnych modeli. W większości doświadczeń stosowano komórki HEK, jako drugie na liście najczęściej wybieranych modeli znalazły się komórki CHO i XO. Pozostałe badania przeprowadzano m.in. na komórkach mięśniówki komórek serc świnek morskich, królików, myszy, szczurów i psów, mysich komórkach AT-1 lub L oraz komórkach COS-7. Dla 239 substancji chemicznych wyznaczono wartość  $IC_{50}$  (łącznie 673 rekordy) z wykorzystaniem systemu HEK, CHO lub XO. Kompletny zestaw informacji zgromadzono dla 598 rekordów. Tabela 2 przedstawia stworzoną bazę danych. Poszczególne rekordy bazy danych odnoszą się do pojedynczych wyznaczonych eksperymentalnie wartości  $IC_{50}$ . Pierwsze dwie kolumny zawierają nazwę międzynarodową związku (lub inne oznaczenie dla niezarejestrowanych substancji) oraz wartość stężenia powodującego 50% blok kanału hERG wyrażone w mikromolach [ $\mu M$ ]. Kolumna „Źródło” zawiera odnośniki do publikacji opisujących oryginalne, przeprowadzone przez autorów doświadczenia, w których dana wartość  $IC_{50}$  została wyznaczona. Kolejne kolumny w bazie danych zawierają informacje na temat warunków eksperymentalnych, które mogą wpływać na wartość uzyskiwanych wyników. Do czynników tych należą: (1) zastosowany system ekspresji; (2) typ transfekcji (stabilna lub przejściowa) dla komórek CHO i HEK; (3) metoda wyznaczenia  $IC_{50}$ ; (4) temperatura, w której przeprowadzono oznaczenie; w przypadku, gdy temperatura nie była precyzyjnie zdefiniowana (warunki określano jako „temperatura pokojowa”) w bazie danych notowano wartość 23°C, tę samą wartość uzupełniano, gdy w publikacji dotyczącej oznaczeń z wykorzystaniem komórek XO nie było żadnego określenia temperatury (komórki XO nie są stabilne w temperaturze fizjologicznej i niemożliwe jest przeprowadzenie doświadczeń w takich warunkach); (5) stężenie jonów potasowych w płynie omywającym; (6) rodzaj protokołu napięć – ramp, step-ramp, step, ramp-step lub potencjał czynnościowy (AP); (7) charakterystyka protokołu napięć: potencjał utrzymania, napięcie depolaryzujące, potencjał błonowy podczas repolaryzacji, przy którym mierzono natężenie prądu  $I_{Kr}$ , oraz czas trwania pulsu depolaryzującego. Wszystkie niepewne lub brakujące informacje oznaczano za pomocą znaku zapytania.

## 1.2.2. Charakterystyka rekordów

Największa liczba eksperymentów przeprowadzona została z wykorzystaniem embrionalnych komórek nerki ludzkiej, jajników chomika chińskiego oraz oocytów żab z *Xenopus laevis* (odpowiednio 40.7%, 24.0% i 22.2%). W 1.3% przypadków model zastosowany do doświadczeń nie był zdefiniowany (Rycina 21).



**Rycina 21.** Modele komórkowe stosowane w oznaczeniach wartości  $IC_{50}$ . Oś odciętych reprezentuje procentowy udział rekordów w bazie danych, oś rzędnych - modele.

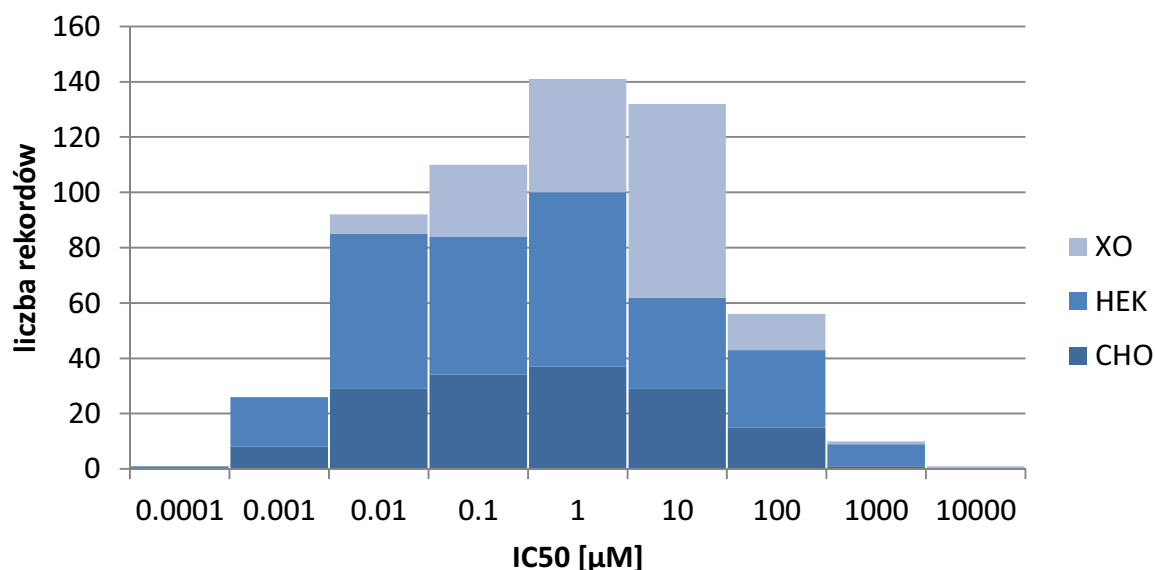
Zdecydowana większość kultur komórkowych była stabilnie transfekowana DNA lub mRNA kodującym podjednostkę  $\alpha$  kanału potasowego hERG, jedynie 4% rekordów odnosiło się do pomiarów wykonanych na komórkach transferowanych przejściowo. Wszystkie pomiary (167 rekordów) z zastosowaniem oocytów żaby *Xenopus* wykonano za pomocą dwuelektrodowej techniki voltage-clamp. Oznaczenia z wykorzystaniem komórek CHO i HEK były wykonywane prawie wyłącznie metodą patch-clamp w konfiguracji whole-cell (405 rekordów). Dwadzieścia wartości  $IC_{50}$  otrzymano techniką PatchXpress.

Na rycinie 22 przedstawiono zakres aktywności biologicznej zgromadzonych związków reprezentowany w zbiorze danych.

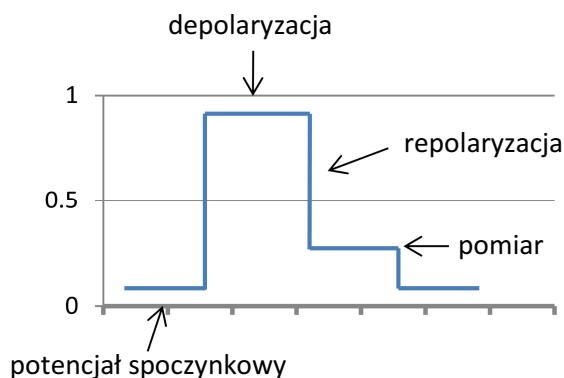
Około dwie trzecie eksperymentów przeprowadzonych zostało w temperaturze pokojowej (19-25°C) i jedna trzecia w temperaturze określanej jako fizjologiczna (34-37°C). Wszystkie doświadczenia wykorzystujące komórki XO prowadzone były w temperaturze pokojowej, ze względu na niestabilność tych komórek podczas pomiarów elektrofizjologicznych w temperaturze 37°C. Połowa doświadczeń na komórkach HEK oraz

prawie jedna trzecia na komórkach CHO prowadzona była w temperaturze fizjologicznej. W kilku przypadkach nie udało się ustalić przy jakiej wartości temperatury dokonano oznaczeń.

W poszczególnych eksperymentach stosowano także różne stężenia jonów potasowych w płynie omywającym komórki (2 – 100 mmol/L). W większości prac (83.3% wszystkich rekordów) stężenie potasu mieściło się w zakresie 2-5.6 mmol/L. Dla 1% oznaczeń nie podano zewnątrzkomórkowego stężenia jonów  $K^+$ .



**Rycina 22.** Dystrybucja wartości  $IC_{50}$  dla kanału hERG w zbiorze danych. Oś odciętych reprezentuje liczbę rekordów, oś rzędnych zakres wartości  $IC_{50}$ .



**Rycina 23.** Schemat protokołu pomiarowego typu step.

Kolejnym czynnikiem różnicującym pomiary opisywane w zgromadzonych publikacjach był rodzaj i charakterystyka protokołu napięć zastosowanego do wyzwolenia prądu hERG. W doświadczeniach stosowano 5 rodzajów protokołów. Zdecydowana większość (91% rekordów) wartości  $IC_{50}$  została wyznaczona z zastosowaniem protokołu typu *step* polegającego na skokowej zmianie napięcia przykadanego do błony

komórkowej z potencjału spoczynkowego (utrzymania) do wartości wywołującej depolaryzację a następnie skokowe obniżenie przykładanego napięcia do wartości powodującej repolaryzację i otwarcie kanałów potasowych (Rycina 23). Potencjał utrzymania dla przeważającej liczby eksperymentów mieścił się w zakresie -95 do -60 mV, w 2% doświadczeń wartość tego potencjału wynosiła od -55 do -40 mV. Stosowane wartości napięć pulsów depolaryzujących mieściły się w zakresie -40 do 80 mV, najczęściej depolaryzację wywoływano impulsem o napięciu 20 mV. Zakres wartości napięć, w których dokonywano pomiaru amplitudy prądu hERG wynosił od -140 mV do 0 mV.

Tabela 2. Natywna baza danych.

Substancja	IC50 [μM]	model	źródło <sup>1</sup>	trans <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
1-(bis(4-fluorofenilo)metylo) piperazyna	1.6	CHO	[363]	stabilna	35	whole cell PC	4	1; 2	-80	20	-50	step
4,4'-difluorobenzhydrol	99	CHO	[363]	stabilna	35	whole cell PC	4	1; 2	-80	20	-50	step
4,4'-difluorobenzofenon	72	CHO	[363]	stabilna	35	whole cell PC	4	1; 2	-80	20	-50	step
4,4'-dimetyl-difenhydramina	18.5	XO	[338]		23	voltage-clamp 2-elektrodowy	10	2; 2	-90	0	-100	step
4-aminopirydyna	4400	HEK	[278]	stabilna	37	whole cell PC	4	1; 1.5	-40	30	-40	step
5-hydroksypropafenon	0.88	CHO	[9]	stabilna	21.5	whole cell PC	4	5; 5	-80	-10	-60	step
acehytizyna	465.95	HEK	[141]	stabilna	23	whole cell PC	4	4; 4	-80	20	20	step
AF3013	>335	HEK	[200]	stabilna	37	whole cell PC	4	0.5; 0.5	-75	10	-40	step
ajmalina	1.04	HEK	[175]	stabilna	21	whole cell PC	5	0.4; 0.4	-80	40	-120	step
ajmalina	42.3	XO	[175]		21	voltage-clamp 2-elektrodowy	5	0.4; 0.4	-80	40	-60	step
ambasilid	3.6	CHO	[376]	stabilna	22	whole cell PC	4.8	3.9; 5	-80	30	-60	step
ambasilid	16.1	CHO	[376]	stabilna	22	whole cell PC	4.8	3.9; 5	-80	-30	-60	step
amiodaron	0.048	CHO	[124]	stabilna	35	whole cell PC	4	1; 0.5	-80	20	-40	step
amiodaron	0.07	HEK	[279]	stabilna	37	whole cell PC	4	1; 4	-40	30	-40	step
amiodaron	0.244	CHO	[12]	stabilna	23	PatchXpress 7000A	4	4; 4	-80	20	-40	step
amiodaron	0.421	CHO	[124]	stabilna	23	PatchXpress 7000A	4	1; 0.5	-80	20	-40	step
amiodaron	0.546	CHO	[397]	stabilna	25	whole cell PC	4	0.4; 0.4	-80	20	-40	step
amiodaron	0.8	HEK	[392]	stabilna	23	whole cell 1-electrode VC	4	4; 5.7		20	-50	step
amiodaron	1.157	CHO	[124]	stabilna	23	PatchXpress 7000A	4	4; 4	-80	20	-40	step
amiodaron	2.347	CHO	[124]	stabilna	23	PatchXpress 7000A	4	1; 0.5	-80	20	-40	step
amiodaron	9.8	XO	[173]		20	voltage-clamp 2-elektrodowy	5	0.4; 0.4	-80	30	-60	step
amiodaron	15	XO	[247]		23	voltage-clamp 2-elektrodowy	?	2; 2	-90	0	-80	step
amiodaron	24.7	XO	[212]		21.5	voltage-clamp 2-elektrodowy	10	1.2; 1.2	-80	40	-40	step

Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
amiodaron	37	XO	[212]		21.5	voltage-clamp 2- elektrodowy	5	1.2; 1.2	-80	40	-40	step
amiodaron	37.9	XO	[158]		23	voltage-clamp 2- elektrodowy	5.4	2; 0	-90	40	-70	step
amitryptylina	1.7	CHO	[124]	stabilna	23	PatchXpress 7000A	4	1; 0.5	-80	20	-40	step
amitryptylina	1.801	CHO	[124]	stabilna	35	whole cell PC	4	1; 0.5	-80	20	-40	step
amitryptylina	3.26	XO	[151]		21.5	voltage-clamp 2- elektrodowy	2	0.5; 0.5	-70	20	-60	step
amitryptylina	4.66	XO	[151]		21.5	voltage-clamp 2- elektrodowy	2	4; 4	-70	30	-60	step
amitryptylina	4.78	XO	[151]		21.5	voltage-clamp 2- elektrodowy	4	0.5; 0.5	-70	20	-60	step
amitryptylina	5.96	XO	[151]		21.5	voltage-clamp 2- elektrodowy	2	4; 4	-70	10	-60	step
amitryptylina	8.71	XO	[151]		21.5	voltage-clamp 2- elektrodowy	2	4; 4	-70	-10	-60	step
amitryptylina	10	CHO	[358]	stabilna	21	whole cell PC	4.8	3.9; 5	-80	30	-60	step
amitryptylina	23	XO	[151]		21.5	voltage-clamp 2- elektrodowy	2	4; 4	-70	-30	-60	step
amoksycylina	>50000	CHO	[397]	stabilna	25	whole cell PC	4	0.4; 0.4	-80	20	-40	step
amsakryna	0.2094	HEK	[353]	stabilna	23.5	whole cell PC	4	2; 2	-80	20	-50	step
amsakryna	2	XO	[353]		23.5	voltage-clamp 2- elektrodowy	5	2; 1.6	-80	20	-40	step
apomorfina	2.4	CHO	[143]	stabilna	35	whole cell PC	4.2	0; 0.3	-80	40	-80	protokół AP
artykaina	224	CHO	[317]	przejściowa	23	whole cell PC	5	0; 1	-80	60	-80	ramp
astemizol	0.001	HEK	[411]	stabilna	35	whole cell PC	4	4; 4	-80	0	-50	step
astemizol	0.0013	HEK	[342]	stabilna	35	whole cell PC	4	1; 200	-80	20	-80	step ramp
astemizol	0.0013	HEK	[378]	stabilna	36	whole cell PC	4	1; 1	-80	40	-50	step
astemizol	0.006	CHO	[124]	stabilna	35	whole cell PC	4	1; 0.5	-80	20	-40	step
astemizol	0.018	CHO	[124]	stabilna	23	PatchXpress 7000A	4	1; 0.5	-80	20	-40	step

Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
astemizol	0.0484	XO	[332]		22	voltage-clamp 2- elektrodowy	2	0.5; 0.5	-80	40	-85	step
astemizol	0.069	XO	[36]		23	voltage-clamp 2- elektrodowy	2	4; 6	-80	20	-60	step
astemizol	0.48	XO	[336]		22.5	voltage-clamp 2- elektrodowy	10	1; 1	-90	0	-100	step
atenolol	>1000	HEK	[170]	stabilna	23	whole cell PC	4	4; 6	-80	20	-50	step
AVE0118	10	CHO	[118]	stabilna	23	whole cell PC	5	2; 2	-80	20	-40	step
AZD7009	0.6	CHO	[262]	stabilna	22	whole cell 1-electrode VC	5.4	2; 2	-80	20	-60	step
azimilid	0.56	CHO	[376]	stabilna	37	whole cell PC	4.8	3.9; 5	-80	30	-60	step
azimilid	0.61	CHO	[376]	stabilna	22	whole cell PC	4.8	3.9; 5	-80	30	-60	step
azimilid	0.9	CHO	[262]	stabilna	22	whole cell 1-electrode VC	5.4	2; 2	-80	20	-60	step
azimilid	1.4	XO	[30]		22	voltage-clamp 2- elektrodowy	2	0.5; 0.5	-80	40	-85	step
azimilid	5.2	XO	[30]		22	voltage-clamp 2- elektrodowy	2	0.4; 0.4	-95	40	-80	step
azimilid	5.8	XO	[211]		21.5	voltage-clamp 2- elektrodowy	5	1.2; 1.2	-80	40	-40	step
azimilid	5.8	XO	[212]		21.5	voltage-clamp 2- elektrodowy	5	1.2; 1.2	-80	40	-40	step
azimilid	29.3	XO	[212]		21.5	voltage-clamp 2- elektrodowy	10	1.2; 1.2	-80	40	-40	step
balofloksacyna	214	CHO	[306]	prześciowa	?	whole cell PC	5.4	2; 3	-80	20	-60	step
benzoiloekgonina	>1000	HEK	[101]	stabilna	22	whole cell PC	4	5; 5	-80	60	-60	step
bepriidil	0.023	HEK	[188]	stabilna	22	whole cell PC	4	2; 2	-80	20	-50	step
bepriidil	0.023	HEK	[188]	stabilna	35	whole cell PC	4	1; 200	-80	20	-80	step ramp
bepriidil	0.023	HEK	[188]	stabilna	35	whole cell PC	4	0.5; 0.5	-80	20	-50	step
bepriidil	0.026	HEK	[107]	stabilna	35	whole cell PC	4	1; 120	-80	20	-40	step ramp
bepriidil	0.035	CHO	[124]	stabilna	35	whole cell PC	4	1; 0.5	-80	20	-40	step
bepriidil	0.035	HEK	[188]	stabilna	35	whole cell PC	4	2; 2	-80	20	-50	step
bepriidil	0.099	CHO	[124]	stabilna	23	PatchXpress 7000A	4	1; 0.5	-80	20	-40	step

Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
bepiridil	3	XO	[159]		23	voltage-clamp 2- elektrodowy	2	5; 0	-90	0	-80	step
berberyna	3.1	HEK	[285]	?	22	whole cell PC	4	2; 0	-80	20	-60	step
berberyna	75	XO	[207]		22	voltage-clamp 2- elektrodowy	2	4; 4	-80	0	-60	step
berberyna	80	XO	[285]		23	voltage-clamp 2- elektrodowy	2	4; 0	-80	0	-70	step
bertosamil	62.7	XO	[412]		20	voltage-clamp 2- elektrodowy	5	2; 1.6	-80	20	-40	step
BRL32872	0.241	XO	[348]		21	whole cell PC	5	2; 2	-80	20	-40	step
bromfeniramina	0.9	CHO	[315]	stabilna	?	whole cell PC	5.4	2; 3	-80	20	-60	step
budygina	10.2	XO	[299]		20	voltage-clamp 2- elektrodowy	5	2; 1.6	-80	20	-40	step
bupiwakaina	12	XO	[318]		23	voltage-clamp 2- elektrodowy	2	2; 0	-80	0	0	step
bupiwakaina	18	CHO	[119]	stabilna	21.5	whole cell PC	4	5; 5	-80	-10	-60	step
bupiwakaina	20	CHO	[109]	?	23	whole cell PC	5	0; 1	-80	60	-80	ramp
bupiwakaina	21.9	CHO	[319]	przejściowa	23	whole cell PC	5	0; 1	-80	60	-80	ramp
bupiwakaina	25	XO	[318]		23	voltage-clamp 2- elektrodowy	2	2; 1	-80	0	-40	step
bupiwakaina	43	CHO	[109]	?	23	whole cell PC	5	1; 0.03	-80	60	-80	ramp- step
bupiwakaina	132	XO	[318]		23	voltage-clamp 2- elektrodowy	2	2; 1	-80	0	-120	step
buprenorfina	7.5	HEK	[168]	stabilna	22	whole cell PC	5.4	2; 0	-80	20	-50	step
cetyryzyna	>100	HEK	[58]	stabilna	23	whole cell PC	?	1; 1	-80	40	-50	step
cetyryzyna	>30	XO	[336]		22.5	voltage-clamp 2- elektrodowy	10	1; 1	-90	0	-100	step
chinidyna	0.32	XO	[265]		23	voltage-clamp 2- elektrodowy	4	0.2; 5	-80	50	-70	step
chinidyna	0.33	HEK	[78]	?	22	whole cell PC	4	1; 1	-80	30	-40	step



Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
chinidyna	0.41	HEK	[256]	stabilna	37	whole cell PC	4	0.5; 0.5	-40	30	-40	step
chinidyna	0.547	CHO	[161]	stabilna	23	whole cell PC	5	4; 4	-80	20	-80	step
chinidyna	0.62	CHO	[227]	stabilna	37	whole cell PC	4	2; 4	-80	20	-40	step (prepuls: -40 mV)
chinidyna	0.686	CHO	[124]	stabilna	35	whole cell PC	4	1; 0.5	-80	20	-40	step
chinidyna	0.745	CHO	[124]	stabilna	23	PatchXpress 7000A	4	1; 0.5	-80	20	-40	step
chinidyna	0.82	HEK	[188]	stabilna	35	whole cell PC	4	2; 2	-80	20	-50	step
chinidyna	0.86	HEK	[188]	stabilna	35	whole cell PC	4	0.5; 0.5	-80	20	-50	step
chinidyna	0.89	HEK	[188]	stabilna	22	whole cell PC	4	2; 2	-80	20	-50	step
chinidyna	1.07	HEK	[188]	stabilna	35	whole cell PC	4	1; 200	-80	20	-80	step ramp
chinidyna	2.5	XO	[364]		23	voltage-clamp 2- elektrodowy	3	4; 6	-80	20	-60	step
chinidyna	4.6	XO	[290]		23	voltage-clamp 2- elektrodowy	2	4; 4	-80	0	-70	step
chinidyna	5.2	XO	[290]		23	voltage-clamp 2- elektrodowy	2	4; 4	-80	0	-140	step
chinidyna	8	XO	[205]		21.5	voltage-clamp 2- elektrodowy	2.5	2; 2	-80	30	-60	step
chinidyna	9.1	XO	[212]		21.5	voltage-clamp 2- elektrodowy	5	1.2; 1.2	-80	40	-40	step
chinidyna	9.9	XO	[211]		21.5	voltage-clamp 2- elektrodowy	5	1.2; 1.2	-80	40	-40	step
chinidyna	16.8	XO	[351]		23	voltage-clamp 2- elektrodowy	2	0.025; ?	-80	0	0	step (interpuls: -110 mV)
chinidyna	27.6	XO	[212]		21.5	voltage-clamp 2- elektrodowy	10	1.2; 1.2	-80	40	-40	step
chinidyna	1.5	CHO	[397]	stabilna	25	whole cell PC	4	0.4; 0.4	-80	20	-40	step
chlorfeniramina	13	HEK	[169]	stabilna	22	whole cell PC	5.1	2; 0	-80	20	-50	step
chlorfeniramina	15.6	XO	[135]		21.5	voltage-clamp 2- elektrodowy	2	4; 1	-70	0	-60	step

Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
chlorfeniramina	19.2	XO	[135]		21.5	voltage-clamp 2- elektrodowy	2	4; 1	-70	20	-60	step
chlorfeniramina	20.9	XO	[332]		22	voltage-clamp 2- elektrodowy	2	0.5; 0.5	-80	40	-85	step
chlorfeniramina	25.5	XO	[135]		21.5	voltage-clamp 2- elektrodowy	2	4; 1	-70	40	-60	step
chlorobutanol	0.0044	HEK	[193]	stabilna	?	whole cell PC	?	0.2; 0.45	-80	40	-100	step
chlorochina	2.5	HEK	[362]	stabilna	21	whole cell PC	4	4.8; 5	-80	20	-50	step
chlorochina	8.4	XO	[292]		23	voltage-clamp 2- elektrodowy	2	4; 4	-80	0	-70	step
chlorochina	19.7	XO	[292]		23	voltage-clamp 2- elektrodowy	2	4; 4	-80	0	-140	step
chloroform	9970	HEK	[301]	stabilna	20	whole cell PC	5	2; 1.6	-80	80	-20	step
chloroform	10440	XO	[301]		20	voltage-clamp 2- elektrodowy	5	2; 1.6	-80	20	-40	step
chlorpromazyna	1.47	CHO	[358]	stabilna	21	whole cell PC	4.8	3.9; 5	-80	30	-60	step
chlorpromazyna	1.562	CHO	[181]	przejściowa	35	whole cell PC	5.4	2; 3	-80	20	-60	step
chlorpromazyna	4.9	XO	[203]		22	voltage-clamp 2- elektrodowy	2	4; 0	-70	30	-60	step
chlorpromazyna	8.8	XO	[203]		22	voltage-clamp 2- elektrodowy	2	4; 0	-70	0	-60	step
chlorpromazyna	10.5	XO	[203]		22	voltage-clamp 2- elektrodowy	2	4; 0	-70	-30	-60	step
chlorpromazyna	21.6	XO	[351]		21	voltage-clamp 2- elektrodowy	5	2; 1.6	-80	20	-40	step
chlortalidon	100	XO	[222]		21	voltage-clamp 2- elektrodowy	2	2; 0	-70	50	-70	step
cibenzolina	3.7	HEK	[132]	stabilna	34	whole cell PC	4	2; 3	-80	30	-50	step
ciprofloksacyna	302	CHO	[23]	stabilna	23	whole cell PC	5.4	1; 1	-80	20	-40	step
ciprofloksacyna	996	CHO	[163]	stabilna	23	whole cell PC	5	2; 2	-80	20	-40	step
ciprofloksacyna	>335	HEK	[90]	stabilna	37	whole cell PC	4	0.5; 0.5	-75	10	-40	step
citalopram	3.97	CHO	[386]	przejściowa	20	whole cell PC	4	2; 2	-80	20	-40	step
citalopram	3.97	HEK	[386]	stabilna	36	whole cell PC	?	2; 2	-80	20	-40	step

Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
cizapryd	0.0065	HEK	[236]	stabilna	22.5	whole cell PC	4	10; 5	-80	10	-50	step (prepuls: +60mV)
cizapryd	0.012	HEK	[3]	stabilna	37	whole cell PC	4	0.4; 0.4	-75	10	-40	step
cizapryd	0.0143	HEK	[287]	stabilna	21.5	whole cell PC	4	2.5; 5	-80	10	-50	step (prepuls: +60mV)
cizapryd	0.0144	HEK	[378]	stabilna	36	whole cell PC	4	1; 1	-80	40	-50	step
cizapryd	0.015	CHO	[124]	stabilna	35	whole cell PC	4	1; 120	-80	20	-40	step
cizapryd	0.015	HEK	[107]	stabilna	35	whole cell PC	4	1; 0.5	-80	20	-40	step ramp
cizapryd	0.016	CHO	[374]	stabilna	21	whole cell PC	4.8	3.9; 5	-75	25	-55	step
cizapryd	0.0182	HEK	[224]	stabilna	37	whole cell PC	5	3; 4	-80	0	-50	step
cizapryd	0.02	CHO	[124]	stabilna	23	PatchXpress 7000A	4	1; 0.5	-80	20	-40	step
cizapryd	0.0209	HEK	[342]	stabilna	35	whole cell PC	4	1; 200	-80	20	-80	step ramp
cizapryd	0.023	HEK	[188]	stabilna	35	whole cell PC	4	2; 2	-80	20	-50	step
cizapryd	0.024	CHO	[374]	stabilna	37	whole cell PC	4.8	3.9; 5	-75	25	-55	step
cizapryd	0.026	HEK	[188]	stabilna	22	whole cell PC	4	2; 2	-80	20	-50	step
cizapryd	0.027	HEK	[188]	stabilna	35	whole cell PC	4	1; 200	-80	20	-80	step ramp
cizapryd	0.045	CHO	[258]	?	23	whole cell PC	5	2; 2	-80	20	-40	step
cizapryd	0.079	HEK	[58]	stabilna	23	whole cell PC	?	1; 1	-80	40	-50	step
cizapryd	0.091	CHO	[397]	stabilna	25	whole cell PC	4	0.4; 0.4	-80	20	-40	step
cizapryd	0.133	XO	[32]		23	voltage-clamp 2- elektrodowy	2	5; 0.5	-90	0	-70	step
cyjamemazyna	0.47	HEK	[51]	stabilna	37	whole cell PC	4	0.5; 0.5	-75	10	-40	step
cymetydyna	>10000	HEK	[169]	stabilna	22	whole cell PC	5.1	2; 0	-80	20	-50	step
cząsteczka A	0.0292	CHO	[165]	stabilna	23	whole cell PC	5.77	5; 0.256	-90	0	-140	step
cząsteczka B	0.005	CHO	[165]	stabilna	23	whole cell PC	5.77	5; 0.256	-90	0	-140	step
dekstrometofan	5.1	CHO	[60]	stabilna	21	whole cell PC	5.4	2; 2	-80	20	-40	step
desbutyllumefantryna	5.5	HEK	[362]	stabilna	21	whole cell PC	4	4.8; 5	-80	20	-50	step
desloratadyna	6.31	HEK	[58]	stabilna	23	whole cell PC	?	1; 1	-80	40	-50	step

Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	trans <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
desmetylastemizol	0.001	HEK	[411]	stabilna	35	whole cell PC	4	4; 4	-80	0	-50	step
desmetylerytromycyna	147.1	HEK	[371]	stabilna	35	whole cell PC	4	1; 120	-80	20	-40	step ramp
difenhydramina	2.6	HEK	[188]	stabilna	35	whole cell PC	4	2; 2	-80	20	-50	step
difenhydramina	3.8	HEK	[188]	stabilna	22	whole cell PC	4	2; 2	-80	20	-50	step
difenhydramina	21.5	XO	[338]		23	voltage-clamp 2- elektrodowy	10	2; 2	-90	0	-100	step
difenhydramina	27.1	XO	[332]		22	voltage-clamp 2- elektrodowy	2	0.5; 0.5	-80	40	-85	step
difenhydramina (-)	32.2	XO	[338]		23	voltage-clamp 2- elektrodowy	10	2; 2	-90	0	-100	step
difenhydramina (+)	28.9	XO	[338]		23	voltage-clamp 2- elektrodowy	10	2; 2	-90	0	-100	step
digitoksyna	0.0111	HEK	[379]	stabilna	21	whole cell PC	5	2; 2	-80	60	-50	step
digoksyna	0.0539	HEK	[379]	stabilna	21	whole cell PC	5	2; 2	-80	60	-50	step
diltiazem	6.19	CHO	[124]	stabilna	23	PatchXpress 7000A	4	1; 0.5	-80	20	-40	step
diltiazem	7.39	CHO	[124]	stabilna	35	whole cell PC	4	1; 0.5	-80	20	-40	step
diltiazem	17.3	HEK	[406]	stabilna	23	whole cell PC	4	4; 6	-80	20	-50	step
dipfluzyna	0.00098	XO	[404]			voltage-clamp 2- elektrodowy	5	1; ?	-80	10	-50	step
disopyramid	7.23	CHO	[255]	prześciowa	36	whole cell PC	4	0.5; 1	-40	30	-40	step
disopyramid	10.66	CHO	[227]	stabilna	37	whole cell PC	4	2; 4	-80	20	-40	step (prepuls: -40 mV)
dofetylid	0.003	CHO	[124]	stabilna	23	PatchXpress 7000A	4	1; 0.5	-80	20	-40	step
dofetylid	0.005	CHO	[124]	stabilna	35	whole cell PC	4	1; 0.5	-80	20	-40	step
dofetylid	0.01	CHO	[258]	?	23	whole cell PC	5	2; 2	-80	20	-40	step
dofetylid	0.011	CHO	[162]	stabilna	23	whole cell PC	5	2; 2	-80	20	-40	step
dofetylid	0.029	HEK	[220]	stabilna	23	whole cell PC	4	0.5; 2	-80	60	-40	step (prepuls: -60 mV)
dofetylid	0.035	XO	[174]		20	inside-out macro patches	100	0.4; 0.4	-80	0	-70	step

Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
dofetylid	0.0473	XO	[252]		23	voltage-clamp 2- elektrodowy	2	5; 0.4	-90	0	-70	step
dofetylid	0.073	XO	[364]		23	voltage-clamp 2- elektrodowy	3	4; 6	-80	20	-60	step
dofetylid	0.11	XO	[380]		23	voltage-clamp 2- elektrodowy	2	0.025; 0	-80	0	0	step
dofetylid	0.125	XO	[205]		21.5	voltage-clamp 2- elektrodowy	2.5	2; 2	-80	30	-60	step
dofetylid	0.158	HEK	[58]	stabilna	23	whole cell PC	?	1; 1	-80	40	-50	step
dofetylid	0.192	XO	[65]		19	voltage-clamp 2- elektrodowy	5	2; 2	-70	0	-50	step
dofetylid	0.31	XO	[103]		20	voltage-clamp 2- elektrodowy	5	1.6; 1.6	-80	0	-70	step (prepuls: 0 mV)
dofetylid	0.42	XO	[107]		23	voltage-clamp 2- elektrodowy	?	5; 0	-90	0	-80	step
dofetylid	0.5	XO	[211]		21.5	voltage-clamp 2- elektrodowy	5	1.2; 1.2	-80	40	-40	step
dofetylid	0.595	XO	[174]		20	voltage-clamp 2- elektrodowy	5	0.4; 0.4	-80	0	-70	step
dofetylid	0.644	XO	[174]		20	voltage-clamp 2- elektrodowy	2	0.4; 0.4	-80	0	-70	step
dofetylid	0.692	XO	[174]		20	voltage-clamp 2- elektrodowy	20	0.4; 0.4	-80	0	-70	step
dofetylid	1.5	XO	[212]		21.5	voltage-clamp 2- elektrodowy	5	1.2; 1.2	-80	40	-40	step
dofetylid	2.7	XO	[212]		21.5	voltage-clamp 2- elektrodowy	10	1.2; 1.2	-80	40	-40	step
doksazosyna	0.3244	HEK	[354]	stabilna	21	whole cell PC	4	2; 2	-80	20	-120	step
doksazosyna	0.5851	HEK	[355]	stabilna	23.5	whole cell PC	4	2; 2	-80	20	-50	step
doksazosyna	18.2	XO	[355]		21.5	voltage-clamp 2- elektrodowy	5	2; 1.6	-80	20	-40	step

Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
doksepina	6.5	HEK	[85]	stabilna	37	whole cell PC	4	2; 4	-80	20	-40	step (prepuls: -40 mV)
dolasetron	5.95	HEK	[197]	stabilna	23	whole cell PC	5	2; 2	-80	20	-40	step
domperidon	0.057	HEK	[45]	przejściowa	21	whole cell PC	5.6	2; 2	-80	10	-40	step
domperidon	0.162	CHO	[70]	?	22.5	whole cell PC	4	4; 6	-80	20	-50	step
dronedaron	0.0591	HEK	[282]	stabilna	37	whole cell PC	4	1; 4	-40	30	-40	step
dronedaron	9.2	XO	[350]		21	voltage-clamp 2- elektrodowy	5	0.4; 0.4	-80	30	-60	step
droperidol	0.0322	HEK	[69]	stabilna	22.5	whole cell PC	4	4; 5.7	-80	20	-50	step
droperidol	0.0773	HEK	[221]		23	whole cell PC	5	4; 4	-80	50	-50	step
droperidol	0.098	HEK	[373]	stabilna	21	whole cell PC	4	1; 0.5	-80	20	-120	step
droperidol	0.307	CHO	[397]	stabilna	25	whole cell PC	4	0.4; 0.4	-80	20	-40	step
DW286a	89	CHO	[180]	przejściowa	35	whole cell PC	5.4	2; 3	-80	20	-60	step
E4031	0.0077	HEK	[410]	stabilna	35	whole cell PC	4	4; 5	-80	0	-50	step
E4031	0.00998	CHO	[180]	przejściowa	35	whole cell PC	5.4	2; 3	-80	20	-60	step
E4031	0.011	HEK	[107]	stabilna	35	whole cell PC	4	1; 120	-80	20	-40	step ramp
E4031	0.012	HEK	[188]	stabilna	35	whole cell PC	4	1; 200	-80	20	-80	step ramp
E4031	0.014	HEK	[188]	stabilna	22	whole cell PC	4	2; 2	-80	20	-50	step
E4031	0.01596	CHO	[227]	stabilna	37	whole cell PC	4	2; 4	-80	20	-40	step (prepuls: -40 mV)
E4031	0.016	HEK	[74]	stabilna	22	whole cell PC	4	4.8; 5	-80	20	-50	step
E4031	0.017	HEK	[188]	stabilna	35	whole cell PC	4	2; 2	-80	20	-50	step
E4031	0.0181	HEK	[51]	stabilna	37	whole cell PC	4	0.5; 0.5	-75	10	-40	step
E4031	0.0264	HEK	[200]	stabilna	37	whole cell PC	4	0.5; 0.5	-75	10	-40	step
E4031	0.032	HEK	[373]	stabilna	21	whole cell PC	4	1; 0.5	-80	20	-120	step
E4031	0.048	HEK	[224]	stabilna	37	whole cell PC	5	3; 4	-80	0	-50	step
E4031	0.079	HEK	[58]	stabilna	23	whole cell PC	?	1; 1	-80	40	-50	step
E4031	0.134	CHO	[375]	stabilna	21	whole cell PC	4.8	3.9; 3.9	-55	25	-35	step

Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
E4031	0.17	XO	[380]		23	voltage-clamp 2- elektrodowy	2	0.025; 0	-80	0	0	step (interpuls)
E4031	0.57	XO	[107]		23	voltage-clamp 2- elektrodowy	?	5; 0	-90	0	-80	step
ebastyna	0.3	XO	[189]		20	voltage-clamp 2- elektrodowy	2	1.2; 1	-80	30	-80	step
EDDP	>50	HEK	[168]	stabilna	22	whole cell PC	5.4	2; 0	-80	20	-50	step
EGCG	6	HEK	[171]	stabilna		whole cell PC	5	0.4; 0.4	-80		-120	step
EGCG	20.5	XO	[171]			voltage-clamp 2- elektrodowy	5	0.4; 0.4	-80		-60	step
epinastyna	>100	XO	[36]		23	voltage-clamp 2- elektrodowy	2	4; 6	-80	20	-60	step
erytromycylamina	273.9	HEK	[371]	stabilna	35	whole cell PC	4	1; 120	-80	20	-40	step ramp
erytromycyna	38.9	HEK	[325]	stabilna	36	whole cell PC	4	0.5; 0.5	-80	10	-40	step
erytromycyna	59.3	HEK	[84]	stabilna	37	whole cell PC	4	2; 2	-80	20	-40	step (prepuls: -40 mV)
erytromycyna	72.2	HEK	[371]	stabilna	35	whole cell PC	4	1; 120	-80	20	-40	step ramp
erytromycyna	101	HEK	[219]	stabilna	23	whole cell PC	4	2; 2	-80	60	-40	step (prepuls: -60 mV)
erytromycyna	115	HEK	[188]	stabilna	35	whole cell PC	4	1; 200	-80	20	-80	step ramp
erytromycyna	158	HEK	[224]	stabilna	37	whole cell PC	5	3; 4	-80	0	-50	step
erytromycyna	199	HEK	[188]	stabilna	35	whole cell PC	4	2; 2	-80	20	-50	step
erytromycyna	1410	HEK	[188]	stabilna	22	whole cell PC	4	2; 2	-80	20	-50	step
estradiol	>30	CHO	[238]		21	whole cell PC	5.4	1; 1	-80	20	-40	step
feksofenadyna	23	CHO	[258]	?	23	whole cell PC	5	2; 2	-80	20	-40	step
feksofenadyna	65.1	HEK	[270]	stabilna	22.5	whole cell PC	4	4; 5.7	-80	20	-50	step
feksofenadyna	>100	XO	[297]		23	voltage-clamp 2- elektrodowy	2	1; 1	-80	40	-80	step

Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
fenobarbital	3000	HEK	[54]	stabilna	22.5	whole cell PC	4	4.8; 4.8	-80	20	-50	step
fentanyl	1.8	HEK	[168]	stabilna	22	whole cell PC	5.4	2; 0	-80	20	-50	step
fenytoina	240	HEK	[54]	stabilna	22.5	whole cell PC	4	4.8; 4.8	-80	20	-50	step
flekainid	1.03	HEK	[78]	?	22	whole cell PC	4	1; 1	-80	30	-40	step
flekainid	3.91	HEK	[256]	stabilna	37	whole cell PC	4	0.5; 0.5	-40	30	-40	step
flunaryzyna	0.0057	CHO	[363]	stabilna	35	whole cell PC	4	1; 2	-80	20	-50	step
fluoksetyna	0.46	HEK	[169]	stabilna	35	whole cell PC	4	1; 200	-80	20	-80	step ramp
fluoksetyna	0.5	CHO	[124]	stabilna	35	whole cell PC	4	1; 0.5	-80	20	-40	step
fluoksetyna	0.5	HEK	[188]	stabilna	35	whole cell PC	4	2; 2	-80	20	-50	step
fluoksetyna	0.55	HEK	[188]	stabilna	35	whole cell PC	4	0.5; 0.5	-80	20	-50	step
fluoksetyna	0.7	HEK	[267]	stabilna	23	whole cell PC	4	4; 5.7	-80	70	-50	step
fluoksetyna	0.7	HEK	[267]	stabilna	37	whole cell PC	4	0; 0.38	-80	40	-80	protokół AP
fluoksetyna	0.71	HEK	[188]	stabilna	22	whole cell PC	?	2; 2	-80	20	-50	step
fluoksetyna	0.72	CHO	[124]	stabilna	23	PatchXpress 7000A	4	1; 0.5	-80	20	-40	step
fluoksetyna	0.99	HEK	[224]	stabilna	37	whole cell PC	5	3; 4	-80	0	-50	step
fluoksetyna	1.5	CHO	[386]	przejściowa	20	whole cell PC	4	2; 2	-80	20	-40	step
fluoksetyna	1.5	HEK	[386]	stabilna	36	whole cell PC	?	2; 2	-80	20	-40	step
fluspirylen	2.34	XO	[316]		23	voltage-clamp 2- elektrodowy	2	?; ?	-80	20	-55	step
fluwoksamina	3.8	HEK	[233]	stabilna	37	whole cell PC	4	2; 4	-80	20	-40	step
gabapentyna	>100	HEK	[55]	stabilna	23	whole cell PC	4	4.8; 4.8	-80	20	-50	step
gambierol	>1	XO	[53]		23	voltage-clamp 2- elektrodowy	2	?; ?	-90	40	-90	step
gatifloksacyna	130	CHO	[163]	stabilna	23	whole cell PC	5	2; 2	-80	20	-40	step
gemifloksacyna	300	CHO	[306]	przejściowa	?	whole cell PC	5.4	2; 3	-80	20	-60	step
GF109203X (bisindolylmaleimid I)	1	HEK	[352]	stabilna	23	whole cell PC	4	2; 2	-80	20	-50	step
GF109203X (bisindolylmaleimid I)	13.2	XO	[352]		23	voltage-clamp 2- elektrodowy	5	2; 1.6	-80	20	-40	step
granisetron	3.73	HEK	[197]	stabilna	23	whole cell PC	5	2; 2	-80	20	-40	step



Substancja	IC50 [μM]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
grepafloksacyna	50	CHO	[163]	stabilna	23	whole cell PC	5	2; 2	-80	20	-40	step
grepafloksacyna	76.5	CHO	[23]	stabilna	23	whole cell PC	5.4	1; 1	-80	20	-40	step
halofantryna	0.0216	HEK	[225]	stabilna	23	whole cell PC	4	4; 5.7	-80	20	-50	step
halofantryna	0.04	HEK	[362]	stabilna	21	whole cell PC	4	4.8; 5	-80	20	-50	step
halofantryna	0.1969	CHO	[358]	stabilna	21	whole cell PC	4.8	3.9; 10	-80	30	-60	step
halofantryna	1	XO	[291]		23	voltage-clamp 2- elektrodowy	2	4; 4	-80	0	-70	step
haloperidol	0.015	CHO	[397]	stabilna	25	whole cell PC	4	0.4; 0.4	-80	20	-40	step
haloperidol	0.019	HEK	[188]	stabilna	35	whole cell PC	4	1; 200	-80	20	-80	step ramp
haloperidol	0.025	HEK	[188]	stabilna	22	whole cell PC	4	2; 2	-80	20	-50	step
haloperidol	0.026	HEK	[188]	stabilna	35	whole cell PC	4	2; 2	-80	20	-50	step
haloperidol	0.027	CHO	[124]	stabilna	35	whole cell PC	4	1; 0.5	-80	20	-40	step
haloperidol	0.0281	HEK	[33]	stabilna	37	whole cell PC	4	0.4; 0.4	-75	10	-40	step
haloperidol	0.032	CHO	[258]	?	23	whole cell PC	5	2; 2	-80	20	-40	step
haloperidol	0.042	CHO	[124]	stabilna	23	PatchXpress 7000A	4	1; 0.5	-80	20	-40	step
haloperidol	0.063	HEK	[169]	stabilna	22	whole cell PC	5.1	2; 0	-80	20	-50	step
haloperidol	0.174	HEK	[224]	stabilna	37	whole cell PC	5	3; 4	-80	0	-50	step
haloperidol	1.1	XO	[331]		22	voltage-clamp 2- elektrodowy	2	0.5; 2?	-80	40	-70	step
haloperidol	1.36	XO	[316]		23	voltage-clamp 2- elektrodowy	2	?; ?	-80	20	-55	step
haloperidol	2.2	XO	[99,331]		22	voltage-clamp 2- elektrodowy	2	0.5; 2?	-80	0	-70	step
hesperetyna	288.8	XO	[414]		23	voltage-clamp 2- elektrodowy	?	0.4; 0.4	-80	30	-60	step
hydroksyhaloperidol	2.6	XO	[99,331]		22	voltage-clamp 2- elektrodowy	2	0.5; 2?	-80	40	-70	step
hydroksyzyna	10.7	XO	[338]		23	voltage-clamp 2- elektrodowy	10	2; 2	-90	0	-100	step
ibutilid	0.028	XO	[261]		23	voltage-clamp 2- elektrodowy	2	5; 0.4	-90	0	-70	step

Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
ibutilid	0.9	XO	[213]		21.5	voltage-clamp 2- elektrodowy	5	1.2; 1.2	-80	40	-40	step
ibutilid	2	XO	[213]		21.5	voltage-clamp 2- elektrodowy	10	1.2; 1.2	-80	40	-40	step
ICI118551	9.2	HEK	[170]	stabilna	23	whole cell PC	4	4; 6	-80	20	-50	step
imipramina	1.212	CHO	[124]	stabilna	35	whole cell PC	4	1; 0.5	-80	20	-40	step
imipramina	1.52	CHO	[124]	stabilna	23	PatchXpress 7000A	4	1; 0.5	-80	20	-40	step
imipramina	3.4	CHO	[344]	przejściowa	23	?perforated PC	4	0.4; 0.4	-80	20	-40	step
indometacyna	>300	HEK	[224]	stabilna	37	whole cell PC	5	3; 4	-80	0	-50	step
irbesartan	193	CHO	[239]	stabilna	21.5	whole cell PC	4	5; 5	-80	60	-60	step
josamycyna	102.4	HEK	[371]	stabilna	35	whole cell PC	4	1; 120	-80	20	-40	step ramp
kapsaicyna	17.45	XO	[395]		22	voltage-clamp 2- elektrodowy	2	2; 3	-90	0	-60	step
kapsaicyna	29.15	XO	[395]		22	voltage-clamp 2- elektrodowy	30	2; 4	-90	0	-90	step
kaptopril	>3000	CHO	[397]	stabilna	25	whole cell PC	4	0.4; 0.4	-80	20	-40	step
karbamazepina	>500	HEK	[54]	stabilna	22.5	whole cell PC	4	4.8; 4.8	-80	20	-50	step
karwedilol	0.51	HEK	[170]	stabilna	23	whole cell PC	4	4; 6	-80	20	-50	step
karwedilol	10.4	XO	[166]		20	voltage-clamp 2- elektrodowy	5	0.4; 0.4	-80	30	-60	step
ketanserin	0.128	HEK	[188]	stabilna	22	whole cell PC	4	2; 2	-80	20	-50	step
ketanseryna	0.107	HEK	[188]	stabilna	35	whole cell PC	4	1; 200	-80	20	-80	step ramp
ketanseryna	0.11	HEK	[340]	stabilna	22.5	whole cell PC	5	1; 0	-80	30	-50	step
ketanseryna	0.121	HEK	[188]	stabilna	35	whole cell PC	4	2; 2	-80	20	-50	step
ketanseryna	0.31	HEK	[340]	stabilna	22.5	whole cell PC	20	1; 0	-80	30	-50	step
ketokonazol	1.7	HEK	[280]	stabilna	37	whole cell PC	4	2; 4	-80	20	-40	step
ketokonazol	1.92	HEK	[339]	stabilna	23	whole cell 1-electrode VC	4	4; 6	-80	20	-50	step
ketokonazol	31.2	XO	[82]		23	voltage-clamp 2- elektrodowy	2.5	0.8; 0.8	-80	0	-80	step
klarytromycyna	32.9	HEK	[371]	stabilna	35	whole cell PC	4	1; 120	-80	20	-40	step ramp

Substancja	IC50 [μM]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
klarytromycyna	45.7	HEK	[325]	stabilna	36	whole cell PC	4	0.5; 0.5	-80	10	-40	step
klarytromycyna	170	HEK	[287]	stabilna	21.5	whole cell PC	4	0.5; 1	-60	60	-40	step
klarytromycyna	750	CHO	[1]	?	?	whole cell PC	?	1; 2	-80	20	-40	step
klebopryde	0.62	CHO	[182]	stabilna	?	whole cell PC	5.4	2; 3	-80	20	-40	step
klemastyna	0.012	HEK	[281]	stabilna	37	whole cell PC	4	2; 4	-80	20	-40	step (prepuls: -40mV)
klobutinol	1.9	CHO	[60]	stabilna	21	whole cell PC	5.4	2; 2	-80	20	-40	step
klomifen	0.18	HEK	[403]	stabilna	36	whole cell PC	4	2; 3	-40	30	-40	step
klomipramina	0.13	HEK	[152]	stabilna	36	whole cell PC	5.4	4.8; 0	-80	20	-60	step
klomipramina	9.94	XO	[152]		21.5	voltage-clamp 2- elektrodowy	2	4; 0	-80	40	-60	step
klomipramina	19.5	XO	[152]		21.5	voltage-clamp 2- elektrodowy	2	4; 0	-80	0	-60	step
klomipramina	31.7	XO	[152]		21.5	voltage-clamp 2- elektrodowy	2	4; 0	-80	-40	-60	step
klotrimazol	3.6	HEK	[357]	prześciowa	21.5	? Whole or perforated	5	1; 1	-80	30	-50	step
klozapina	2.5	HEK	[204]	stabilna	35	whole cell PC	5.4	4.8; 4.8	-80	20	-60	step
klozapina	2.63	CHO	[358]	stabilna	21	whole cell PC	4.8	3.9; 10	-80	30	-60	step
klozapina	22.9	XO	[204]		21.5	voltage-clamp 2- elektrodowy	2	4; 4	-70	40	-60	step
klozapina	28.3	XO	[204]		21.5	voltage-clamp 2- elektrodowy	2	4; 4	-70	0	-60	step
klozapina	39.9	XO	[204]		21.5	voltage-clamp 2- elektrodowy	2	4; 4	-70	-40	-60	step
kodeina	97	CHO	[60]	stabilna	21	whole cell PC	5.4	2; 2	-80	20	-40	step
kodeina	>300	HEK	[168]	stabilna	22	whole cell PC	5.4	2; 0	-80	20	-50	step
kokaetylen	1.2	HEK	[101]	stabilna	22	whole cell PC	4	5; 12	-80	60	-60	step
kokaina	4.4	HEK	[101]	stabilna	22	whole cell PC	4	5; 12	-80	60	-60	step
kokaina	7.2	HEK	[407]	stabilna	23	whole cell PC	4	4; 5.7	-80	20	-50	step
kokaina	11.1	HEK	[123]	stabilna	23	whole cell PC	5	5; 5	-80	50	-50	step
kwac acetylosalicylowy	>100000	CHO	[397]	stabilna	25	whole cell PC	4	0.4; 0.4	-80	20	-40	step
kwac kanrenowy	104.2	CHO	[31]	stabilna	24.5	whole cell PC	4	5; 5	-80	60	-60	step

Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
kwetiapina	5.765	CHO	[191]	?	23	whole cell PC	5	2; 2	-80	20	-40	step
L-702,958	0.0143	CHO	[165]	stabilna	23	whole cell PC	5.33	5; 0.256	-90	0	-140	step
laam	2.2	HEK	[168]	stabilna	22	whole cell PC	5.4	2; 0	-80	20	-50	step
lamotrygina	229.1	HEK	[55]	stabilna	23	whole cell PC	4	4.8; 4.8	-80	20	-50	step
lewokromakalim	>10	HEK	[220]	stabilna	23	whole cell PC	4	0.5; 2	-80	60	-40	step (prepuls: -60 mV)
lewobupiwakaina	10	CHO	[109]	?	23	whole cell PC	5	0; 1	-80	60	-80	ramp
lewobupiwakaina	10.2	CHO	[119]	stabilna	21.5	whole cell PC	4	5; 5	-80	-10	-60	step
lewobupiwakaina	12.7	CHO	[319]	przejściowa	23	whole cell PC	5	0; 1	-80	60	-80	ramp
lewobupiwakaina	32	CHO	[109]	?	23	whole cell PC	5	1; 0.03	-80	60	-80	ramp- step
lewofloksacyna	430	HEK	[4]	stabilna	22	whole cell PC	4	1; 200	-80	20	-80	step ramp
lewofloksacyna	915	CHO	[163]	stabilna	23	whole cell PC	5	2; 2	-80	20	-40	step
lidoflazyna	0.0166	HEK	[279]	stabilna	37	whole cell PC	4	2; 4	-80	20	-40	step
lidoflazyna	0.037	HEK	[169]	stabilna	22	whole cell PC	5.1	2; 0	-80	20	-50	step
lidokaina	262.9	HEK	[169]	stabilna	37	whole cell PC	4	0.5; 0.5	-40	30	-40	step
lomefloksacyna	2400	HEK	[415]	przejściowa	21	whole cell PC	5.6	2; 2	-80	10	-40	step
loratadyna	0.173	HEK	[50]	stabilna	36	whole cell PC	4	0.4; 0.4	-75	10	-40	step
loratadyna	2.3	HEK	[188]	stabilna	35	whole cell PC	4	1; 200	-80	20	-80	step ramp
loratadyna	3.9	HEK	[188]	stabilna	22	whole cell PC	4	2; 2	-80	20	-50	step
loratadyna	4	HEK	[169]	stabilna	22	whole cell PC	5.1	2; 0	-80	20	-50	step
loratadyna	7.3	HEK	[188]	stabilna	35	whole cell PC	4	2; 2	-80	20	-50	step
loratadyna	7.94	HEK	[58]	stabilna	23	whole cell PC	?	1; 1	-80	40	-50	step
loratadyna	101	XO	[336]		22.5	voltage-clamp 2- elektrodowy	10	1; 1	-90	0	-100	step
lowastatyna	7	HEK	[169]	stabilna	22	whole cell PC	5.1	2; 0	-80	20	-50	step
lumefantryna	8.1	HEK	[362]	stabilna	21	whole cell PC	4	4.8; 5	-80	20	-50	step
maprotylina	0.13	HEK	[150]	stabilna	36	whole cell PC	5.4	4.8; 4.8	-80	20	-60	step
maprotylina	5.2	HEK	[102]	stabilna	23	whole cell PC	4	2; 2	-80	0	-40	step

Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
maprotylina	8.2	HEK	[176]	stabilna	20	whole cell PC	5	0.4; 0.4	-80	80	-60	step
maprotylina	23.7	XO	[102]		23	voltage-clamp 2- elektrodowy	2	4; 4	-80	0	-70	step
maprotylina	29.2	XO	[176]		20	voltage-clamp 2- elektrodowy	5	0.4; 0.4	-80	80	-60	step
maprotylina	39.5	XO	[150]		21.5	voltage-clamp 2- elektrodowy	2	4; 2	-70	-40	-60	step
maprotylina	43.6	XO	[150]		21.5	voltage-clamp 2- elektrodowy	2	4; ?2	-70	40	-60	step
MCI-154	41.55	XO	[150]		21.5	voltage-clamp 2- elektrodowy	2	4; 4	-70	40	-60	step
MDL74156	12.1	HEK	[197]	stabilna	23	whole cell PC	5	2; 2	-80	20	-40	step
meflochina	2.6	HEK	[362]	stabilna	21	whole cell PC	4	4.8; 5	-80	20	-50	step
meflochina	5.6	CHO	[161]	stabilna	23	whole cell PC	5	4; 4	-80	20	-80	step
meperydyna	75	HEK	[168]	stabilna	22	whole cell PC	5.4	2; 0	-80	20	-50	step
mepiwakaina	156.2	CHO	[319]	przejściowa	23	whole cell PC	5	0; 1	-80	60	-80	ramp
metadon	9.8	HEK	[168]	stabilna	22	whole cell PC	5.4	2; 0	-80	20	-50	step
metadon	20	HEK	[193]	stabilna	?	whole cell PC	?	0.2; 0.45	-80	40	-100	step
metadon (RS)	0.21	XO	[210]			voltage-clamp 2- elektrodowy		?; ?				step
metadon (RS)	2	HEK	[95]			Qpatch	4	2; 2	-85	20	-50	step
metadon(R)	7	HEK	[88]	stabilna	37	whole cell PC	5	2.5; 2.5	-80	40	-50	step
metadon(R)	29	HEK	[88]	stabilna	25	whole cell PC	5	2.5; 2.5	-80	40	-50	step
metadon(RS)	3	HEK	[88]	stabilna	37	whole cell PC	5	2.5; 2.5	-80	40	-50	step
metadon(RS)	19	HEK	[88]	stabilna	25	whole cell PC	5	2.5; 2.5	-80	40	-50	step
metadon(S)	2	HEK	[88]	stabilna	37	whole cell PC	5	2.5; 2.5	-80	40	-50	step
metadon(S)	12	HEK	[88]	stabilna	25	whole cell PC	5	2.5; 2.5	-80	40	-50	step
metoklopramid	5.4	HEK	[45]	przejściowa	21	whole cell PC	5.6	2; 2	-80	10	-40	step
metoprolol	145	HEK	[170]	stabilna	23	whole cell PC	4	4; 6	-80	20	-50	step
metylekgonidyna	171.7	HEK	[101]	stabilna	22	whole cell PC	4	5; 5	-80	60	-60	step
metyloester ekgoniny	>1000	HEK	[101]	stabilna	22	whole cell PC	4	5; 5	-80	60	-60	step
mezorydazyna	0.55	HEK	[329]	stabilna	37	whole cell PC	5	3; 3	-80	0	-50	step

Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
mianseryna	14.78	CHO	[358]	stabilna	21	whole cell PC	4.8	3.9; 10	-80	30	-60	step
mikonazol	2.1	HEK	[178]	stabilna	23	whole cell PC	4	4; 6	-80	20	-50	step
misolastyna	0.35	HEK	[337]	stabilna	22.5	whole cell PC	10	1; 1	-80	?	-100	step
mizolastyna	3.4	XO	[337]		22.5	voltage-clamp 2- elektrodowy	10	1; 0.4	-90	0	-100	step
MK499	0.0216	HEK	[387]	stabilna	36	whole cell PC	4	1; 1	-80	40	-50	step
MK499	0.034	XO	[32]		23	voltage-clamp 2- elektrodowy	2	5; 0.5	-90	0	-70	step
MK499	0.0898	CHO	[165]	stabilna	23	whole cell PC	5.33	5; 0.256	-90	0	-140	step
MK499	0.12	XO	[32]		23	voltage-clamp 2- elektrodowy	96	5; 0.5	-90	0	-70	step
moksifloksacyna	29	HEK	[4]	stabilna	36	whole cell PC	4	1; 200	-80	20	-80	step ramp
moksifloksacyna	35.7	HEK	[39]	stabilna	35	whole cell PC	4	1; 200	-80	20	-80	step ramp
moksifloksacyna	58.5	HEK	[224]	stabilna	37	whole cell PC	5	3; 4	-80	0	-50	step
moksifloksacyna	61	HEK	[4]	stabilna	?	whole cell PC	4	2; 4	-80	20	-40	step
moksifloksacyna	65	HEK	[4]	stabilna	22	whole cell PC	4	1; 200	-80	20	-80	step ramp
moksifloksacyna	74.7	HEK	[200]	stabilna	37	whole cell PC	4	0.5; 0.5	-75	10	-40	step
moksifloksacyna	99	HEK	[219]	stabilna	23	whole cell PC	4	2; 2	-80	60	-40	step (prepuls: -60 mV)
moksifloksacyna	102.6	CHO	[23]	stabilna	23	whole cell PC	5.4	1; 1	-80	20	-40	step
moksifloksacyna	129	CHO	[163]	stabilna	23	whole cell PC	5	2; 2	-80	20	-40	step
moksifloksacyna	235	CHO	[397]	stabilna	25	whole cell PC	4	0.4; 0.4	-80	20	-40	step
morfina	>1000	HEK	[168]	stabilna	22	whole cell PC	5.4	2; 0	-80	20	-50	step
moryna	111.4	XO	[414]		23	voltage-clamp 2- elektrodowy	?	0.4; 0.4	-80	30	-60	step
naryngenina	102.3	XO	[414]		23	voltage-clamp 2- elektrodowy	?	0.4; 0.4	-80	30	-60	step
naryngenina	102.6	XO	[300]		20	voltage-clamp 2- elektrodowy	5	0.4; 0.4	-80	30	-60	step

Substancja	IC50 [μM]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
naryngenina	173.3	XO	[214]		21.5	voltage-clamp 2-elektrodowy	5	1.2; 1.2	-80	40	-40	step
n-desbutylhalofantryna	0.0717	HEK	[225]	stabilna	23	whole cell PC	4	4; 5.7	-80	20	-50	step
neferyna	7.419	HEK	[121]	stabilna	23	whole cell PC	5.4	4; 2	-80	0	-50	step
nifedypina	>50	HEK	[406]	stabilna	23	whole cell PC	4	4; 6	-80	20	-50	step
nifekalant	5.8	XO	[107]		23	voltage-clamp 2-elektrodowy	?	5; 0	-90	0	-80	step
nifekalant	7.9	XO	[199]		23	voltage-clamp 2-elektrodowy	2	3.6; 3.6	-80	40	-60	step
nikorandyl	>100	HEK	[220]	stabilna	23	whole cell PC	4	0.5; 2	-80	60	-40	step (prepuls: -60 mV)
nikotyna	16.6	XO	[377]		22.5	voltage-clamp 2-elektrodowy	5	2; 2	-70	50	-50	step
norastemizol	0.0277	HEK	[411]	stabilna	35	whole cell PC	4	4; 4	-80	0	-50	step
norfluoksetyna	2.3	HEK	[268]	stabilna	23	whole cell PC	4	4; 5.7	-80	70	-50	step
norfluoksetyna	2.5	HEK	[268]	stabilna	37	whole cell PC	4	0; 0.38	-82	42	-82	protokół AP
norpropoksyfen	33.2	XO	[366]		21	voltage-clamp 2-elektrodowy	2	2; 2	-70	-10	-70	step
ofloksacyna	1420	CHO	[163]	stabilna	23	whole cell PC	5	2; 2	-80	20	-40	step
oksykodon	171	HEK	[95]			Qpatch	4	2; 2	-85	20	-50	step
oksykodon	>30	XO	[95]		23	voltage-clamp 2-elektrodowy	4	1; 1.5	-80	40	-60	step
olanzapina	1.78	CHO	[124]	stabilna	23	PatchXpress 7000A	4	1; 0.5	-80	20	-40	step
olanzapina	2.01	CHO	[124]	stabilna	35	whole cell PC	4	1; 0.5	-80	20	-40	step
olanzapina	6.013	CHO	[191]	?	23	whole cell PC	5	2; 2	-80	20	-40	step
oleandomycyna	339.6	HEK	[371]	stabilna	35	whole cell PC	4	1; 120	-80	20	-40	step ramp
ondasetron	0.81	HEK	[197]	stabilna	23	whole cell PC	5	2; 2	-80	20	-40	step
orfenadryna	0.85	HEK	[302]	stabilna	22	whole cell PC	5	0.4; 0.4	-80	30	-120	step
orfenadryna	6.7	XO	[302]		22	voltage-clamp 2-elektrodowy	5	0.4; 0.4	-80	30	-120	step

Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
orfenadryna	13.1	XO	[338]		23	voltage-clamp 2- elektrodowy	10	2; 2	-90	0	-100	step
ouabaina	0.0152	HEK	[379]	stabilna	21	whole cell PC	5	2; 2	-80	60	-50	step
papaweryna	4.43	HEK	[187]		36	voltage-clamp 2- elektrodowy	5.4	4.8; 5	-80	20	-60	step
papaweryna	24.8	XO	[187]		22	voltage-clamp 2- elektrodowy	2	4; 0	-70	40	-60	step
papaweryna	30	XO	[187]		22	voltage-clamp 2- elektrodowy	2	4; 0	-70	20	-60	step
papaweryna	38.8	XO	[187]		22	voltage-clamp 2- elektrodowy	2	4; 0	-70	-10	-60	step
papaweryna	71.03	XO	[179]		23	voltage-clamp 2- elektrodowy	2	0.5; 0.5	-80	0	-40	step
pentamidyna	5.1	HEK	[198]	stabilna	21	whole cell PC	5	2; 1.5	-80	20	-40	step
pentamidyna	>1000	HEK	[169]	stabilna	22	whole cell PC	5.1	2; 0	-80	20	-50	step
pentamidyna	252	HEK	[49]	stabilna	35	whole cell PC	4	1; 200	-80	20	-80	step ramp
pentobarbital	1580	XO	[15]		23	voltage-clamp 2- elektrodowy	2	0.25; 0.6	-80	40	-40	step
pentoksyweryna	3	CHO	[60]	stabilna	21	whole cell PC	5.4	2; 2	-80	20	-40	step
perfenazyna	1.003	CHO	[181]	przejściowa	35	whole cell PC	5.4	2; 3	-80	20	-60	step
pergolid	0.12	CHO	[143]	stabilna	35	whole cell PC	4.2	0; 0.3	-80	40	-80	protokół AP
perheksylina	7.8	CHO	[375]	stabilna	21	whole cell PC	4.8	3.9; 3.9	-55	25	-35	step
pilsikainid	20.4	HEK	[391]	stabilna	34	whole cell PC	4	4; 6	-60	50	-50	step
pimozyd	0.001	HEK	[188]	stabilna	22	whole cell PC	4	2; 2	-80	20	-50	step
pimozyd	0.001	HEK	[188]	stabilna	35	whole cell PC	4	1; 200	-80	20	-80	step ramp
pimozyd	0.0015	HEK	[188]	stabilna	35	whole cell PC	4	2; 2	-80	20	-50	step
pimozyd	0.0026	HEK	[188]	stabilna	35	whole cell PC	4	0.5; 0.5	-80	20	-50	step
pimozyd	0.018	CHO	[164]	?	23	whole cell PC	5	2; 2	-80	20	-40	step
pimozyd	0.018	CHO	[191]	?	23	whole cell PC	5	2; 2	-80	20	-40	step
pimozyd	0.018	CHO	[258]	?	23	whole cell PC	5	2; 2	-80	20	-40	step



Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
pimozyd	1.74	XO	[315]		23	voltage-clamp 2- elektrodowy	2	?; ?	-80	20	-55	step
prazosyna	1.57	HEK	[355]	stabilna	23.5	whole cell PC	4	2; 2	-80	20	-50	step
prazosyna	10.1	XO	[355]		21.5	voltage-clamp 2- elektrodowy	5	2; 1.6	-80	20	-40	step
prenylamina	0.59	HEK	[169]	stabilna	22	whole cell PC	5.1	2; 0	-80	20	-50	step
prochloroperazyna	37.4	XO	[186]		21.5	voltage-clamp 2- elektrodowy	2	4; 4	-70	0	-60	step
prochlorperazyna	22.6	XO	[186]		21.5	voltage-clamp 2- elektrodowy	2	4; 4	-70	40	-60	step
prochlorperazyna	42.1	XO	[186]		21.5	voltage-clamp 2- elektrodowy	2	4; 4	-70	-40	-60	step
prokainamid	138.7	HEK	[277]	stabilna	37	whole cell PC	4	2; 4	-80	20	-40	step
prometazyna	1.46	HEK	[154]		36	whole cell PC	5.4	4.8; 0	-80	20	-60	step
prometazyna	35	XO	[154]		21.5	voltage-clamp 2- elektrodowy	2	4; 2	-70	40	-60	step
prometazyna	41.9	XO	[154]		21.5	voltage-clamp 2- elektrodowy	2	4; 2	-70	20	-60	step
prometazyna	52.1	XO	[154]		21.5	voltage-clamp 2- elektrodowy	2	4; 2	-70	-20	-60	step
propafenon	0.44	HEK	[256]	stabilna	37	whole cell PC	4	0.5; 0.5	-40	30	-40	step
propafenon	0.55	CHO	[9]	stabilna	21.5	whole cell PC	4	5; 5	-80	-10	-60	step
propafenon	2	HEK	[169]	stabilna	22	whole cell PC	5.1	2; 0	-80	20	-50	step
propafenon	14	XO	[230]		22	voltage-clamp 2- elektrodowy	2	?; ?	-80	20	-80	step
propiweryna	10	HEK	[42]	stabilna	35	whole cell PC	4	5; 5	-80	0	-50	step
propoksyfen	44.7	XO	[366]		21	voltage-clamp 2- elektrodowy	2	2; 2	-70	-10	-70	step
propranolol	3.9	HEK	[170]	stabilna	23	whole cell PC	4	4; 6	-80	20	-50	step
propranolol	9.9	CHO	[398]	stabilna	25	whole cell PC	4	0.4; 0.4	-80	20	-40	step
propranolol	81.1	XO	[86]		23	whole cell PC	?	2; 2	-80	-10	-60	step
protryptylina	1.18	HEK	[153]	stabilna	36	whole cell PC	5.4	4; 0	-80	20	-60	step

Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
protryptylina	52.9	XO	[153]		21.5	voltage-clamp 2- elektrodowy	2	4; 0	-70	40	-60	step
protryptylina	91.7	XO	[153]		21.5	voltage-clamp 2- elektrodowy	2	4; 0	-70	0	-60	step
protryptylina	142	XO	[153]		21.5	voltage-clamp 2- elektrodowy	2	4; 0	-70	-40	-60	step
prukalopryd	4.1	HEK	[37]	przejściowa	23	whole cell PC	5.4	2; 2	-80	40	-60	step (prepuls: -60 mV)
pyrilamina	6	HEK	[169]	stabilna	22	whole cell PC	5.1	2; 0	-80	20	-50	step
ranolazyna	12	HEK	[271]	stabilna	23	whole cell PC	4	4; 5.7	-80	20	-50	step
ranolazyna	106	XO	[303]		23	voltage-clamp 2- elektrodowy	2	4; 4	-80	0	-50	step
risperidon	0.074	HEK	[107,401]	stabilna	35	whole cell PC	4	1; 120	-80	20	-40	step ramp
risperidon	0.167	CHO	[191]	?	23	whole cell PC	5	2; 2	-80	20	-40	step
risperidon	0.261	CHO	[71,72]	?	22.5	whole cell PC	5	1; 1	-80	20	-40	step
roksitromycyna	36.5	HEK	[371]	stabilna	35	whole cell PC	4	1; 120	-80	20	-40	step ramp
ropinirol	1.2	CHO	[143]	stabilna	35	whole cell PC	4.2	1; 1	-80	40	-80	protokół AP
ropiwacaina	28	CHO	[108]	?	?	whole cell PC	?	0; 1	-80	60	-80	ramp
ropiwakaina	20	CHO	[109]	?	23	whole cell PC	5	0; 1	-80	60	-80	ramp
ropiwakaina	20.3	CHO	[119]	stabilna	21.5	whole cell PC	4	5; 5	-80	-10	-60	step
ropiwakaina	50	CHO	[109]	?	23	whole cell PC	5	1; 0.03	-80	60	-80	ramp- step
ropiwakaina (S)	23.9	CHO	[319]	przejściowa	23	whole cell PC	5	0; 1	-80	60	-80	ramp
sertindol	0.003	CHO	[191]	?	23	whole cell PC	5	2; 2	-80	20	-40	step
sertindol	0.003	CHO	[258]	?	23	whole cell PC	5	2; 2	-80	20	-40	step
sertindol	0.064	CHO	[356]	stabilna	37?	whole cell PC	5.4	0.5; 0.25	-80	20	-40	step
sertindol	0.21	HEK	[169]	stabilna	22	whole cell PC	5.1	2; 0	-80	20	-50	step

Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
sewofluran	650	XO	[396]		23	voltage-clamp 2- elektrodowy	?	2; 2	-70	40	-70	step
sibutramina	2.5	HEK	[185]	przejściowa	22	whole cell PC	5.4	2; 3	-80	40	-40	step
sibutramina	3.755	CHO	[124]	stabilna	35	whole cell PC	4	1; 0.5	-80	20	-40	step
sibutramina	4.94	CHO	[124]	stabilna	23	PatchXpress 7000A	4	1; 0.5	-80	20	-40	step
sibutramina	7.42	CHO	[183]	stabilna/ przejściowa?	22	whole cell PC	5.4	2; 3	-80	20	-60	step
sildenafil	33.3	HEK	[87]	stabilna	37	whole cell PC	4	0.5; 0.5	-75	10	-40	step
sildenafil	100	HEK	[114]	przejściowa	22.5	whole cell PC	4	4; 4	-80	20	-50	step
sofokarpina	>100	HEK	[409]		37	whole cell PC	5.4	4; 2	-80	40	-50	step
soforydyna	>300	HEK	[409]		37	whole cell PC	5.4	4; 2	-80	40	-50	step
sotalol	5100	XO	[212]		21.5	voltage-clamp 2- elektrodowy	5	1.2; 1.2	-80	40	-40	step
sotalol	7200	XO	[212]		21.5	voltage-clamp 2- elektrodowy	10	1.2; 1.2	-80	40	-40	step
sotalol	69	HEK	[78]		22	whole cell PC	4	1; 1	-80	30	-40	step
sotalol	268	HEK	[188]	stabilna	35	whole cell PC	4	1; 200	-80	20	-80	step ramp
sotalol	278	HEK	[188]	stabilna	35	whole cell PC	4	2; 2	-80	20	-50	step
sotalol	320	HEK	[188]	stabilna	35	whole cell PC	4	0.5; 0.5	-80	20	-50	step
sotalol	810	HEK	[188]	stabilna	22	whole cell PC	4	2; 2	-80	20	-50	step
sotalol	1200	HEK	[373]	stabilna	21	whole cell PC	4	1; 0.5	-80	20	-120	step
sotalol	556	CHO	[397]	stabilna	25	whole cell PC	4	0.4; 0.4	-80	20	-40	step
sparfloksacyna	34.4	CHO	[23]	stabilna	23	whole cell PC	5.4	1; 1	-80	20	-40	step
sparfloksacyna	18	CHO	[163]	stabilna	23	whole cell PC	5	2; 2	-80	20	-40	step
sparfloksacyna	24	CHO	[397]	stabilna	25	whole cell PC	4	0.4; 0.4	-80	20	-40	step
sparfloksacyna	33.2	HEK	[416]	przejściowa	21	whole cell PC	5.6	2; 2	-80	10	-40	step
sparfloksacyna	44	HEK	[219]	stabilna	23	whole cell PC	4	2; 2	-80	60	-40	step (prepuls: -60 mV)
spironolakton	23	CHO	[31]	stabilna	24.5	whole cell PC	4	5; 5	-80	60	-60	step
sulfametoksazol	2200	HEK	[287]	stabilna	21.5	whole cell PC	4	0.5; 1	-60	60	-40	step

Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
tadalafil	>10	HEK	[90,296]	stabilna	37	whole cell PC	4	0.5; 0.5	-75	10	-40	step
tamoksyfen	0.777	CHO	[397]	stabilna	25	whole cell PC	4	0.4; 0.4	-80	20	-40	step
tamoksyfen	45.3	XO	[349]		21	voltage-clamp 2- elektrodowy	5	2; 1.6	-80	20	-40	step
telitromycyna	56.21	HEK	[224]	stabilna	37	whole cell PC	5	3; 4	-80	0	-50	step
telitromycyna	100	HEK	[219]	stabilna	23	whole cell PC	4	2; 2	-80	60	-40	step (prepuls: -60 mV)
telmisartan	24.35	XO	[365]		23	voltage-clamp 2- elektrodowy	?	2; 0	-80	0	-40	step
teobromina	>100	CHO	[60]	stabilna	21	whole cell PC	5.4	2; 2	-80	20	-40	step
terazosyna	17.7	HEK	[355]	stabilna	23.5	whole cell PC	5	2; 2	-80	20	-50	step
terazosyna	113.2	XO	[355]		21	voltage-clamp 2- elektrodowy	5	2; 1.6	-80	20	-40	step
terfenadyna	0.0066	HEK	[188]	stabilna	35	whole cell PC	4	0.5; 0.5	-80	20	-50	step
terfenadyna	0.0066	HEK	[188]	stabilna	35	whole cell PC	4	2; 2	-80	20	-50	step
terfenadyna	0.007	HEK	[188]	stabilna	22	whole cell PC	4	2; 2	-80	20	-50	step
terfenadyna	0.0084	HEK	[188]	stabilna	35	whole cell PC	4	1; 200	-80	20	-80	step ramp
terfenadyna	0.0087	HEK	[342]	stabilna	35	whole cell PC	4	1; 200	-80	20	-80	step ramp
terfenadyna	0.0094	HEK	[378]	stabilna	36	whole cell PC	4	1; 1	-80	40	-50	step
terfenadyna	0.01	CHO	[124]	stabilna	35	whole cell PC	4	1; 0.5	-80	20	-40	step
terfenadyna	0.011	HEK	[107]	stabilna	35	whole cell PC	4	1; 120	-80	20	-40	step ramp
terfenadyna	0.016	HEK	[224]	stabilna	37	whole cell PC	5	3; 4	-80	0	-50	step
terfenadyna	0.0501	HEK	[58]	stabilna	23	whole cell PC	?	1; 1	-80	40	-50	step
terfenadyna	0.052	HEK	[169]	stabilna	22	whole cell PC	5.1	2; 0	-80	20	-50	step
terfenadyna	0.056	CHO	[258]	?	23	whole cell PC	5	2; 2	-80	20	-40	step
terfenadyna	0.061	CHO	[397]	stabilna	25	whole cell PC	4	0.4; 0.4	-80	20	-40	step
terfenadyna	0.077	CHO	[124]	stabilna	23	PatchXpress 7000A	4	1; 0.5	-80	20	-40	step
terfenadyna	0.134	XO	[32]		23	voltage-clamp 2- elektrodowy	2	5; 0.5	-90	0	-70	step

Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
terfenadyna	0.204	HEK	[50]	stabilna	36	whole cell PC	4	0.4; 0.4	-75	10	-40	step
terfenadyna	0.208	HEK	[52]	stabilna	37	whole cell PC	4	0.4; 0.4	-75	10	-40	step
terfenadyna	0.246	XO	[332]		22	voltage-clamp 2- elektrodowy	2	0.5; 0.5	-80	40	-85	step
terfenadyna	0.33	XO	[336]		22.5	voltage-clamp 2- elektrodowy	10	1; 1	-90	0	-100	step
terfenadyna	0.35	XO	[297]		23	voltage-clamp 2- elektrodowy	2	1; 1	-80	40	-80	step
terfenadyna	0.4	XO	[189]		20	voltage-clamp 2- elektrodowy	2	1.2; 1	-80	30	-80	step
terfenadyna	0.431	XO	[36]		23	voltage-clamp 2- elektrodowy	2	4; 6	-80	20	-60	step
terikalant	0.25	XO	[155]		23	voltage-clamp 2- elektrodowy	4	4; 3.8	-90	0	-60	step
terodilina	0.375	HEK	[223]	stabilna	37	whole cell PC	5	3; 4	-80	0	-50	step
tioridazyna	0.0803	HEK	[232]	stabilna	37	whole cell PC	4	2; 4	-80	20	-40	step (prepuls: -40mV)
tioridazyna	0.096	HEK	[188]	stabilna	35	whole cell PC	4	1; 200	-80	20	-80	step ramp
tioridazyna	0.116	CHO	[124]	stabilna	35	whole cell PC	4	1; 0.5	-80	20	-40	step
tioridazyna	0.116	HEK	[188]	stabilna	35	whole cell PC	4	2; 2	-80	20	-50	step
tioridazyna	0.191	CHO	[191]	?	23	whole cell PC	5	2; 2	-80	20	-40	step
tioridazyna	0.2	HEK	[188]	stabilna	22	whole cell PC	4	2; 2	-80	20	-50	step
tioridazyna	0.224	CHO	[181]	przejściowa	35	whole cell PC	5.4	2; 3	-80	20	-60	step
tioridazyna	0.31	CHO	[124]	stabilna	23	PatchXpress 7000A	4	1; 0.5	-80	20	-40	step
tioridazyna	0.39	HEK	[169]	stabilna	22	whole cell PC	5.1	2; 0	-80	20	-50	step
tioridazyna	1.07	CHO	[358]	stabilna	21	whole cell PC	4.8	3.9; 5	-80	30	-60	step
tolterodyna	0.0096	HEK	[223]	stabilna	37	whole cell PC	5	3; 4	-80	0	-50	step
tolterodyna	0.017	CHO	[162]	stabilna	23	whole cell PC	5	2; 2	-80	20	-40	step
topiramet	>1000	HEK	[55]	stabilna	23	whole cell PC	4	4.8; 4.8	-80	20	-50	step
trazodon	0.69	HEK	[342]	stabilna	35	whole cell PC	4	1; 200	-80	20	-80	step ramp

Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
trazodon	2.9	HEK	[413]	stabilna	21	whole cell PC	5	2; 1.6	-80	20	-40	step
trazodon	13.2	XO	[413]		21	voltage-clamp 2- elektrodowy	5	0.4; 0.4	-80	20	-120	step
trifluoperazyna	1.406	CHO	[181]	prześciowa	35	whole cell PC	5.4	2; 3	-80	20	-60	step
trifluoperazyna	9.29	XO	[41]		21.5	voltage-clamp 2- elektrodowy	2	4; 4	-70	40	-60	step
trifluoperazyna	16.6	XO	[41]		21.5	voltage-clamp 2- elektrodowy	2	4; 4	-70	0	-60	step
trifluoperazyna	21.6	XO	[41]		21.5	voltage-clamp 2- elektrodowy	2	4; 4	-70	-40	-60	step
trimetoprim	1	HEK	[287]	stabilna	21.5	whole cell PC	4	0.5; 1	-60	60	-40	step
wardenafil	12.8	HEK	[87]	stabilna	37	whole cell PC	4	0.5; 0.5	-75	10	-40	step
werapamil	0.094	HEK	[107]	stabilna	35	whole cell PC	4	1; 120	-80	20	-40	step ramp
werapamil	0.136	HEK	[188]	stabilna	35	whole cell PC	4	1; 200	-80	20	-80	step ramp
werapamil	0.143	HEK	[406]	stabilna	23	whole cell PC	4	4; 6	-80	20	-50	step
werapamil	0.192	CHO	[124]	stabilna	35	whole cell PC	4	1; 0.5	-80	20	-40	step
werapamil	0.239	CHO	[124]	stabilna	23	PatchXpress 7000A	4	1; 0.5	-80	20	-40	step
werapamil	0.37	HEK	[188]	stabilna	35	whole cell PC	4	2; 2	-80	20	-50	step
werapamil	0.41	HEK	[224]	stabilna	37	whole cell PC	5	3; 4	-80	0	-50	step
werapamil	0.444	HEK	[188]	stabilna	22	whole cell PC	4	2; 2	-80	20	-50	step
werapamil	0.94	CHO	[397]	stabilna	25	whole cell PC	4	0.4; 0.4	-80	20	-40	step
werapamil	1.055	CHO	[124]	stabilna	23	PatchXpress 7000A	4	1; 0.5	-80	20	-40	step
werapamil	5.1	XO	[75]		21.5	voltage-clamp 2- elektrodowy	2	2; 2	-90	0	-60	step
werapamil	114	XO	[75]		21.5	voltage-clamp 2- elektrodowy	30	2; 2	-90	40	-90	step
werapamil	0.215	HEK	[279]	stabilna	37	whole cell PC	4	2; 4	-80	20	-40	step
wesnarinon	1.1	HEK	[167]	?	23.5	whole cell PC	5.4	4; 4	-80	20	-60	step
wesnarinon	17.7	XO	[157]		23	voltage-clamp 2- elektrodowy	2	2; 0	-90	0	-80	step
ziprasidon	0.12	HEK	[330]	stabilna	37	whole cell PC	5	3; 3	-80	0	-50	step

Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
ziprasidon	0.165	CHO	[397]	stabilna	25	whole cell PC	4	0.4; 0.4	-80	20	-40	step
ziprasidon	0.169	CHO	[191]	?	23	whole cell PC	5	2; 2	-80	20	-40	step
ziprasidon	0.24	HEK	[80]	?	20	whole cell PC	4	1; 1	-80	30	-40	step
ziprasidon	2.8	XO	[330]		23	voltage-clamp 2- elektrodowy	2	5; 5	-90	0	-90	step

<sup>1</sup> oryginalna publikacja

<sup>2</sup> rodzaj transfekcji komórek

<sup>3</sup> technika pomiarowa zastosowana w doświadczeniu

<sup>4</sup> stężenie jonów potasowych w płynie omywającym komórki

<sup>5</sup> potencjał utrzymania

<sup>6</sup> wartość napięcia depolaryzującego

<sup>7</sup> napięcie pomiarowe

<sup>8</sup> rodzaj protokołu napięć

## 2. WSPÓŁCZYNNIKI EKSTRAPOLACJI

Podczas budowy bazy danych i przeprowadzonej analizy zgromadzonych informacji zwracał uwagę fakt znaczącej zależności uzyskiwanych eksperymentalnie wyników pomiarów  $IC_{50}$  od warunków doświadczeń. Jednocześnie brak jest ścisłych regulacji dotyczących wyboru modelu, standardowych warunków oznaczeń czy stosowanego protokołu. Analiza danych literaturowych wykazała, że największe różnice obserwowanych wartości  $IC_{50}$  dla danej substancji występują dla oznaczeń wykonanych z użyciem różnych linii komórkowych (modeli *in vitro*) oraz temperatury utrzymywanej podczas oznaczeń [188,385, Tabela 2]. W publikacjach dotyczących predykcji interakcji lek-kanal potasowy hERG, w zbiorze danych stosowanych do konstrukcji modelu zwykle znajdują się wartości  $IC_{50}$  uzyskane w różnych warunkach eksperymentalnych [11,16,33,76,172,208,267,283,333,336,346,359,401], co może skutkować błędną interpretacją wyników i ograniczeniem możliwości praktycznego zastosowania danego modelu. Konieczne jest więc opracowanie metod umożliwiających rzetelne porównanie danych uzyskanych w toku niejednorodnych doświadczeń i racjonalną ekstrapolację wyników. Jednym z potencjalnych rozwiązań mogą być zaproponowane w niniejszej pracy współczynniki ekstrapolacji. Współczynniki pozwalające na międzysystemową oraz temperaturową ekstrapolację wartości  $IC_{50}$  umożliwiają dogodny wybór niektórych z wspomnianych powyżej parametrów eksperymentalnych bez ryzyka deprecjacji otrzymywanych wyników i mogą stanowić jeden z pierwszych modułów narzędzia opartego o realizowane komputerowo techniki eksploracji danych.

### 2.1. Metodyka

Źródłem danych, które posłużyły do wyznaczenia współczynników ekstrapolacji była stworzona wcześniej baza. Ostateczny zbiór informacji zawierający 599 rekordów poddano dalszej optymalizacji. Wykluczono rekordy, w których brakowało precyzyjnie zdefiniowanej temperatury oznaczeń (z wyjątkiem oznaczeń z zastosowaniem komórek XO, których wykonanie możliwe jest wyłącznie w temperaturze pokojowej). Usunięto również wartości  $IC_{50}$  wyznaczone techniką patch clamp w konfiguracji innej niż whole cell oraz te, dla których technika wyznaczenia nie była jednoznacznie określona. Ze zbioru danych wykluczono także wartości  $IC_{50}$  dla form izomerycznych związków stosowanych również w formie racematu, oznaczone przy wyższym niż fizjologiczne stężeniu potasu w płynie omywającym oraz rekordy z niejednoznacznie określonym modelem doświadczalnym. W kilku publikacjach wartości  $IC_{50}$  podane w tekście różniły się znacząco od wartości przedstawionych na zamieszczonych wykresach. W przypadku, gdy udało się skontaktować z autorem publikacji i uzyskać informacje na temat rzeczywistych wyników doświadczeń, wartości  $IC_{50}$  poprawiono, natomiast, gdy taki kontakt był niemożliwy rekordy zostały usunięte. Z bazy danych wykluczono również wpisy odnoszące się do pentamidyny. Powodem jej wyeliminowania był znacząco dłuższy niż w innych oznaczeniach czas inkubacji komórek z lekiem poprzedzający pomiar. W przypadku



pentamidyny, a także np. trójtlenku arsenu czy pochodnych geldamycyny, tego rodzaju zmiana może mieć istotne znaczenie, ze względu na specyficzny mechanizm wpływu na prąd potasowy hERG. Wykazano, że pentamidyna w standardowej procedurze nie powoduje bloku kanału hERG, jednak po dłuższej inkubacji obserwuje się zmniejszenie amplitudy prądu potasowego, co związane jest z jej wpływem na przemieszczanie podjednostek kanału do błony komórkowej [27,61,83]. W pewnych doniesieniach autorzy wskazują także, że podczas doświadczeń z wykorzystaniem kanałów hERG transfekowanych do komórek XO np. sotalol czy pochodne erytromycyny nie wykazują właściwego potencjału blokowania kanału hERG [385]. W niektórych pracach dotyczących eksperymentów przeprowadzanych w warunkach *in silico* autorzy eliminowali ze zbioru danych sotalol, erytromycynę oraz jej pochodne [90,346]. Argumentacja przemawiająca za wykluczeniem nie była jednak wystarczająco przekonująca, stąd zdecydowano się na pozostawienie tych związków w bazie danych. W tabeli 3 przedstawiono wszystkie rekordy wykluczone z bazy danych. Ostatecznie baza zawierała 512 wartości  $IC_{50}$  209 różnych substancji chemicznych. Podczas wyznaczania współczynników ekstrapolacji nie brano pod uwagę także 31 rekordów opisujących eksperymenty, w których testowane stężenia były niższe od wartości  $IC_{50}$  (w tabeli 2 rekordy oznaczone symbolem „>”).

W celu przygotowania odpowiedniego dla wyznaczenia współczynników ekstrapolacji zbioru danych z oryginalnej bazy wyodrębniono informacje o wartości  $IC_{50}$  oraz modelu i temperaturze danego oznaczenia. Ciągłe wartości temperatury podzielono na dwie kategorie: „temperatura pokojowa” (19-25°C) oraz „temperatura fizjologiczna” (34-37°C). Przygotowane wstępnie dane pogrupowane zostały według modelu zastosowanego w doświadczeniach oraz temperatury oznaczeń na pięć podzbiorów (‘XO’; ‘CHO<sub>r.t.</sub>’; ‘CHO<sub>fizj.</sub>’; ‘HEK<sub>r.t.</sub>’ oraz ‘HEK<sub>fizj.</sub>’) koniecznych do wyznaczenia współczynników ekstrapolacji. Z każdego podzbioru wyselekcjonowano dane, dla których znana była standardowa wartość  $IC_{50}$ , za jaką przyjęto stężenie otrzymane z wykorzystaniem linii komórkowej HEK w temperaturze fizjologicznej. Warunki standardowe wybrano opierając się na założeniu, że komórki HEK, ze względu na ludzkie pochodzenie, oraz temperatura 34-37°C będą najlepiej spośród wszystkich stosowanych systemów ekspresji naśladowały warunki interakcji leku z kanałem hERG w warunkach klinicznych [410]. Oszacowanie współczynnika temperaturowego polegało na obliczeniu stosunków wartości  $IC_{50}$  uzyskanych eksperymentalnie z użyciem komórek HEK w temperaturze pokojowej i wartości standardowych - otrzymanych na tym samym modelu w temperaturze fizjologicznej. Porównanie proporcji  $IC_{50}$  dla poszczególnych par modeli (XO/HEK oraz CHO/HEK) w homogennej temperaturze pozwoliło z kolei na oszacowanie współczynników ekstrapolacji międzysystemowej. Wyznaczone współczynniki posłużyły następnie do przeliczenia wszystkich rekordów na warunki standardowe (wg schematu przedstawionego poniżej – przykład przeliczenia wartości  $IC_{50}$  otrzymanych na modelu XO w temperaturze pokojowej).

$$\frac{mIC50_{XO_{r.t.}}^{S_1}}{MW^{S_1}} \xrightarrow{F_{XO_{r.t.}/HEK_{fizj}}} \frac{cIC50_{HEK_{fizj}}^{S_1}}{MW^{S_1}}$$

$$\frac{mIC50_{XO_{r.t.}}^{S_n}}{MW^{S_n}} \xrightarrow{F_{XO_{r.t.}/HEK_{fizj}}} \frac{cIC50_{HEK_{fizj}}^{S_n}}{MW^{S_n}}$$

m – zmierzona eksperymentalnie wartość  $IC_{50}$

c – obliczona wartość  $IC_{50}$

S – substancja

MW – masa molowa

F – współczynnik ekstrapolacji

**Rycina 24.** Metoda przeliczania wartości  $IC_{50}$  otrzymanych na modelu XO w temperaturze pokojowej na warunki standardowe.

W celu wybrania najlepszych współczynników (zestawów współczynników), dla wartości uzyskanych z zastosowaniem komórek XO lub CHO oraz na komórkach HEK w temperaturze pokojowej odpowiadającą im standardową wartość  $IC_{50}$  ( $HEK_{fizj}$ ) wyznaczono za pomocą wszystkich możliwych ścieżek przeliczeń.

Ze względu na opisywany w literaturze naukowej błąd mierzonych wartości  $IC_{50}$  dla dużych cząsteczek (o masie powyżej 600Da), widoczny również w zgromadzonych danych (Rycina 25), w celu zwiększenia spójności danych zdecydowano na standaryzację poprzez przyrównanie  $IC_{50}$  ( $\mu M$ ) do masy molowej (M).

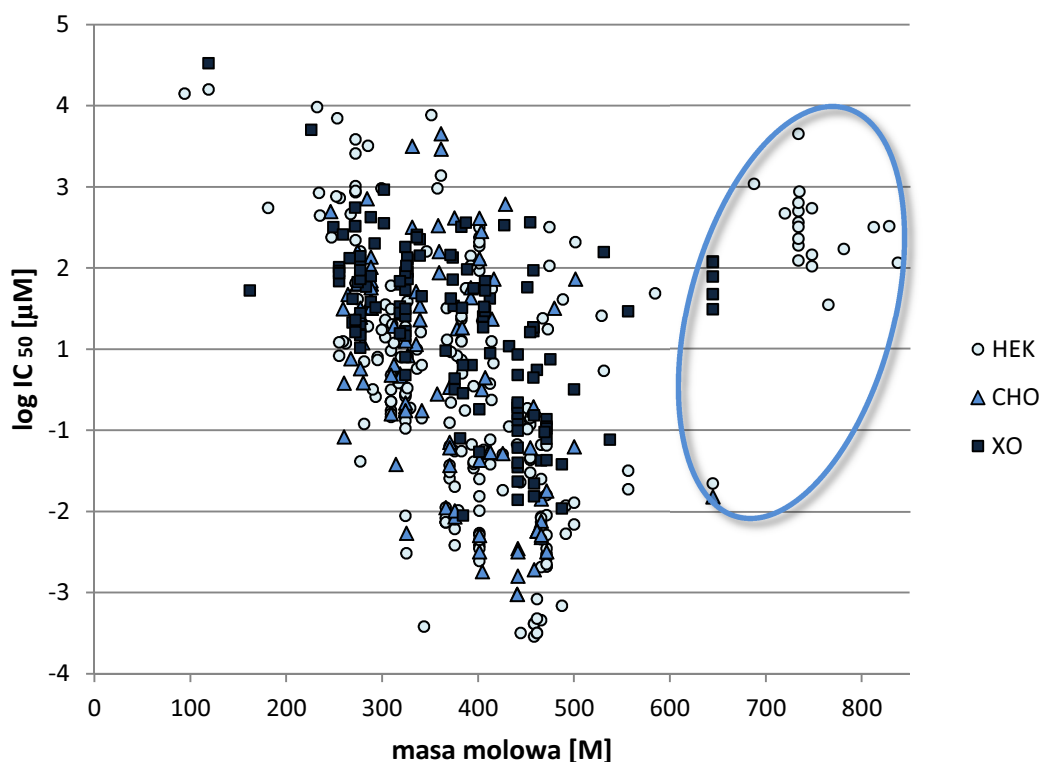
Jako wartość oczekiwaną zmiennej losowej wybrano medianę, aby uniknąć wrażliwości wartości średniej na skrajne wartości  $IC_{50}$ .

Tabela 3. Rekordy wykluczone z natywnej bazy danych.

No	Substancja	Przyczyna wykluczenia	Źródła
1	4.4'-dimetyl-difenhydramina	niefizjologiczne stężenie potasu	[337]
2	amiodaron	niestandardowa technika pomiaru (PatchXpress 7000A)	[124]
3	amiodaron	niefizjologiczne stężenie potasu	[212]
4	amitriptylina	niestandardowa technika pomiaru (PatchXpress 7000A)	[124]
5	astemizol	niestandardowa technika pomiaru (PatchXpress 7000A)	[124]
6	astemizol	niefizjologiczne stężenie potasu	[336]
7	azimilid	niefizjologiczne stężenie potasu	[212]
8	balofloksacyna	niezdefiniowana temperatura	[306]
9	bepidil	niestandardowa technika pomiaru (PatchXpress 7000A)	[124]
10	bromfeniramina	niezdefiniowana temperatura	[315]
11	cetyryzyna	niefizjologiczne stężenie potasu	[336]
12	chinidyna	niestandardowa technika pomiaru (PatchXpress 7000A)	[124]
13	chinidyna	niefizjologiczne stężenie potasu	[212]
14	chlorobutanol	niezdefiniowana temperatura	[193]
15	chloroform	niestandardowa technika eksperymentalna ze względu na stan skupienia	[301]
16	chlortalidon	dane niepewne	[222]
17	cizapryd	niestandardowa technika pomiaru (PatchXpress 7000A)	[124]
18	częsteczka A i B	błędne dane	[165]
19	dekstrometorfan	forma izomeryczna	[60]
20	difenhydramina	niefizjologiczne stężenie potasu	[337]
21	difenhydramina (-)	forma izomeryczna i niezdefiniowana temperatura	[337]
22	difenhydramina (+)	forma izomeryczna i niezdefiniowana temperatura	[337]
23	diltiazem	niestandardowa technika pomiaru (PatchXpress 7000A)	[124]
24	dipfluzyna	niezdefiniowana temperatura	[404]
25	dofetylid	niepewna technika pomiaru (inside-out macro patch) oraz niefizjologiczne stężenie potasu	[16,174]
26	dofetylid	niestandardowa technika pomiaru (PatchXpress 7000A)	[124]
27	dofetylid	niefizjologiczne stężenie potasu	[211,212]

28	ebastyna	błędne dane	[189]
29	EGCG	niezdefiniowana temperatura oraz potencjał depolaryzujący	[171]
30	fluoksetyna	niestandardowa technika pomiaru (PatchXpress 7000A)	[124]
31	fluoksetyna	dane niepewne	[269]
32	gemifloksacyna	niezdefiniowana temperatura	[306]
33	haloperidol	niestandardowa technika pomiaru (PatchXpress 7000A)	[124]
34	hydroksyzyna	niefizjologiczne stężenie potasu	[337]
35	ibutilid	niefizjologiczne stężenie potasu	[213]
36	imipramina	niepewna technika pomiaru (perforowany PatchClamp?)	[99,344]
37	imipramina	niestandardowa technika pomiaru (PatchXpress 7000A)	[124]
38	kapsaicyna	niefizjologiczne stężenie potasu	[395]
39	ketanseryna	niefizjologiczne stężenie potasu	[340]
40	klarytromycyna	niezdefiniowana temperatura	[1]
41	klebopryd	niezdefiniowana temperatura	[182]
42	klotrimazol	niepewna technika pomiaru (whole-cell lub perforowany patch clamp)	[357]
43	lewobupiwakaina	forma izomeryczna	[104,119,319]
44	lewofloksacyna	forma izomeryczna	[4,161]
45	loratadyna	niefizjologiczne stężenie potasu	[336]
46	metadon	niezdefiniowana temperatura	[193,214]
47	metadon	niestandardowa technika pomiaru (Qpatch) oraz niezdefiniowana temperatura	[95]
48	metadon (S)	forma izomeryczna	[88]
49	metadon(R)	forma izomeryczna	[88]
50	mizolastyna	niefizjologiczne stężenie potasu	[338,401]
51	moksifloxacyna	niezdefiniowana temperatura	[4]
52	norfluoksetyna	dane niepewne	[269]
53	oksykodon	niestandardowa technika pomiaru (Qpatch) oraz niezdefiniowana temperatura	[95]
54	olanzapina	niestandardowa technika pomiaru (PatchXpress 7000A)	[124]
55	orfenadryna	błędne dane oraz niepewna temperatura	[302]
56	orfenadryna	niefizjologiczne stężenie potasu	[337]

57	pentamidyna	niestandardowy czas inkubacji i niepewna temperatura	[49,90,198]
58	ropiwakaina	niezdefiniowana temperatura	[108]
59	ropiwakaina (S)	forma izomeryczna	[319]
60	sertindol	niepewna temperatura	[356]
61	sibutramina	niestandardowa technika pomiaru (PatchXpress 7000A)	[124]
62	sotalol	niefizjologiczne stężenie potasu	[212]
63	terfenadyna	niestandardowa technika pomiaru (PatchXpress 7000A)	[124]
64	terfenadyna	niefizjologiczne stężenie potasu	[336]
65	tioridazyna	niestandardowa technika pomiaru (PatchXpress 7000A)	[124]
66	werapamil	niefizjologiczne stężenie potasu	[75]
67	werapamil	niestandardowa technika pomiaru (PatchXpress 7000A)	[124]



Rycina 25. Stosunek masy molowej do wartości  $IC_{50}$  uzyskanych z wykorzystaniem różnych systemów *in vitro*.

## 2.2. Wyniki

### 2.2.1. Współczynniki ekstrapolacji

W tabelach 4 i 5 przedstawiono wyznaczone wartości współczynników ekstrapolacji. Jako miarę zróżnicowania wykorzystano 95% przedział ufności dla mediany oraz oparty ma medianie odpowiednik odchylenia standardowego - medianowe odchylenie bezwzględne (MAD) liczony wg wzoru:

$$MAD = \text{median}(|X_i - \text{mediana} X_i|) \quad \text{Równanie 27.}$$

Tabela 4. Współczynniki ekstrapolacji - modele XO i HEK.

	$HEK_{(r.t.)}/HEK_{(fizj)}$	$XO_{(r.t.)}/HEK_{(r.t.)}$	$XO_{(r.t.)}/HEK_{(fizj)}$
rekordy (n)	189	112	151
substancje (n)	27	31	24
zakres	0.03-63.08	0.3-1740	0.644-1740
mediana	1.4545	9.1447	20.4878
MAD	1.00	6.59	15.07
$CI_{95\%}$	1.16 – 1.75	6.56 – 11.73	13.01 – 28.24

Tabela 5. Współczynniki ekstrapolacji – model CHO.

	CHO <sub>(r.t.)</sub> /CHO <sub>(fizj)</sub>	CHO <sub>(r.t.)</sub> /HEK <sub>(r.t.)</sub>	CHO <sub>(r.t.)</sub> /HEK <sub>(fizj)</sub>	CHO <sub>(fizj)</sub> /HEK <sub>(fizj)</sub>
rekordy (n)	31	81	110	72
substancje (n)	15	25	18	11
zakres	0.56-13.43	0.014-18	0.086-18	0.048-6.67
mediana	2.2929	1.6576	3.0000	1.0000
MAD	1.41	1.11	2.09	0.37
CI <sub>95%</sub>	0.87 – 3.72	1.14– 2.17	2.11 – 3.89	0.86 – 1.15

n – liczba par rekordów

MAD – medianowe odchylenie bezwzględne

CI<sub>95%</sub> - 95% przedział ufności liczony wg metody zaproponowanej przez Bonett & Price [24].

### 2.2.2. Walidacja wewnętrzna

Weryfikacji współczynników dokonano w oparciu o struktury cząsteczek, dla których w bazie danych znajdowały się co najmniej dwa rekordy, z których przynajmniej jeden dotyczył pomiarów wykonanych na komórkach HEK w temperaturze fizjologicznej. Rozbieżność pomiędzy wartością szacowaną a rzeczywistą oceniano za pomocą pierwiastka błędu średniokwadratowego (RMSE) obliczonego dla poszczególnych przeliczeń (XO/HEK<sub>fizj</sub>; CHO<sub>r.t.</sub>/HEK<sub>fizj</sub>; CHO<sub>fizj</sub>/HEK<sub>fizj</sub>; HEK<sub>r.t.</sub>/HEK<sub>fizj</sub>). W przypadku gdy w bazie znajdowała się więcej niż jedna natywna, wyznaczona w warunkach standardowych, wartość IC<sub>50</sub> różnica liczona była w stosunku do każdej z nich. W celu wybrania najlepszych współczynników (zestawów współczynników) obliczono i porównano wartości błędu RMSE dla wszystkich możliwych ścieżek przeliczeń (Tabela 6).

Tabela 6. Wartości błędu dla poszczególnych schematów ekstrapolacji.

	XO <sub>(r.t.)</sub> ↓ HEK <sub>(fizj)</sub>	XO <sub>(r.t.)</sub> ↓ HEK <sub>(r.t.)</sub> ↓ HEK <sub>(fizj)</sub>	CHO <sub>(r.t.)</sub> ↓ HEK <sub>(fizj)</sub>	CHO <sub>(r.t.)</sub> ↓ HEK <sub>(r.t.)</sub> ↓ HEK <sub>(fizj)</sub>	CHO <sub>(r.t.)</sub> ↓ CHO <sub>(fizj)</sub> ↓ HEK <sub>(fizj)</sub>	CHO <sub>(fizj)</sub> ↓ HEK <sub>(fizj)</sub>	HEK <sub>(r.t.)</sub> ↓ HEK <sub>(fizj)</sub>
rekordy (n)	151		110			72	191
substancje (n)	23		17			11	28
zakres IC <sub>50</sub> /MW	7E-5 - 18.7		3.4E-5 – 2.04			1.3E-5-0.002	2.2E-6-4.4
średnia	0.42		0.11			0.0005	0.23
mediana	0.032		0.0004			7.4E-5	0.003
RMSE	<b>0.024</b>	0.051	0.07	0.05	<b>0.047</b>	<b>0.0004</b>	<b>0.37</b>

n - liczba rekordów

RMSE – pierwiastek błędu średniokwadratowego (ang. *Root Meansquare Error*); wg wzoru:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (S_1^i - S_n^i)^2}{n}}$$

Równanie 28.

Wartości błędów obliczone dla ścieżki  $HEK_{r.t.}/HEK_{fizj}$  oraz  $CHO_{fizj}/HEK_{fizj}$  mają jedynie wartość informacyjną, ponieważ są to jedyne możliwe opcje przeliczenia i brak jest alternatyw do porównania. Wartości błędu RMSE opisujące różnice pomiędzy eksperymentalną, wyznaczoną na komórkach HEK w temperaturze fizjologicznej wartością  $IC_{50}$  poszczególnych substancji a wartością  $IC_{50}$  uzyskaną z wykorzystaniem najlepszych współczynników zestawiono z wartością błędu RMSE porównania eksperymentalnych wartości  $IC_{50}$  dla  $HEK_{fizj}$  z  $IC_{50}$  uzyskanymi na innym modelu i/lub w odmiennej temperaturze (Tabela 7). Zastosowanie współczynników ekstrapolacji spowodowało zmniejszenie błędu RMSE.

**Tabela 7.** Zestawienie wartości błędu RMSE dla porównania wartości  $IC_{50}$  eksperymentalnych i obliczonych za pomocą najlepszych faktorów z wartościami standardowymi.

	XO/ $HEK_{(fizj)}$	CHO <sub>(r.t.)</sub> / $HEK_{(fizj)}$	CHO <sub>(fizj)</sub> / $HEK_{(fizj)}$	HEK <sub>(r.t.)</sub> / $HEK_{(fizj)}$
<b>RMSE <math>IC_{50}</math> eksperymentalne</b>	2.49	0.20	0.0004	0.63
<b>RMSE <math>IC_{50}</math> obliczone</b>	0.024	0.047	0.0004	0.37

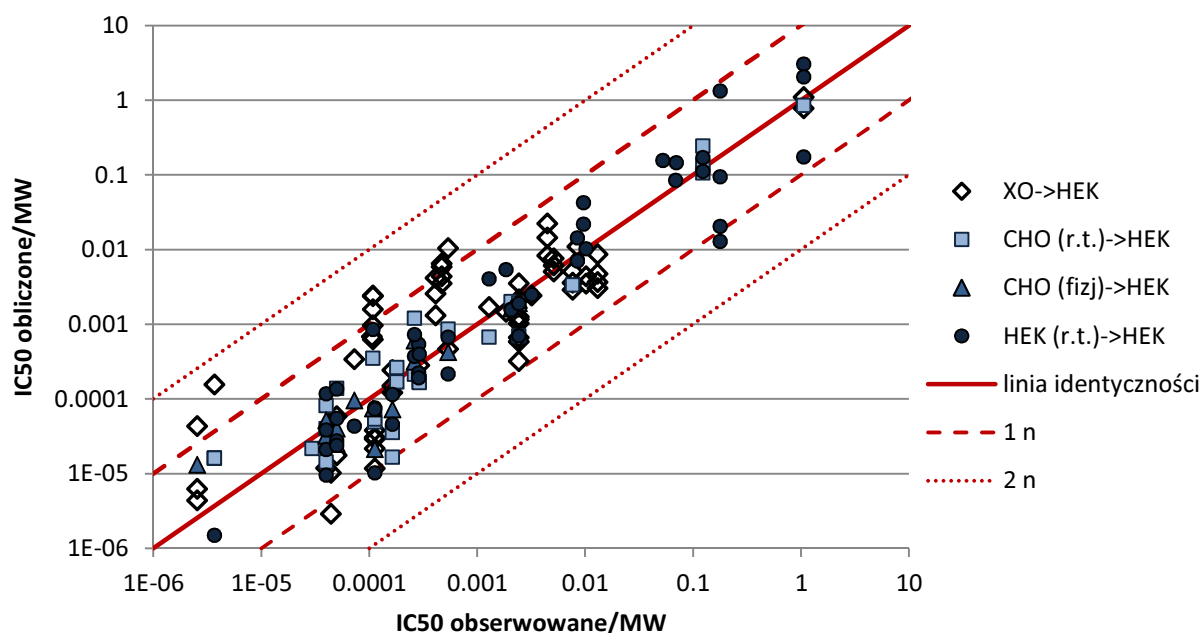
Istotność różnic pomiędzy poszczególnymi wartościami RMSE sprawdzono testem kolejności par Wilcozona (zmienne mają rozkład odbiegający od normalnego – test Shapiro-Wilka). Wykazano istotne statystycznie różnice przy poziomie ufności 0.05 (z wyjątkiem przeliczenia  $CHO_{(fizj)}/HEK_{(fizj)}$ , dla którego współczynnik ekstrapolacji jest równy 1). Wyniki te potwierdzają zasadność stosowania współczynników ekstrapolacji w celu standaryzacji danych otrzymywanych w różnych warunkach doświadczalnych.

Weryfikacji współczynników dokonano również przez porównanie wartości stężeń powodujących połowiczny blok kanału potasowego hERG obliczonych przy zastosowaniu najlepszych współczynników ekstrapolacji z wartościami natywnymi (Rycina 26). W przypadku gdy w bazie znajdowała się więcej niż jedna wartość  $IC_{50}$  wyznaczona w warunkach przyjętych za standardowe za wartość referencyjną przyjmowano ich średnią arytmetyczną. Zdecydowana większość wartości  $IC_{50}$  obliczonych za pomocą współczynników ekstrapolacji nie różniła się od wartości zmierzonych doświadczalnie na komórkach HEK w temperaturze fizjologicznej więcej niż o jeden rząd wielkości.

### 2.2.3. Walidacja zewnętrzna

Współczynniki ekstrapolacji poddano walidacji za pomocą zewnętrznego, niezależnego zbioru danych uzyskanych zgodnie z metodyką zastosowaną na etapie budowy pierwotnej bazy danych. W zbiorze walidacyjnym znalazły się rekordy charakteryzujące eksperymenty opisane, opublikowane i udostępnione po zakończeniu gromadzenia informacji do bazy danych uczących użytych do opracowania współczynników ekstrapolacji (Tabela 9).





Rycina 26. Porównanie eksperymentalnych i obliczonych wartości  $IC_{50}$ .

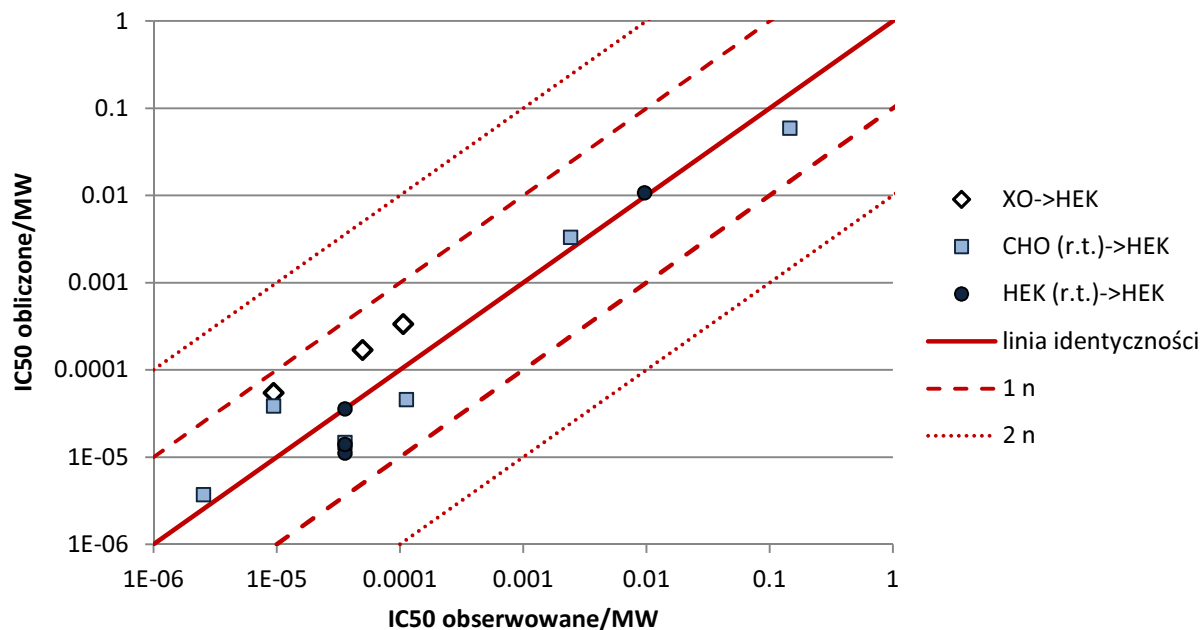
Proces walidacji polegał na przeliczeniu wartości  $IC_{50}$  otrzymanych w różnych warunkach temperaturowych oraz z wykorzystaniem różnych modeli na warunki standardowe, a następnie, podobnie jak w przypadku walidacji wewnętrznej, porównaniu wartości obliczonych z wartościami wyznaczonymi doświadczalnie z wykorzystaniem modelu HEK w temperaturze fizjologicznej. Wartości błędu RMSE dla poszczególnych przeliczeń przedstawiono w tabeli 8.

Tabela 8. Wartości błędu dla poszczególnych schematów ekstrapolacji – zbiór walidacyjny.

	XO <sub>(r.t.)</sub> ↓ HEK <sub>(fizj.)</sub>	CHO <sub>(r.t.)</sub> ↓ HEK <sub>(fizj.)</sub>	HEK <sub>(r.t.)</sub> ↓ HEK <sub>(fizj.)</sub>	razem
rekordy (n)	35	11	22	68
substancje (n)	3	6	2	9
zakres $IC_{50}/MW$	5.4E-5 – 0.0003	3.7E-6 – 0.39	1.1E-5 – 0.01	3.7E-6 - 0.39
średnia	0.0002	0.065	0.0027	0.033
mediana	0.0002	4.53E-05	2.47E-05	4.97E-05
RMSE	0.0001	0.048	0.002	0.031

W wyniku zastosowania procedury standaryzacji z wykorzystaniem wyznaczonych współczynników ekstrapolacji wartość błędu RMSE dla rekordów ze zbioru walidacyjnego zmalała o 36.3% w porównaniu do błędu średniokwadratowego liczonego na podstawie różnic pomiędzy wartością standardową  $IC_{50}$  danej substancji chemicznej a oryginalną wartością  $IC_{50}$  otrzymaną w innych warunkach doświadczalnych. Porównanie wartości stężeń powodujących połowiczny blok kanału potasowego hERG obliczonych przy

zastosowaniu współczynników ekstrapolacji z wartościami natywnymi przedstawiono również na rycinie 27.



Rycina 27. Porównanie eksperymentalnych i obliczonych wartości  $IC_{50}$  – zbiór walidacyjny.

Tabela 9. Zewnętrzny zbiór walidacyjny.

Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	trans <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
17- $\beta$ -estradiol	>0.1	HEK	[196]		22	perforowany PC	4.8	2; 0	-80	20	-40	step
4-aminopirydyna	3800	HEK	[276]	stabilna	35	whole cell PC	4	1; 200	-80	20	-80	step ramp
akonityna	1.801	XO	[209]		23	voltage clamp 2- elektrodowy	2	2; 1.5	-80	40	-30	step
amoksapina	21.6	XO	[249]		21	voltage clamp 2- elektrodowy	5	2; 1.6	-80	20	-40	step
amoksapina	5.1	HEK	[249]	stabilna	21	whole cell PC	5	2; 2	-80	20	-50	step
antofloksacyna	460.37	HEK	[122]		22	whole cell PC	5.4	2; 2	-80	20	-40	step
astemizol	0.0051	CHO	[259]		23	whole cell PC	5	3; 0	-80	20	20	step
atomoksetyna	6.26	HEK	[298]	stabilna	22	whole cell PC	5	0.4; 0.4	-80	20	-120	step
atomoksetyne	>10	XO	[298]		22	voltage clamp 2- elektrodowy	5	0.4; 0.4	-80	40	-60	step
bromfeniramina	1.75	HEK	[254]	przejściowa	23	whole cell PC	5.4	2; 3	-80	40	-40	step
chinidyna	3.2	CHO	[259]		23	whole cell PC	5	3; 0	-80	20	20	step
cizapryd	<0.1	HEK	[111]		37	whole cell PC	4	0.75; 0.75	-70	0	-50	step
cizapryd	0.0075	HEK	[215]	przejściowa	23	whole cell PC	0	4; 0	-80	50	-50	step
cizapryd	0.0241	HEK	[215]	przejściowa	23	whole cell PC	5	4; 0	-80	50	-50	step
cizapryd	0.1088	HEK	[215]	przejściowa	23	whole cell PC	135	4; 0	-80	50	-50	step
cizapryd	0.0075	HEK	[234]	stabilna	37	whole cell PC	4	2; 4	-80	20	-40	step
cizapryd	0.016	HEK	[234]	stabilna	37	whole cell PC	4	2; 0	-80	-20	-40	step
cizapryd	0.013	HEK	[234]	stabilna	37	whole cell PC	4	1; 200	-80	20	-80	step ramp
cizapryd	0.0205	CHO	[259]		23	whole cell PC	5	3; 0	-80	20	20	step
cizapryd	0.0094	HEK	[360]	stabilna	23	whole cell PC	4	1; 240	-80	40	-80	step ramp
CJ-033466	2.6	HEK	[360]	stabilna	23	whole cell PC	4	1; 240	-80	40	-80	step ramp
cząsteczka 1	0.95	HEK	[322]	stabilna	23	whole cell PC	?	4; 4	-80	10	-50	step
cząsteczka 13	0.6	HEK	[322]	stabilna	23	whole cell PC	?	4; 4	-80	10	-50	step
cząsteczka 17	0.42	HEK	[322]	stabilna	23	whole cell PC	?	4; 4	-80	10	-50	step

Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
cząsteczka 22	1.67	HEK	[322]	stabilna	23	whole cell PC	?	4; 4	-80	10	-50	step
cząsteczka 23	1.86	HEK	[322]	stabilna	23	whole cell PC	?	4; 4	-80	10	-50	step
cząsteczka 24	1.01	HEK	[322]	stabilna	23	whole cell PC	?	4; 4	-80	10	-50	step
cząsteczka 26	0.65	HEK	[322]	stabilna	23	whole cell PC	?	4; 4	-80	10	-50	step
cząsteczka 27	2.28	HEK	[322]	stabilna	23	whole cell PC	?	4; 4	-80	10	-50	step
cząsteczka 4	4.17	HEK	[322]	stabilna	23	whole cell PC	?	4; 4	-80	10	-50	step
cząsteczka 7	0.285	HEK	[322]	stabilna	23	whole cell PC	?	4; 4	-80	10	-50	step
deksrazoksan	>30	HEK	[79]		23	whole cell PC	4	1; 1	-80	30	-40	step
dezipramina	28.5	XO	[134]		21.5	voltage clamp 2- elektrodowy	2	4; 0	-70	0	-60	step
dezipramina	177.3	XO	[134]		21.5	voltage clamp 2- elektrodowy	2	4; 0	-70	0	-140	step
dihydrotestosteron	>1	HEK	[196]		22	perforowany PC	4.8	2; 0	-80	20	-40	step
dofetylid	0.49	XO	[145]		21	voltage clamp 2- elektrodowy	2	2; 0	-80	0	-60	step
dofetylid	0.0035	HEK	[234]	stabilna	37	whole cell PC	4	2; 4	-80	20	-40	step
dofetylid	0.005	HEK	[234]	stabilna	37	whole cell PC	4	2; 4	-80	-20	-40	step
dofetylid	0.004	HEK	[234]	stabilna	37	whole cell PC	4	1; 200	-80	20	-80	step ramp
dofetylid	0.051	CHO	[259]		23	whole cell PC	5	3; 0	-80	20	20	step
doksorubicyna	>30	HEK	[79]		23	whole cell PC	4	1; 1	-80	30	-40	step
dronaderon	3.8	XO	[314]		23	voltage clamp 2- elektrodowy	?	2; 0	-90	0	-80	step
E-4031	1.39	XO	[145]		21	voltage clamp 2- elektrodowy	2	2; 0	-80	0	-60	step
E-4031	>0.1	CHO	[241]		37	whole cell PC	4	0.75; 0.75	-70	0	-50	step
erytromycyna	129.5	CHO	[259]		23	whole cell PC	5	3; 0	-80	20	20	step
estriol	>0.001	HEK	[156]	stabilna	22	perforowany PC	4.8	2; 2	-80	20	-40	step
estron	>0.001	HEK	[156]	stabilna	22	perforowany PC	4.8	2; 2	-80	20	-40	step
famotydyna	>10	CHO	[241]		37	whole cell PC	4	0.75; 0.75	-70	0	-50	step
flukonazol	>10	HEK	[129]	przejściowa	23	whole cell PC	5.4	3; 0	-80	10	10	step

Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
flukonazol	48.2	HEK	[129]	przejściowa	23	whole cell PC	5.4	3; 2	-80	20	-40	step
heroina	427	HEK	[416]	stabilna	21	whole cell PC	5.6	2; 2	-80	20	-40	step
hydroksyzyna	0.62	CHO	[288]	przejściowa	37	whole cell PC	5.4	2; 0	-80	40	-80	protokół AP
klofilium	~0.03	XO	[260]		23	voltage clamp 2- elektrodowy	2	5; 0.4	-90	0	-70	step
lobelina	0.34	HEK	[148]	stabilna	23	whole cell PC	4	4; 0	-80	20	-60	step
metadon	4.8	HEK	[416]	stabilna	21	whole cell PC	5.6	2; 2	-80	20	-40	step
mleczan lewofloksacyny	43.86	HEK	[122]		22	whole cell PC	5.4	2; 2	-80	20	-40	step
mozapryd	4.8	HEK	[360]	stabilna	23	whole cell PC	4	1; 240	-80	40	-80	step ramp
perheksylina	5.9	CHO	[259]		23	whole cell PC	5	3; 0	-80	20	20	step
PNU-0068611A	0.08084	XO	[260]		23	voltage clamp 2- elektrodowy	2	5; 0.4	-90	0	-70	step
Prunicus fructus ekstrakt wodny	>100	HEK	[313]	stabilna	23	whole cell PC	5	2; 4	-80	40	-50	step
prymachina	21.5	HEK	[184]	przejściowa	23	whole cell PC	5.4	2; 2	-80	40	-40	step
prymachina	>100	HEK	[184]	przejściowa	23	whole cell PC	94	2; 2	-80	40	-40	step
raloksyfen	1.8	HEK	[217]	stabilna	?	whole cell PC	5.4	1; 0	-80	20	-50	step
raloksyfen	1.1	HEK	[217]	stabilna	?	whole cell PC	5.4	1; 0	-80	20	20	step
siarczan IV estronu	>0.01	HEK	[156]	stabilna	22	perforowany PC	4.8	2; 2	-80	20	-40	step
sotalol	515.5	CHO	[259]		23	whole cell PC	5	3; 0	-80	20	20	step
sparfloksacyna	2.69	HEK	[122]		22	whole cell PC	5.4	2; 2	-80	20	-40	step
SSR149744C	10	CHO	[112]		?	whole cell PC	4	?				step
terfenadyna	0.0641	CHO	[259]		23	whole cell PC	5	3; 0	-80	20	20	step
UK-062711-13	0.003	XO	[260]		23	voltage clamp 2- elektrodowy	2	5; 0.4	-90	0	-70	step
UK-064962-01	0.273	XO	[260]		23	voltage clamp 2- elektrodowy	2	5; 0.4	-90	0	-70	step
UK-064963-01	0.329	XO	[260]		23	voltage clamp 2- elektrodowy	2	5; 0.4	-90	0	-70	step
UK-068097	0.922	XO	[260]		23	voltage clamp 2- elektrodowy	2	5; 0.4	-90	0	-70	step

Substancja	IC50 [ $\mu$ M]	model	źródło <sup>1</sup>	transf <sup>2</sup>	temp. [°C]	technika <sup>3</sup>	K <sup>+</sup> <sup>4</sup> [mM]	t <sub>1</sub> ;t <sub>2</sub> puls [s]	hold <sup>5</sup> [mV]	depol <sup>6</sup> [mV]	pomiar <sup>7</sup> [mV]	protokół <sup>8</sup>
UK-068522	2.59	XO	[260]		23	voltage clamp 2- elektrodowy	2	5; 0.4	-90	0	-70	step
UK-068523	0.081	XO	[260]		23	voltage clamp 2- elektrodowy	2	5; 0.4	-90	0	-70	step
UK-068526-01	0.044	XO	[260]		23	voltage clamp 2- elektrodowy	2	5; 0.4	-90	0	-70	step
W-13 [N-(4- aminobutylo)-5-chloro- 2-naftaleno sulfonamid HCl]	19.8	HEK	[408]	stabilna	22.5	whole cell PC	5	2.5; 2.5	-80	30	-50	step
W-7 [N-(6- aminoheksylo)-5- chloro-1-naftaleno sulfonamid]	3.5	HEK	[408]	stabilna	22.5	whole cell PC	5	3; 3	-80	30	-50	step
Z-338 [akotiamid]	>10	HEK	[111]		37	whole cell PC	4	0.75; 0.75	-70	0	-50	step

<sup>1</sup> oryginalna publikacja

<sup>2</sup> rodzaj transfekcji komórek

<sup>3</sup> technika pomiarowa zastosowana w doświadczeniu

<sup>4</sup> stężenie jonów potasowych w płynie omywającym komórki

<sup>5</sup> potencjał utrzymania

<sup>6</sup> wartość napięcia depolaryzującego

<sup>7</sup> napięcie pomiarowe

<sup>8</sup> rodzaj protokołu napięć

### 3. MODELE

#### 3.1. Metodyka

##### 3.1.1. Oprogramowanie i sprzęt komputerowy

###### a. Oprogramowanie wykorzystane do tworzenia modeli

Podstawowe oprogramowanie wykorzystane w pracy stanowiły specjalistyczne narzędzia umożliwiające zastosowanie różnego rodzaju technik uczenia maszynowego: system Nets2010 [229] oraz WEKA 3.6 [138].

###### Nets2010

Oprogramowanie napisane w języku Borland Delphi<sup>®</sup> 5.0 i kompilowane w pakiecie Lazarus IDE umożliwia symulację, testowanie i analizę sztucznych sieci neuronowych na komputerze PC pod kontrolą środowiska Windows lub Linux [229]. Symulator *Nets2010* w wersji 64-bitowej może przetwarzać gigabajtowe zbiory danych zawierające od jednego do kilkunastu milionów rekordów w zależności od wielkości wektora wejściowego. Dane wsadowe przygotowane są w formie pliku tekstowego, w którym wartości kolejnych zmiennych oddzielane są znakiem tabulacji. Każdy rekord uczący zapisany jest w osobnej linii.

Graficzny interfejs użytkownika (GUI) pozwala na szybkie zdefiniowanie architektury sieci neuronowej i parametrów jej uczenia. System *Nets2010* oferuje również możliwość pracy w trybie wsadowym tj. treningu i analizowania zestawów sieci bez ingerencji użytkownika. W tym celu konieczne jest odpowiednie przygotowanie zbiorów uczących i testowych oraz stworzenie plików konfiguracyjnych determinujących parametry uczenia sieci neuronowych.

Opcje programu umożliwiają wybranie:

- typu sieci – perceptron wielowarstwowy (MLP), sieć z kontrapropagacją lub sieć rozmyta;
- algorytmu uczenia – propagacja wsteczna (BP) z modyfikacjami takimi jak: technika momentu, zanikające współczynniki uczenia, algorytm delta-bar-delta; algorytm Kohonena, warstwy kontekstu, sieci hybrydowe;
- architektury sieci – liczba komórek wejściowych i wyjściowych, liczba warstw ukrytych i budujących je komórek;
- ilość iteracji, wartość epoki;
- parametry uczenia i funkcja aktywacji –  $\delta$ ,  $\alpha$ , funkcja sigmoidalna,  $a*x$ ,  $a/x$ ,  $\ln(|x|)$ ,  $\sin(x)$ ,  $\cos(x)$ ,  $\tanh(x)$ , funkcja fsr wg Bilskiego [22]; istnieje możliwość zdefiniowania indywidualnych parametrów dla każdej warstwy;
- funkcji kosztu – funkcja kwadratowa, logarytmiczna lub funkcja Hampela;

- modyfikacje metody defuzyfikacji w uczeniu sieci neuronowo-rozmytych propagacją wsteczną (COA, MOM, M-SLIDE).

Ponadto program *Nets2010* posiada możliwość wykonania analizy wrażliwościowej sieci typu MLP. Zaimplementowano również opcję gazu neuronowego dla uczenia sieci z kontrapropagacją i kryterium Fishera redukcji ilości reguł rozmytych.

Aktualnie dostępne są następujące algorytmy uczenia sieci:

- propagacja wsteczna i jej modyfikacje (momentum, delta-bar-delta),
- kontrapropagacja (algorytm WTM oraz gaz neuronowy),
- lokalne przestrajanie,
- modyfikacje metody defuzyfikacji w uczeniu sieci rozmytych z propagacją wsteczną (COA, MOM, M-SLIDE),
- modyfikacja propagacji wstecznej dla sieci częściowo rekurencyjnych budowanych za pomocą tzw. warstw kontekstu (sieci typu Jordana-Elmana).

Simulator *Nets2010* jest przystosowany do pracy w trybie tekstowym, dzięki czemu możliwe jest rozpraszanie obliczeń w sieci komputerowej, co znacznie przyspiesza uczenie zestawu architektur sieci neuronowych.

### WEKA 3.6

WEKA (Waikato Environment for Knowledge Analysis) [390] jest zbiorem algorytmów uczenia maszynowego przeznaczonych do drążenia danych (*ang. data mining*) zaimplementowanym w języku Java. System Weka złożony jest z czterech modułów:

- Explorer - zawiera narzędzia do wstępnego przetwarzania danych (*data preprocessing*), ich klasyfikacji, regresji, klasteryzacji, asocjacji i wizualizacji, w tym drzewa decyzyjne i algorytmy oparte o logikę bayesowską,
- Experimenter – środowisko do projektowania eksperymentów i przeprowadzania testów statystycznych pomiędzy różnymi schematami uczenia,
- KnowledgeFlow – oferuje funkcjonalności podobne do Explorera wraz z interfejsem typu przeciągnij i upuść (*ang. drag-and-drop*), umożliwia uczenie inkrementalne,
- Simple CLI (*ang. simple command-line interface*) – interfejs linii komend.

Weka jest oprogramowaniem typu Open Source udostępnianym w ramach licencji publicznej GNU [138].



Weka przetwarza dane prezentowane w formie tablicy relacyjnej. Natywnym formatem odczytywanych plików jest ARFF (*ang. attribute related file format*). Dane uczące i testowe wykorzystywane przez algorytmy systemu muszą więc być specjalnie przygotowane w tym formacie. Plik ARFF posiada dwie sekcje: nagłówek oraz dane. Nagłówek zawiera nazwę zadania, relacji i deklaracje atrybutów w postaci uporządkowanej sekwencji: @attribute <nazwa atrybutu><typ danych>. Weka obsługuje 4 typy danych: numeryczne, nominalne, łańcuchowe (string) i daty. Kolejność, w jakiej atrybuty są deklarowane wskazuje pozycje kolumny w sekcji 'dane' pliku. Sekcję zawierającą dane rozpoczyna deklaracja @data. Każdy rekord, przykład uczący zapisany jest w osobnej linii ze znakiem karetki oznaczającym koniec linii. Wartości atrybutów rozdzielane są przecinkami, znak zapytania koduje brakujące wartości parametrów.

b. Inne programy

W badaniach poza systemami uczenia maszynowego zastosowano także specjalistyczne programy konieczne do przygotowania danych czy analizy wyników testowania modeli:

Marvin 5.4.0.1 - ChemAxon Calculator (cxcalc) - kalkulator ChemAxon (cxcalc), jeden z modułów platformy cheminformatycznej Marvin Beans i JChem [137], to uruchamiany za pomocą wiersza poleceń (*ang. command line interface - CLI*) program konsolowy wykonujący obliczenia chemiczne. Wykorzystywane przez program wtyczki (Calculator Plugins) umożliwiają wyznaczenie różnorodnych deskryptorów właściwości fizyko-chemicznych na podstawie struktury chemicznej cząsteczek. Cxcalc pozwala na wyznaczenie prostych właściwości takich jak masa molowa, zawartość procentowa poszczególnych rodzajów atomów oraz bardziej złożonych jak współczynniki podziału czy deskryptory topologiczne czy właściwości elektryczne i magnetyczne [137]. Obliczenia mogą być prowadzone indywidualnie, osobno dla każdej cząsteczki lub w trybie wsadowym. Program pozwala na pracę z cząsteczkami zapisanymi w różnorodnych formatach (m.in. SMILES, mol, sdf). Plik konfiguracyjny niezbędny do uruchomienia obliczeń jest plikiem zgodnym ze standardem plików właściwości dla języka java (.properties).

Marvin 5.4.0.1 – MolConverter – element platformy cheminformatycznej Marvin Beans i JChem uruchamiany za pomocą wiersza poleceń. Umożliwia konwersję różnych formatów zapisu struktury cząsteczek oraz uporządkowanie struktur w przestrzeni trójwymiarowej [137].

Marvin 5.4.0.1 – MarvinSketch – element platformy cheminformatycznej Marvin Beans, edytor umożliwiający rysowanie, edycję struktur chemicznych oraz zapis w różnych formatach.

Eddan3 - arkusz kalkulacyjny do wstępnej obróbki danych dla sieci [229].

Preproc – aplikacja przeznaczona do wstępnego przetwarzania danych; umożliwia m.in. wykonanie normalizacji, binaryzacji, balansowania klas [229].

Classif\_compute – zaawansowany system analizy wyników otrzymanych podczas testowania modeli neuronowych [229].

Fingerprint – narzędzie generujące binarny odcisk cząsteczki (fingerprint) o zdefiniowanej długości i dokładności na podstawie dwuwymiarowej struktury chemicznej kodowanej w formacie sdf [334].

FileConverter – system automatycznego tworzenia par zbiorów testowy-uczący koniecznych w procedurze 10-krotnego wzajemnego sprawdzanie modeli.

Microsoft Excel 2007 – arkusz kalkulacyjny, zastosowany obróbki danych, przygotowania zbiorów uczących i testowych oraz zestawienia i analizy wyników modelowania.

#### c. Sprzęt komputerowy

Obliczenia wykonywane były z wykorzystaniem lokalnej sieci GRID (22 jednostki wieloprocesorowe) pracującej pod kontrolą środowiska Linux. Centralny system zarządzania mocą obliczeniową, stworzony i zarządzany przez dra Aleksandra Mendyka, pozwala zautomatyzować proces przygotowania pakietów, rozpraszania zadań obliczeniowych oraz akwizycji wyników i ich scalania. Dzięki wykorzystaniu możliwości sprzętowych procesorów z rodziny Xeon oraz serwerowych płyt głównych i obudów, a także wspomnianego systemu, całość systemu pracuje w trybie 24/7.

#### 3.1.2. Wektor wejściowy

Zbiory treningowe dla testowanych algorytmów uczenia maszynowego skonstruowano wykorzystując zgromadzone dane eksperymentalne, deskryptory właściwości fizyko-chemicznych wygenerowane przez program Marvin Calculator Plugins (Cxcalc) i tzw. odciski cząsteczek w różnych konfiguracjach.

##### a. dane eksperymentalne

Podstawę budowy modeli stanowiła baza wiedzy zgromadzona i opisana w rozdziale III.1 „Baza danych”. Ostateczny zbiór informacji został poddany optymalizacji podobnie jak w przypadku przygotowywania danych potrzebnych do wyznaczenia współczynników ekstrapolacji (rozdz. III.2 „Współczynniki ekstrapolacji” str. 86). Wykluczono rekordy, w których wartość stężenia IC<sub>50</sub> wyznaczono

niestandardową techniką, a także rekordy odnoszące się do pentamidyny, oznaczeń wykonanych dla form izomerycznych substancji stosowanych równocześnie w postaci racematu oraz dane niepewne (rozdz. III.2 „Współczynniki ekstrapolacji” str. 89). W bazie pozostały wyłącznie rekordy zawierające pełen zestaw informacji (model doświadczalny, temperatura, stężenie jonów potasowych w płynie omywającym komórki, czas trwania kolejnych pulsów oraz wartości potencjału utrzymania, pulsu depolaryzującego, napięcie pomiaru oraz zastosowany rodzaj protokołu). Z pierwotnego zbioru usunięto także rekordy zdublikowane. W przypadku, kiedy dla danej substancji w dwóch lub więcej eksperymentach wszystkie parametry doświadczenia były jednakowe a otrzymane wartości  $IC_{50}$  różne, rekordy takie były kompilowane. Finalna baza zawierała 494 wartości  $IC_{50}$  202 różnych substancji chemicznych. Stworzony został odrębny zbiór rekordów wykorzystywany do zewnętrznej walidacji modeli. Metodyka zastosowana podczas jego budowy była identyczna jak w przypadku pierwotnej bazy danych. Rekordy stanowiące zbiór walidacyjny charakteryzują eksperymenty, których opis został opublikowany po zakończeniu etapu gromadzenia danych uczących lub do których uzyskano dostęp po tym etapie (Tabela 9). Zbiór walidacyjny poddano optymalizacji zgodnie z kryteriami przyjętymi dla natywnej bazy danych. Ostatecznie zawierał on 55 rekordów dla 40 różnych substancji chemicznych, z których 26 (30 rekordów) było nieznanymi systemowi.

#### b. deskryptory fizyko-chemiczne

Jednym z elementów wektora wejściowego stanowiącego zbiór uczący był zestaw deskryptorów generowanych przez program Calculator Plugins (Cxcalc). Procedura ich uzyskania obejmowała:

- zgromadzenie plików kodujących strukturę chemiczną cząsteczek znajdujących się w bazie danych w formacie sdf; większość struktur została odnaleziona w bazie PubChem Compound [139], struktury niezamieszczone w tej bazie zostały narysowane w programie MarvinSketch i zapisane w formacie sdf,
- uporządkowanie przestrzenne (3D) cząsteczek – wykonane za pomocą programu Marvin MolConverter; opcja *Clean3D* tworzy konformery z fragmentów, dla których energia konformeru, mierzona za pomocą prostych funkcji energii jest najniższa,
- obliczenie deskryptorów z wykorzystaniem programu Cxcalc; w tabeli 10 przedstawiono wykorzystane wtyczki; eliminacja deskryptorów nieistotnych (jednakowe wartości dla wszystkich leków),
- wykonanie analizy wrażliwościowej i redukcja liczby deskryptorów.

Tabela 10. Zestawienie wykorzystanych wtyczek programu Cxcalc.

No.	grupa deskryptorów	nazwa wtyczki
1.		formula
2.	analiza elementarna	composition
3.		icomposition
4.		mass
5.		ładunek
6.	konformacja	hasvalidconformer
7.		asa
8.		asymmetricatomcount
9.		balabanindex
10.		bondcount
11.		chiralcentercount
12.		connectedgraph
13.		cyclomaticnumber
14.		dreidingenergy
15.		fragmentcount
16.		hararyindex
17.	geometria	hyperwienerindex
18.		maximalprojectionarea
19.		maximalprojectionradius
20.		minimalprojectionarea
21.		minimalprojectionradius
22.		molecularsurfacearea
23.		plattindex
24.		psa
25.		randicindex
26.		szegedindex
27.		topologyanalysisstable
28.		wienerindex
29.		wienerpolarity
30.		izomery
31.	resonantcount	
32.	struktury Markush'a	enumerationcount
33.	współczynniki podziału	logd
34.		logp
35.	protonizacja	pi
36.		pka
37.		acceptorcount
38.		donorcount
39.	inne	msacc
40.		msdon
41.		refractivity

## c. struktura cząsteczki (fingerprint)

Dwuwymiarowe binarne odciski (2D, hashed lub substructure fingerprint) wykorzystywane są do szybkiego wyboru odpowiednich cząsteczek na podstawie analizy ich podstruktur, klasteryzacji chemicznych baz danych, a także do wyznaczania miar podobieństwa par cząsteczek czy różnorodności chemicznej zbiorów danych [m.in. 383,405]. Pierwsze opisy zastosowania tego rodzaju reprezentacji struktury publikowane były już połowie lat 80' ubiegłego wieku [383]. Od tego czasu zaproponowano wiele nowych sposobów kodowania struktury chemicznej, w tym grafy chemiczne, wskaźniki topologiczne, modele farmakoforów 3D, pola molekularne i in. Proste odciski 2D nadal pozostają jednak najczęściej wybieraną metodą reprezentacji, nie tylko ze względu na wydajność obliczeniową, ale także na udowodnioną w wielu analizach porównawczych efektywność [382].

Odciski molekularne są rodzajem abstrakcyjnej reprezentacji określonych cech cząsteczek. Są to binarne (lub booleanowskie) wektory kodujące obecność (1) lub brak (0) danych atomów lub dwuwymiarowych fragmentów struktury. Istnieje kilka typów *fingerprintów*. Jedną z pierwszych koncepcji to tzw. klucze struktury (*ang. structural keys*). Generacja klucza struktury polega na zdefiniowaniu ograniczonego zbioru pewnych cech struktury (fragmentów), a następnie przełączaniu poszczególnych bitów odpowiadających tym fragmentom w bitmapie. Aprioryczny wybór fragmentów wpływa na szybkość tworzenia fingerprintów oraz przeszukiwania baz danych czy działania algorytmów, które je wykorzystują, a także na ich efektywność. Stąd pojawia się konieczność optymalizacji zbioru wybranych fragmentów. Dwuwymiarowe fingerprinty charakteryzują się większym stopniem ogólności, ponieważ nie wymagają predefiniowania zbioru poszukiwanych w cząsteczkach fragmentów. Wszystkie możliwe podstruktury generowane są na podstawie budowy konkretnej cząsteczki. Z drugiej jednak strony konkretne bity, w przeciwieństwie do kluczy struktury, nie mają ściśle przypisanego znaczenia.

W pracy do wygenerowania fingerprintów wykorzystano biblioteki CDL (Chemical Descriptor Library) [334]. W implementacji CDL cząsteczki przedstawiane są w postaci grafów molekularnych, w których wierzchołki reprezentują atomy, a krawędzie – wiązania. Grafy molekularne CDL są modelem walencyjnym, krawędzie przedstawiają wiązanie kowalencyjne, rząd wiązania jest kodowany jako jego właściwość, a nie przez równoległe krawędzie łączące dane wierzchołki (atomy). Algorytm fingerprint koduje wszystkie sekwencje atomów wzdłuż najkrótszej ścieżki łączącej dwa wskazane atomy cząsteczki. Rozważane są długości ścieżek od 1 wiązania do maksymalnej zdefiniowanej liczby wiązań rozdzielających te atomy (zwykle 7). W celu wygenerowania odcisku algorytm inicjuje każdy element wektora binarnego, o długości określonej przez użytkownika (zwykle 1024) jako 0. Następnie algorytm iteruje po wierzchołkach grafu molekularnego obliczając ścieżki. Uwzględniane są wyłącznie najkrótsze ścieżki, bez rozgałęzień i cykli, co realizowane

jest poprzez zastosowanie algorytmu BFS (*ang. breadth-first search*). Obliczone ścieżki są dalej przekształcane w tablice liczb całkowitych, w których każdy element reprezentuje liczbę atomową odczytaną z poszczególnych wierzchołków ścieżki. W kolejnym kroku funkcja haszująca przekształca tablice w liczby całkowite, podawane następnie do generatora liczb pseudolosowych (PRNG). PRNG generuje liczby z zakresu podanego przez użytkownika jako długość wektora binarnego wskazujące na pozycję bitu, którego wartość należy przełączyć.

### 3.1.3. Wartość wyjściowa

Celem modelowania była klasyfikacja cząsteczek jako potencjalnie bezpieczne i blokujące kanał potasowy hERG. Opisywane w literaturze binarne modele klasyfikacyjne blokowania hERG bazują na różnych wartościach progowych separujących cząsteczki pod względem bezpieczeństwa kardiologicznego. Proponowane wartości progowe wahają się w zakresie od 130 nM do 40  $\mu$ M [346]. Analiza danych literaturowych wykazuje, że cząsteczki o wartości  $IC_{50}$  większej niż 10  $\mu$ M lub 30  $\mu$ M uważane są za bezpieczne, podczas gdy  $IC_{50}$  niższe niż 1  $\mu$ M jest kryterium dyskwalifikacji danego związku [16,172,283,307,346].

W pracy jako wartość progową różnicującą substancje bezpieczne i niebezpieczne wybrano stężenie  $IC_{50}$  wynoszące 1  $\mu$ M. Przy tak ustalonym progu 242 substancje ze zbioru treningowego sklasyfikowano jako silnie blokujące kanał hERG ( $IC_{50} \leq 1$ ) a 233 jako nieaktywne. Według przyjętego kryterium podziału dodatkowy zbiór testowy stanowiło 29 substancji o wysokim potencjale arytmogennym i 26 sklasyfikowanych jako bezpieczne.

### 3.1.4. Budowa modeli – testowane parametry

Modele predykcyjne działające w oparciu o logikę Bayesa oraz drzewa decyzyjne konstruowane były przy pomocy programu WEKA. Podczas badań testowano sieci przekonań Bayesa (BayesNet), naiwne klasyfikatory bayesowskie (NaiveBayes) oraz drzewa decyzyjne, z modyfikacją parametrów algorytmów uczenia i doboru struktury sieci/drzewa. W tabelach 11 i 12 przedstawiono parametry modyfikowalne wybranych algorytmów. Testowano zarówno zmiany pojedynczych parametrów jak i ich kombinacji.

a. algorytmy Bayesa

Tabela 11. Parametry algorytmów Bayesa.

algorytm	parametry	opis
<b>NaiveBayes</b>		standardowy algorytm, z estymatorem jądra lub nadzorowaną dyskretyzacją atrybutów
	useKernelEstimator [K]	dla parametrów numerycznych zamiast rozkładu normalnego wykorzystuje estymator jądra
	useSupervisedDiscretization [SD]	wykorzystuje nadzorowaną dyskretyzację w celu konwersji atrybutów numerycznych na nominalne
<b>BayesNet</b>		algorytm uczenia sieci Bayesa, dostarcza m.in. struktury sieci i warunkowe rozkłady prawdopodobieństwa
	SimpleEstimator	szacuje tabele prawdopodobieństw warunkowych sieci Bayesa bezpośrednio z danych po ustaleniu struktury sieci, współczynnik $\alpha$ może być interpretowany jako początkowa wartość zmiennych (A0.5)
	K2 searchAlgorithm	jeden z algorytmów wspinaczki ( <i>ang. hill climbing</i> ) działających zgodnie ze strategią zachłanną
	initAsNaiveBayes [T/N]	inicjowanie sieci jako NaiveBayes – efektem wyboru tej opcji jest wykorzystanie struktury naiwnej sieci Bayesa (z krawędziami łączącymi węzeł klasyfikujący ze wszystkimi pozostałymi węzłami) jako punktu wyjścia do poszukiwania struktury optymalnej; ustawienie wartości tej opcji jako False powoduje, że sieć początkowa jest siecią pustą
	maxNrOfParents [P 1-100]	pozwała określić maksymalną liczbę rodziców, jaką może posiadać węzeł; w wyniku ustawienia wartości tego parametru jako 1 przy inicjowaniu sieci jako NaiveBayes otrzymuje się naiwny klasyfikator bayesowski; ustawienie jego wartości jako 2 daje w efekcie sieć TAN ( <i>ang. Tree Augmented Bayes Network</i> ), a

algorytm	parametry	opis
		wartości powyżej 2 sieci BAN ( <i>ang. Bayes Net Augmented Bayes Network</i> ); wybór wartości znacznie większych niż liczba węzłów w sieci (zazwyczaj 100000) nie definiuje liczby rodziców
	scoreType	determinuje miarę wykorzystywaną do oceny jakości struktury sieci: Bayes, entropia lub Akaike Information Criterion (AIC)
HillClimber searchAlgorithm		algorytm wspinaczki dodający, usuwający i odwracający kierunek krawędzi sieci. W przeciwieństwie do algorytmu K2 nie jest ograniczony kolejnością zmiennych
	initAsNaiveBayes [T/N]	jw.
	maxNrOfParents [P 1-100]	jw.
	scoreType	jw.
SimulatedAnnealing searchAlgorithm		symulowane wyżarzanie ( <i>ang. simulated annealing</i> ) jest metodą poszukiwania optymalnej struktury sieci łączącą błędnie przypadkowe z tzw. szukaniem przypadkowym; jest to modyfikacja algorytmu wspinaczki, w której dopuszcza się przejście ze stanu bieżącego do stanów gorszych, co w efekcie daje możliwość przebadania znacznie większego obszaru przestrzeni rozwiązań
	TStart [A 1-10]	temperatura początkowa określa prawdopodobieństwo akceptacji zmiany w "złym" kierunku (im wyższa temperatura, tym większe prawdopodobieństwo)
	delta	współczynnik zmiany temperatury w kolejnych iteracjach
	runs [R 10 000 – 100 000]	określa liczbę iteracji realizowanych podczas wyżarzania
	scoreType	determinuje miarę wykorzystywaną do oceny jakości struktury sieci: Bayes, entropia lub Akaike Information Criterion (AIC)
	seed [S 1-100]	wartość początkowa dla generatora liczb losowych; ustawienie wartości seed gwarantuje powtarzalność eksperymentów



b. drzewa decyzyjne

Tabela 12. Parametry algorytmów konstrukcji drzew decyzyjnych.

algorytm	parametry	opis
<b>ADTree</b>		algorytm oparty na procedurze zmiennego drzewa decyzyjnego ( <i>ang. alternating decision tree learning algorithm</i> )
	numOfBoosting Iterations. [B]	liczba iteracji; parametr ten wymaga ręcznego dostosowania do zbioru danych oraz pożądanego współczynnika złożoności systemu/efektywności; im więcej iteracji tym większe i potencjalnie dokładniejsze drzewo, ale jednocześnie wolniejsze działanie algorytmu
	search Path [E]	wybór sposobu budowy drzewa; opcja domyślna „rozwiń wszystkie ścieżki” ( <i>ang. Expand all paths</i> ) gwarantuje wyczerpujące przeszukiwanie przestrzeni drzew, inne zaimplementowane metody („expand the haviest path”, „Expand z-pure path”, „Expanr random path”) są heurystyczne i nie gwarantują znalezienia optymalnego rozwiązania, są jednak dużo szybsze
<b>J48</b>		algorytm generowania przycinanych lub nie drzew decyzyjnych typu C4.5
	binarySplits [B]	wybór binarnego podziału atrybutów nominalnych
	confidenceFactor [C]	współczynnik ufności dla funkcji przycinającej drzewo; mniejsza wartość wymusza mocniejsze przycinanie
	minNumObj [M]	ustawienie minimalnej liczby rekordów w liściach drzewa
	numFolds [F]	określa liczbę danych używanych do przycinania reduced-error pruning. jednokrotność zbioru uczącego wykorzystywana jest do przycinania, pozostałe krotności do rozbudowy drzewa
	reducedErrorPruning [R]	wybór procedury przycinania: <i>reduced-error pruning</i> lub C.4.5
	subtreeRaising [S]	pozwała na budowę poddrzewa podczas przycinania
unpruned [U]	opcja wyłączenia przycinania drzewa	

algorytm	parametry	opis
<b>RandomTree</b>		algorytm konstrukcji drzew decyzyjnych, który w każdym węźle rozważa K losowo wybranych atrybutów, przycinanie wygenerowanego drzewa nie jest możliwe
	KValue [K]	wybór liczby atrybutów w węźle
	maxDepth [depth]	ustawienie maksymalnej wielkości drzewa
	minNum [M]	określenie minimalnej liczby przypadków w liściach
<b>RandomForest</b>		algorytm konstruuje lasy z drzew losowych
	maxDepth [depth]	ustawienie maksymalnej wielkości drzew
	numFeatures [K]	liczba losowo wybieranych do węzła atrybutów
	numTrees [I]	liczba drzew

## c. sieci neuronowe

Modele predykcyjne, których podstawę stanowiły sztuczne sieci neuronowe tworzone były przy pomocy programu Nets2010. Przetestowano ponad 2000 architektur (ok. 130 000 sieci neuronowych) uczonych metodą wstecznej propagacji błędu (BP) oraz sieci działających w oparciu o logikę rozmytą [417]. W procesie trenowania sieci BP zastosowano technikę momentu i algorytm delta-bar-delta [417]. Sieci typu MLP (*ang. MultiLayer Perceptron*) posiadały od 1 do 6 warstw ukrytych. Testowano trzy rodzaje funkcji aktywacji: tangens hiperboliczny, fsr [22] oraz sigmoidę unipolarną. Wszystkie komórki danej sieci posiadały identyczną funkcję aktywacji. Komórki w warstwach sąsiadujących połączone zostały zgodnie z zasadą „każdy z każdym”. Każda sieć trenowana była przez 10 000 do 5 000 000 iteracji. Wartość epoki wynosiła 1. Współczynnik momentu  $\alpha$  wynosił 0.3, a początkowa wartość  $\delta$  była równa 0.65. Zastosowano także technikę jog-of-weights z kryterium cierpliwości 100 000 iteracji [253]. Preprocessing danych treningowych dla sieci neuronowych obejmował skalowanie do zakresu 0.2-0.8 lub -0.8-0.8 (scale) i/lub zaszumianie (rand). Wprowadzenie szumu do wzorców uczących mające na celu zwiększenie zdolności uogólniania sieci, polegało na multiplikacji rekordów zbioru treningowego z dokładaniem szumu o rozkładzie jednostajnym i amplitudzie  $\pm 5\%$  wartości danej zmiennej. W przypadku sieci rozmytych do automatycznego tworzenia architektury wybrano metodę tablic przeglądowych (*ang. lookup table*) [218], defuzyfikacji dokonywano metodą M-SLIDE. Ciągłe wartości wyjść podawane przez wytrenowaną sieć binaryzowano, jako próg pobudzenia stosując środek przedziału pobudzeń.

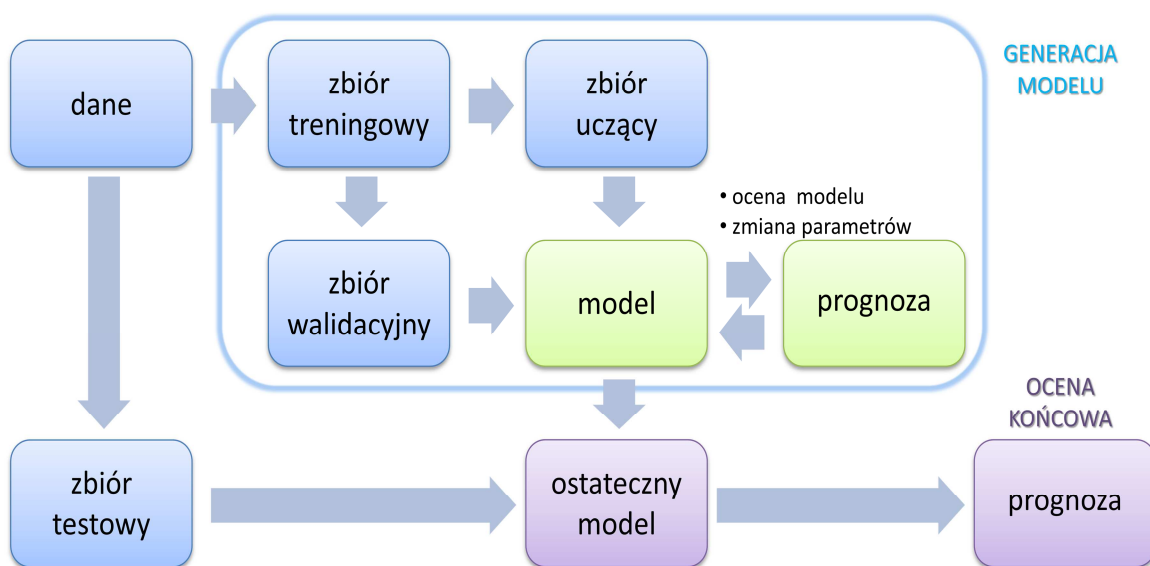
W pracy oprócz modeli budowanych w oparciu o pojedyncze architektury testowano także systemy modularne, tzw. komitety ekspertów utworzone ze zbioru wyselekcjonowanych modeli. Poszczególne moduły wchodzące w skład komitetu mogą różnić się pomiędzy sobą zarówno strukturą jak i technikami zastosowanymi do przetwarzania danych, proces uczenia każdego z nich może być przeprowadzony niezależnie. Zakłada się, że połączenie różnych sieci neuronowych czy modeli budowanych w oparciu o różne techniki uczenia maszynowego może zwiększyć dokładność klasyfikacji w stosunku do pojedynczej sieci/algorytmu. Związane jest to wzajemnym uzupełnianiem się poszczególnych modułów komitetu, w procesie przetwarzania zwracanych przez nie wyników w bloku interpretera. Zaproponowanych zostało wiele metod łączenia klasyfikatorów, wśród których najczęściej stosowane jest głosowanie większościowe i uśrednianie [242].

## 3.1.5. Weryfikacja modeli

## a. zbiory testowe

Wytrenowany model powinien charakteryzować się zdolnością uogólniania zdobytej wiedzy. Wiarygodna ocena tej właściwości modelu wymaga, aby do jego budowy i testowania zastosowane zostały różne zbiory danych. Zapobiega to również nadmiernemu dopasowaniu parametrów modelu do danych i ograniczeniu jego zdolności generalizacji. Proces generowania modelu powinien więc w optymalnych warunkach wyglądać następująco (Rycina 28):

- podział zbioru danych na zbiór treningowy i testowy,
- podział zbioru treningowego na uczący i walidacyjny,
- budowa modelu na podstawie danych uczących,
- walidacja modelu na zbiorze walidacyjnym, optymalizacja parametrów modelu,
- opcjonalnie – wytrenowanie modelu na całym zbiorze treningowym,
- ocena efektywności ostatecznego modelu z wykorzystaniem zbioru testowego



Rycina 28. Proces budowy modelu – podział danych.

Taka procedura budowy modelu, w której do estymacji miar oceny sprawności klasyfikatora stosuje się osobny zbiór testowy (*ang. hold-out*), wymaga jednak dużej liczby (przyjmuje się, że więcej niż kilka tysięcy) przykładów w zbiorze danych [327]. W przypadku realnych problemów pradykcyjnych zwykle dostępny jest ograniczony zbiór danych, który traktuje się jako próbę losową pobraną z populacji ogólnej możliwych przykładów. Dla takich zbiorów w celu określenia jakości klasyfikacji modelu oraz doboru jego parametrów stosuje się podziały wielokrotne i przeprowadza ocenę dla każdego podziału.

Podczas prezentowanych w tej pracy badań zastosowano dwa podejścia do oceny i porównania sprawności modeli: walidację wewnętrzną i zewnętrzną (testowanie).

Walidacja wewnętrzna miała na celu wybranie najbardziej optymalnych parametrów i wytypowanie najlepszych modeli. Została ona przeprowadzona zgodnie z najczęściej stosowaną procedurą 10-krotnego wzajemnego sprawdzania (*ang. 10-fold cross-validation*). Zmodyfikowano jednak sposób przygotowania par treningowych. Standardowa procedura przewiduje losowy wybór około 10% rekordów do zbioru testowego, co powoduje, że w zbiorze testowym i uczącym potencjalnie mogą znaleźć się rekordy niemal identyczne, dla tych samych cząsteczek. Zbiory walidacyjne wykorzystywane w procedurze 10-krotnego wzajemnego sprawdzania podczas badań zostały natomiast skonstruowane przy pomocy aplikacji FileConverter specjalnie przygotowanej do tego celu. Aplikacja ta dzieli zbiór treningowy na 10 par zbiorów: uczący – walidacyjny. Algorytm wybierający rekordy przeznaczone do zbioru testowego rozwiązuje problem wieloplecakowy (*ang. multiple knapsack problem*) przydzielając do każdego zbioru maksymalnie 10% rekordów zbioru treningowego. W pierwszym etapie podziału algorytm dzieli zbiór wg leków i zlicza liczbę rekordów opisujących każdą z substancji, a następnie łączy powstałe podzbiory rozpoczynając od najliczniejszych. Zastosowana procedura ma na celu zapewnienie różnorodności struktur obecnych w zbiorze testowym oraz podobną liczbę leków. Rekordy wybrane do zbioru walidacyjnego są automatycznie usuwane ze zbioru uczącego. Rezultat działania algorytmu przedstawiono w tabeli 13.

**Tabela 13.** Charakterystyka zbiorów walidacyjnych wykorzystanych w procedurze 10-krotnego wzajemnego sprawdzania.

nr zbioru testowego	liczba rekordów	liczba substancji	substancje
1	54	16	1-(bis(4-fluorofenilo)metylo)piperazyna, amsakryna, bertosamil, chlorpromazyna, fluwoksamina, klaritromycyna, MDL74156, moksifloksacyna, norpropoksyfen, perfenazyna, protryptylina, pyrilamina, sibutramina, sildenafil, sofokarpina, terfenadyna
2	49	13	4,4'-difluorobenzofenon, amitryptylina, chinidyna, flekainid, halofantryna, klomifen, klozapina, kokaina, kwas acetylosalicylowy, lidoflazyna, metoprolol, olanzapine, prochlorperazyne
3	49	16	benzoiloekgonina, ciprofloksacyna, cizapryd, domperidon, EDDP, erytromycylamina, erytromycyna, fenobarbital, gabapentyna, klomipramina, norastemizol, norfluoksetyna, papaweryna, tamoksyfen, wardenafil, wesnarinon
4	49	18	4-aminopirydyna, amiodaron, cyjamemazyna, dofetylid, grepafloksacyna, ibutyliid, mizolastyna, naryngenina, N-desbutylhalofantryna, orfenadryna, pentoksyweryna,

			pilsikainid, prometazyna, propoksyfen, ranolazyna, sparfloksacyna, spironolakton, terikalant
5	49	19	azimilid, chlorfeniramina, cymetydyna, difenhydramina, E4031, fentanyl, kokaetylen, LAAM, meflochina, mianseryna, neferyna, oksykodon, pentamidyna, prenylamina, propranolol, tadalafil, topiramát, trazodon
6	49	21	AF3013, ajmalina, ambasilid, cibenzolina, desmetylastemizol, desmetylerytromycyna, dronedaron, flunaryzyna, haloperidol, lamotrygina, lidokaina, lomefloksacyna, maprotylina, mezorydazyna, MK499, morfina, nikotyna, perheksylina, pimozyd, risperidon, ropinirol
7	49	22	AVE0118, acehityzyna, bepridil, berberyina, cząsteczka B, desbutylumefantryna, digoxyna, diltiazem, doksepina, ICI118551, indometacyna, josamycyna, kodeina, kwetiapina, L-702958, metadon, metylekgonidyna, mikonazol, soforydyna, terazosyna, werapamil, ziprasidon
8	49	24	4,4'-difluorobenzhydrol, astemizol, buprenorfina, chlorochina, cząsteczka A, digitoxyna, dizopyramid, DW286a, estradiol, fenytoina, granisetron, kapsaicyna, karbamazepina, ketanseryna, lewkromakalim, lumefantryna, MCI-154, nikorandyl, propafenon, roksitromycyna, sertindol, sotalol, teobromina, terodilina
9	49	26	artycaina, BRL32872, bupiwakaina, cetyryzyna, doksazosyna, feksofenadyna, gatifloksacyna, GF109203X_(bisindolilmaleimid_I), hydroksyzyna, imipramina, irbesartan, kaptopryl, klobutinol, lowastatyna, meperydyna, nifedypina, nifekalant, ondansetron, ouabaina, prazosyna, prokainamid, propiweryna, sulfametoksazol, telitromycyna, tiorydazyna, trifluoperazyna
10	48	28	4,4'-dimetylodifenhydramina, 5-hydroksypropafenon, amoksycylina, apomorfina, atenolol, AZD7009, budypina, citalopram, dolasetron, droperidol, epinastyna, fluoksetyna, hydroksyhaloperidol, karwedilol, ketokonazol, klemastyna, kwas kanrenowy, loratadyna, mepiwakaina, metoklopramid, metyloester ekgoniny, ofloksacyna, oleandomycyna, pentobarbital, pergolid, prukalopryd, ropiwakaina, tolterodyna

Walidacja zewnętrzna miała na celu ocenę zdolności predykcyjnych modeli wybranych na podstawie wyników procedury 10-ktornego wzajemnego sprawdzania. Walidacja zewnętrzna wykonana została z zastosowaniem specjalnie przygotowanego zbioru testowego, w którym znalazły się rekordy niewykorzystywane podczas budowy modelu (Tabela 9, str 97). Dane użyte do testów zawierały zarówno rekordy opisujące nowe doświadczenia wykonane dla przedstawianych wcześniej systemowi substancjach (25 rekordów; 13 cząsteczek), jak i dla nowych, nieznanymi systemowi struktur (30 rekordów; 26 cząsteczek).

## 3.1.6. Ocena błędu

Sztuczne sieci neuronowe jako wyjście podają wartości ciągłe, dlatego kontrola procesu uczenia oraz ocena błędu na zbiorze treningowym i testowym opierała się na wartości błędu średniokwadratowego (*ang. Root Mean Squared Error - RMSE*):

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^j (z^{(j)} - y^{(j)})^2}$$

Równanie 29.

gdzie:

$n$  - ilość rekordów zbioru uczącego

$j$  - numer rekordu danych

$z^{(j)}$  - wartość podawana przez sieć

$y^{(j)}$  - wartość oczekiwana wyjścia

Ogólna trafność modeli klasyfikacyjnych (sieci Bayesa, drzewa decyzyjne, ANN) oceniana była poprzez porównanie wartości przewidzianych przez model z rzeczywistymi, a więc odsetek odpowiedzi prawidłowych i odsetek odpowiedzi błędnych.

$$\eta = \frac{n_T}{n} \text{ lub } \varepsilon = \frac{n - n_T}{n}$$

Równanie 30.

gdzie:

$\eta$  - trafność klasyfikacji (*ang. accuracy*)

$\varepsilon$  - błąd klasyfikacji

$n$  - liczba rekordów w zbiorze testowym

$n_T$  - liczba rekordów sklasyfikowanych poprawnie

Wyniki klasyfikacji przedstawione są w postaci tzw. tablic kontyngencji (macierzy pomyłek). Dla modelu binarnego istnieją cztery możliwości odpowiedzi. Macierz pomyłek dla takiego modelu przedstawiono w tabeli 14.

Jak wcześniej wspomniano, w przypadku modeli dających na wyjściu wartości ciągłe przyjęty standardowy próg dyskryminacji, przy którym dokonywano oceny wyników wynosił 0.5 (50% wartości pobudzenia).

Tabela 14. Macierz pomyłek.

		wartość rzeczywista	
		1	0
wartość przewidziana	1	TP	FP
	0	FN	TN

**TP** (ang. *true positive*) – liczba poprawnie sklasyfikowanych przykładów z wybranej klasy  
(1 sklasyfikowane jako 1)

**FN** (ang. *false negative*) – liczba błędnie sklasyfikowanych przykładów z tej klasy  
(1 sklasyfikowane jako 0)

**TN** (ang. *true negative*) – liczba poprawnie sklasyfikowanych przykładów z drugiej klasy  
(0 sklasyfikowane jako 0)

**FP** (ang. *false positive*) – liczba błędnie sklasyfikowanych przykładów klasy drugiej  
(0 sklasyfikowane jako 1)

Na podstawie macierzy pomyłek wyliczane są następujące wskaźniki jakości klasyfikatora binarnego:

- trafność całkowita (AG lub AG%) - odsetek poprawnie sklasyfikowanych rekordów względem całości zbioru testowego

$$AG = \frac{TP + TN}{TP + FP + FN + TN} \quad \text{Równanie 31.}$$

- czułość (SE)

$$SE = \frac{TP}{TP + FN} \quad \text{Równanie 32.}$$

- specyficzność (SP)

$$SP = \frac{TN}{FP + TN} \quad \text{Równanie 33.}$$

- wartość predykcyjna dodatnia (PPV)

$$PPV = \frac{TP}{FP + TP} \quad \text{Równanie 34.}$$

- wartość predykcyjna ujemna (NPV)

$$NPV = \frac{TN}{FN + TN} \quad \text{Równanie 35.}$$

- AUROC – pole pod krzywą charakterystyki odbiorcy - ROC (ang. *the area under a receiver operating characteristic/curve*). Krzywa ROC, przedstawiająca zależność czułości systemu od jego specyficzności służy do oceny poprawności działania klasyfikatorów binarnych. W optymalnym przypadku AUROC wynosi od 0.9 do 1.0. Parametr wykorzystany podczas selekcji sieci neuronowych biorących udział w analizie wrażliwościowej.



### 3.1.7. Analiza wrażliwościowa

Analizę wrażliwościową wykonano w celu redukcji rozmiaru wektora wejściowego dla algorytmów uczenia maszynowego i identyfikacji parametrów istotnych dla efektywnego przewidywania przez te systemy klasy (bezpieczny/niebezpieczny), do której należy testowana substancja. Analizę przeprowadzono na modelach neuronowych po wstępnym etapie uczenia przeprowadzonym z wykorzystaniem całego zbioru treningowego. Zastosowana procedura polegała na wytrenowaniu zestawu sieci na pełnowymiarowym wektorze wejściowym złożonym z danych eksperymentalnych oraz deskryptorów wygenerowanych przy użyciu programu Cxcalc Marvin, a następnie wyborze najlepszych sieci, które wykorzystane były do przeprowadzenia analizy wrażliwościowej. Kryterium doboru było AUROC (wartość pola powierzchni pod krzywą ROC) obliczane dla zbioru uczącego. Za graniczną wartość uznawano AUROC równe 90% maksymalnej wartości uzyskanej dla trenowanego zbioru sieci. Z tego powodu liczba wybranych modeli nie była ściśle określona a priori i wahała się od kilku do nawet kilkudziesięciu sieci neuronowych. Ostateczny ranking istotności wejść jest średnią wszystkich wybranych modeli. Redukcja wejść została przeprowadzona według metody opracowanej przez Żuradę [417] bazującej na analizie składników wrażliwości  $S_{ki}$  określających wpływ poszczególnych wejść na wartość wyjściową sieci.

$$S_{ki} = \frac{\partial y_k}{\partial x_i} \quad \text{Równanie 36.}$$

gdzie:

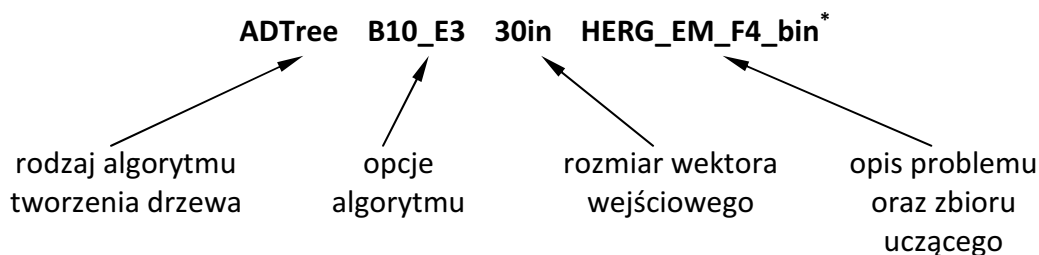
$S_{ki}$  – wrażliwość k-tego wyjścia względem i-tego wejścia, współczynnik ważności  
 $y$  – wartość wyjściowa  
 $x$  – wartość wejściowa

W czasie analizy tworzony jest ranking ważności wejść. W przypadku gdy pomiędzy dwoma kolejnymi wejściami widoczna jest znaczna różnica wartości współczynników można dokonać eliminacji części wejść. Po zredukowaniu wymiaru wektora wejściowego proces uczenia został powtórzony.

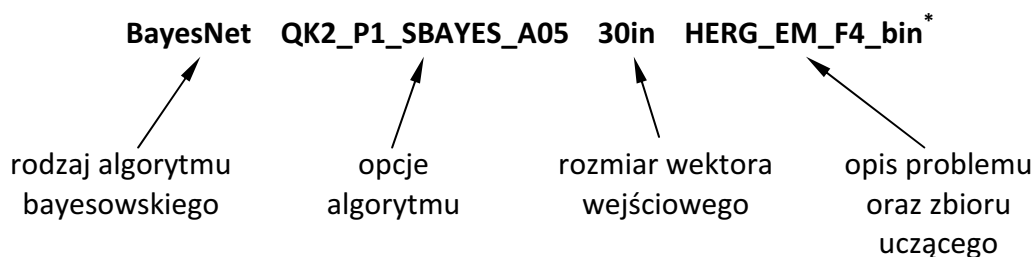
### 3.1.8. Notacja

Znaczna liczba testowanych podczas pracy modeli wymagała opracowania jednolitego systemu ich opisu. Opis poszczególnych algorytmów znajduje się w rozdziale I.5 – „Techniki modelowania”, str. 29.

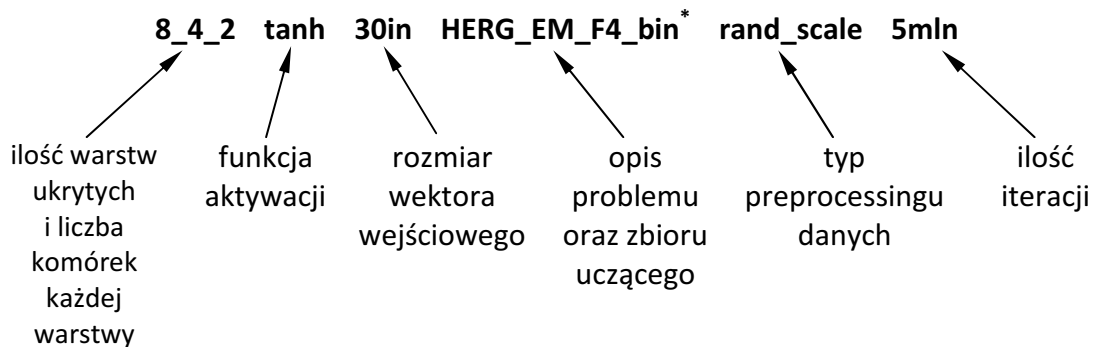
Notacja modeli skonstruowanych z wykorzystaniem drzew decyzyjnych:



Notacja modeli, których podstawę stanowiły algorytmy bayesowskie:



Notacja modeli zbudowanych w oparciu o sztuczne sieci neuronowe:



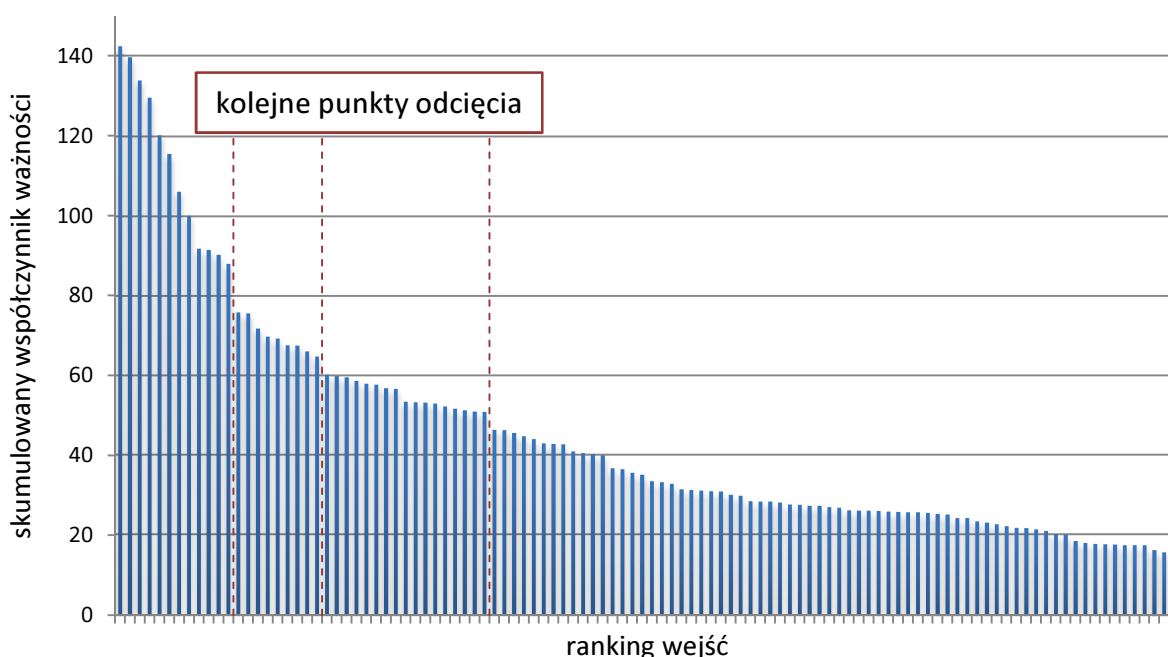
\* - blokowanie kanału potasowego hERG (HERG), dane binarne (bin), rekordy uczące opisane za pomocą deskryptorów charakteryzujących warunki doświadczalne (E), deskryptorów struktury cząsteczki wygenerowanych przez program Marvin (M), binarnego odcisku cząsteczki (F) utworzonego w oparciu o fragmenty zawierające maksymalnie 4 wiązania

## 3.2. Wyniki

### 3.2.1. Analiza wrażliwości

Wektor wejściowy poddany analizie wrażliwościowej posiadał 108 składowych. 9 z nich opisywało warunki eksperymentalne podczas wyznaczania  $IC_{50}$ , pozostałe 99 to deskryptory wygenerowane przez program Cxcalc.

Analiza rankingu wejść (Rycina 29) pozwoliła określić trzy punkty odcięcia pozwalające na ograniczenie pierwotnego wektora wejściowego odpowiednio do 38, 21 oraz 12 wejść. Liczbę deskryptorów charakteryzujących poszczególne rekordy kodowano w nazwie modelu jako 12in, 21in oraz 38in.



Rycina 29. Ranking wejść na podstawie wyników analizy wrażliwościowej.

Na podstawie wyników analizy wrażliwości przygotowano trzy uproszczone, opisywane za pomocą zredukowanego wektora wejściowego, zbiory danych. Dla każdego zbioru stworzonych zostało ponad 360 architektur sieci neuronowych. Sieci trenowane były przez 5 000, 10 000, 50 000, 100 000, 500 000, 1 000 000, 2 000 000 lub 5 000 000 iteracji. Uzyskane wyniki testowania najlepszych architektur sieci w procedurze 10-krotnego wzajemnego sprawdzania przedstawiono w tabelach 15 - 17.

Tabela 15. Wyniki uczenia sieci z wykorzystaniem 12 elementowego wektora wejściowego – 10 najlepszych architektur.

Architektura	iteracje	AG	SE	SP	PPV	NPV	AUROC
10_sigma_12in_HERG_bin_orig	500 000	78.11	0.89	0.67	0.75	0.84	0.85
15_fsr_12in_HERG_bin_rand_scale	200 000	77.70	0.82	0.73	0.77	0.78	0.85

5_sigma_12in_HERG_bin_rand	200 000	77.48	0.89	0.64	0.74	0.84	0.85
5_tanh_12in_HERG_bin_rand	500 000	76.89	0.82	0.71	0.76	0.78	0.84
5_tanh_12in_HERG_bin_rand_scale	500 000	76.35	0.82	0.70	0.75	0.78	0.80
10_tanh_12in_HERG_bin_rand	1 000 000	75.94	0.80	0.71	0.76	0.76	0.84
5_3_sigma_12in_HERG_bin_orig	500 000	75.53	0.82	0.69	0.75	0.77	0.84
5_tanh_12in_HERG_bin_orig_scale	10 000	75.07	0.81	0.69	0.74	0.76	0.84
3_2_tanh_12in_HERG_bin_orig_scale	100 000	74.47	0.80	0.69	0.74	0.75	0.78
3_2_tanh_12in_HERG_bin_rand	50 000	74.18	0.90	0.57	0.70	0.83	0.83
<b>średnia</b>		<b>76.17</b>	<b>0.84</b>	<b>0.69</b>	<b>0.75</b>	<b>0.79</b>	

**Tabela 16.** Wyniki uczenia sieci z wykorzystaniem 21 elementowego wektora wejściowego – 10 najlepszych architektur.

Architektura	iteracje	AG	SE	SP	PPV	NPV	AUROC
10_sigma_21in_HERG_bin_rand	200 000	77.01	0.87	0.67	0.74	0.82	0.84
5_3_sigma_21in_HERG_bin_rand	300 000	77.01	0.80	0.74	0.78	0.77	0.82
5_3_2_tanh_21in_HERG_bin_orig_scale	1 500 000	77.01	0.78	0.76	0.79	0.75	0.78
10_tanh_21in_HERG_bin_orig	750 000	76.92	0.80	0.74	0.77	0.76	0.81
5_tanh_21in_HERG_bin_rand	300 000	76.83	0.81	0.73	0.77	0.77	0.82
15_fsr_21in_HERG_bin_rand_scale	100 000	76.18	0.79	0.74	0.77	0.75	0.82
5_3_2_tanh_21in_HERG_bin_rand_scale	2 000 000	76.09	0.80	0.72	0.76	0.76	0.79
5_tanh_21in_HERG_bin_rand_scale	10 000	76.05	0.81	0.71	0.76	0.77	0.81
5_3_2_sigma_21in_HERG_bin_rand	500 000	76.05	0.77	0.76	0.78	0.74	0.82
5_sigma_21in_HERG_bin_rand	50 000	75.99	0.87	0.64	0.73	0.81	0.83
<b>średnia</b>		<b>76.51</b>	<b>0.81</b>	<b>0.72</b>	<b>0.76</b>	<b>0.77</b>	

**Tabela 17.** Wyniki uczenia sieci z wykorzystaniem 38 elementowego wektora wejściowego – 10 najlepszych architektur.

Architektura	iteracje	AG	SE	SP	PPV	NPV	AUROC
15_7_5_tanh_38in_HERG_bin_rand_scale	1 000 000	78.91	0.84	0.73	0.78	0.81	
3_2_tanh_38in_HERG_bin_orig_scale	1 000 000	78.30	0.79	0.77	0.80	0.77	
5_3_2_tanh_38in_HERG_bin_rand_scale	500 000	78.26	0.79	0.77	0.80	0.77	
15_7_5_fsr_38in_HERG_bin_rand_scale	50 000	78.07	0.80	0.76	0.79	0.77	
5_3_2_sigma_38in_HERG_bin_rand	200 000	78.07	0.81	0.75	0.79	0.78	
15_7_5_sigma_38in_HERG_bin_orig	200 000	78.05	0.79	0.77	0.79	0.77	
5_3_tanh_38in_HERG_bin_orig_scale	1 500 000	77.68	0.80	0.75	0.78	0.77	
3_2_sigma_38in_HERG_bin_orig	300 000	77.66	0.79	0.76	0.79	0.77	
3_2_tanh_38in_HERG_bin_orig	100 000	77.65	0.80	0.75	0.78	0.77	
15_7_5_tanh_38in_HERG_bin_orig_scale	100 000	77.48	0.81	0.73	0.77	0.78	
<b>średnia</b>		<b>78.01</b>	<b>0.80</b>	<b>0.75</b>	<b>0.79</b>	<b>0.78</b>	

Zbiór deskryptorów opisujących warunki doświadczalne oraz właściwości fizykochemiczne cząsteczek wybranych w procesie analizy wrażliwości przedstawia tabela 18.

Najlepsze wyniki uzyskano dla sieci trenowanych z wykorzystaniem rekordów opisanych 38 elementowym zbiorem deskryptorów. Sieci uczone na tym zbiorze charakteryzowały się lepszą zdolnością klasyfikacji w porównaniu zarówno z tymi uczonymi na danych opisanych zredukowanymi wektorami wejść (12 oraz 21 deskryptorów).

Dodatkowo sieci 38in, które charakteryzowały się najwyższą zdolnością generalizacji uczono z wykorzystaniem zbioru danych zawierającego rekordy opisane za pomocą pełnego 108-elementowego zestawu deskryptorów. Wytrenowane sieci przetestowano następnie w procedurze 10-krotnego wzajemnego sprawdzania, aby ocenić wpływ redukcji wymiarowości wektora wejść, opisującego przykłady uczące, na zdolności generalizacyjne. Średnia wartość AG% dla 10 testowanych sieci wyniosła 74.3%. Najwyższy odsetek poprawnych decyzji uzyskano dla sieci 3\_2\_fsr\_108in\_HERG\_bin\_orig (AG% = 76.71). Wyniki te potwierdzają zasadność przyjęcia do dalszych badań trzydziestoosmioelementowego wektora cech. Redukcja wektora wejściowego jest konieczna nie tylko ze względu na wydajność obliczeniową budowanego modelu, ale także dla zapewnienia jego prostoty.

**Tabela 18.** Zmienne kluczowe wyłonione podczas analizy wrażliwościowej.

Nr	nazwa deskryptora	opis
1.	model	komórki wykorzystane w doświadczeniach
2.	temp	temperatura podczas eksperymentu
3.	K+_[mM]_bath	stężenie jonów potasowych w płynie omywającym komórki
4.	t1_puls_[s]	czas pierwszego pulsu
5.	t2_puls_[s]	czas drugiego pulsu
6.	holding_pot	potencjał utrzymania
7.	depol_puls_to	wartość potencjału depolaryzacji
8.	mV_measure	potencjał pomiaru
9.	Aromatic atom count	liczba atomów w cząsteczce posiadającej wiązania aromatyczne
10.	Rotatable bond count	liczba wiązań obrotowych
11.	Heteroaromatic ring count	liczba pierścieni heteroaromatycznych
12.	Smallest ring size	rozmiar najmniejszego pierścienia
13.	Largest ring size	rozmiar największego pierścienia
14.	a(yy)	składowa yy tensora polaryzowalności
15.	ASA-	powierzchnia atomów o ujemnym ładunku cząstkowym dostępna dla rozpuszczalnika
16.	ASA_H	powierzchnia atomów hydrofobowych dostępna dla rozpuszczalnika
17.	ASA_P	powierzchnia atomów polarnych dostępna dla rozpuszczalnika
18.	Asymmetric atom count	liczba atomów asymetrycznych
19.	Balaban index	indeks konektywności Balabana
20.	Chiral center count	liczba centrów chiralnych
21.	Dreiding energy	energia związana ze stabilnością danego konformeru cząsteczki
22.	Maximal projection area	maksymalna powierzchnia projekcji konformeru (3D) na podstawie promienia van der Waalsa
23.	Maximal projection radius	promień maksymalnej powierzchni projekcji konformeru
24.	Minimal projection area	minimalna powierzchnia projekcji konformeru na podstawie promienia van der Waalsa
25.	Minimal projection radius	promień minimalnej powierzchni projekcji konformeru
26.	stereoisomercount	liczba stereoizomerów cząsteczki
27.	pH=0.00(logd)	logD przy pH = 0
28.	pH=6.00(logd)	logD przy pH = 6
29.	pH=9.00(logd)	logD przy pH = 9
30.	pH=10.00(logd)	logD przy pH = 10
31.	pH=11.00(logd)	logD przy pH = 11
32.	logP	logP

33.	bpKa1	I stała dysocjacji zasadowej (-log[K <sub>a</sub> ])
34.	bpKa2	II stała dysocjacji zasadowej (-log[K <sub>a</sub> ])
35.	donorcount	liczba donorów wiązań wodorowych
36.	pH=8.00(msdon)	liczba donorów wiązań wodorowych formy dominującej - pH=8
37.	pH=9.00(msdon)	liczba donorów wiązań wodorowych formy dominującej - pH=9
38.	pH=14.00(msdon)	liczba donorów wiązań wodorowych formy dominującej - pH=14

### 3.2.2. Modelowanie

Na podstawie wyników analizy wrażliwościowej przygotowano 3 rodzaje zbiorów treningowych dla wybranych algorytmów konstruujących modele. Pierwszy zawierał rekordy opisane wyłącznie zestawem deskryptorów wyłonionych w toku analizy wrażliwościowej (38 wejść), w drugim i trzecim dodatkowo dołączono opis struktury cząsteczki w postaci binarnych odcisków (fingerprint) o długości i wielkości fragmentów (maksymalna ilość wiązań oddzielających atomy), odpowiednio 20-4 oraz 1024-7.

#### a. odciski cząsteczkowe (fingerprint)

Budowę modelu *in silico* działania proarytmicznego potencjalnych substancji leczniczych rozpoczęto od oceny wpływu dodatkowego opisu struktury chemicznej cząsteczek, w postaci binarnych odcisków cząsteczek (fingerprint), na sprawność testowanych algorytmów klasyfikacyjnych. Wyniki testowania najlepszych modeli w procedurze 10-krotnego wzajemnego sprawdzania przedstawione zostały w tabelach 19-21.

**Tabela 19.** Wyniki testowania (procedura 10-CV) najlepszych algorytmów uczonych na zbiorze 38in.

Architektura	iteracje	AG	SE	SP	PPV	NPV
15_7_5_tanh_38in_HERG_EM_bin_rand_scale	1 000 000	78.91	0.84	0.73	0.78	0.80
3_2_tanh_38in_HERG_EM_bin_orig_scale	1 000 000	78.30	0.79	0.77	0.80	0.77
5_3_2_tanh_38in_HERG_EM_bin_rand_scale	500 000	78.26	0.79	0.77	0.80	0.77
15_7_5_fsr_38in_HERG_EM_bin_rand_scale	50 000	78.07	0.80	0.76	0.79	0.77
5_3_2_sigma_38in_HERG_EM_bin_rand	200 000	78.07	0.81	0.75	0.79	0.78
15_7_5_sigma_38in_HERG_EM_bin_orig	200 000	78.05	0.79	0.77	0.79	0.77
5_3_tanh_38in_HERG_EM_bin_orig_scale	1 500 000	77.68	0.80	0.75	0.78	0.77
3_2_sigma_38in_HERG_EM_bin_orig	300 000	77.66	0.79	0.76	0.79	0.77
3_2_tanh_38in_HERG_EM_bin_orig	100 000	77.65	0.80	0.75	0.78	0.77
15_7_5_tanh_38in_HERG_EM_bin_orig_scale	100 000	77.48	0.81	0.73	0.77	0.78
...						
RandomForest_I10_K0_S1_38in_HERG_EM_bin_orig		75.96	0.74	0.79	0.79	0.73
NaiveBayes_K_38in_HERG_EM_bin_orig		75.46	0.78	0.73	0.76	0.75
NaiveBayes_38in_HERG_EM_bin_orig		74.14	0.85	0.62	0.72	0.78
BayesNet_QK2_P1_SENTROPY_EsimpleEstimator_A05_38in_HERG_EM_bin_orig		71.72	0.64	0.81	0.79	0.66
BayesNet_QHillClimber_P1_SBAYES_ESimpleEstimator_A05_38in_HERG_EM_bin_orig		71.72	0.64	0.81	0.79	0.66

Tabela 20. Wyniki testowania (procedura 10-CV) najlepszych algorytmów uczonych na zbiorze 58in.

Architektura	iteracje	AG	SE	SP	PPV	NPV
7_5_sigma_58in_HERG_EMF4_bin_orig	50 000	78.05	0.88	0.67	0.75	0.83
7_5_3_sigma_58in_HERG_EMF4_bin_rand	300 000	77.07	0.84	0.70	0.76	0.79
7_5_3_tanh_58in_HERG_EMF4_bin_rand_scale	10 000	77.03	0.83	0.70	0.76	0.79
5_3_fsr_58in_HERG_EMF4_bin_rand	1 000 000	76.85	0.79	0.75	0.78	0.76
5_3_2_tanh_58in_HERG_EMF4_bin_rand_scale	750 000	76.64	0.82	0.70	0.76	0.78
7_5_3_sigma_58in_HERG_EMF4_bin_orig	300 000	76.46	0.86	0.66	0.74	0.81
3_2_fsr_58in_HERG_EMF4_bin_orig	500 000	76.24	0.83	0.69	0.75	0.78
3_2_fsr_58in_HERG_EMF4_bin_rand_scale	50 000	76.07	0.80	0.72	0.76	0.76
7_5_tanh_58in_HERG_EMF4_bin_rand_scale	300 000	75.87	0.82	0.70	0.75	0.77
3_2_fsr_58in_HERG_EMF4_bin_rand	300 000	75.85	0.81	0.70	0.75	0.77
...						
NaiveBayes_58in_HERG_EMF4_bin_orig		75.96	0.86	0.65	0.73	0.80
NaiveBayes_K_58in_HERG_EMF4_bin_orig		75.96	0.78	0.74	0.77	0.75
RandomForest_I25_K0_S1_58in_HERG_EMF4_bin_orig		73.89	0.64	0.85	0.83	0.68
RandomForest_I10_K0_S1_58in_HERG_EMF4_bin_orig		73.74	0.68	0.80	0.79	0.69
BayesNet_QK2_P1_SBAYES_ESimpleEstimator_A05_58in_HERG_EMF4_bin_orig		71.72	0.64	0.81	0.79	0.66
BayesNet_QK2_P10_SBAYES_ESimpleEstimator_A05_58in_HERG_EMF4_bin_orig		71.72	0.64	0.81	0.79	0.66

Tabela 21. Wyniki testowania (procedura 10-CV) najlepszych algorytmów uczonych na zbiorze 1062in.

Architektura	iteracje	AG	SE	SP	PPV	NPV
5_3_2_tanh_1062in_HERG_EMF7_bin_orig	300 000	74.83	0.69	0.81	0.81	0.70
3_2_fsr_1062in_HERG_EMF7_bin_rand	10 000	74.77	0.71	0.80	0.80	0.71
5_3_fsr_1062in_HERG_EMF7_bin_orig	10 000	74.57	0.73	0.76	0.78	0.72
7_5_sigma_1062in_HERG_EMF7_bin_orig	10 000	74.57	0.70	0.80	0.80	0.70
ADTree_B25_E3_1062in_HERG_EMF7_bin_orig		74.21	0.70	0.79	0.80	0.69
3_2_sigma_1062in_HERG_EMF7_bin_rand	10 000	73.99	0.67	0.82	0.81	0.69
15_7_5_fsr_1062in_HERG_EMF7_bin_rand_scale	300 000	73.93	0.72	0.77	0.78	0.71
15_7_5_tanh_1062in_HERG_EMF7_bin_rand	750 000	73.90	0.69	0.79	0.79	0.70
3_2_fsr_1062in_HERG_EMF7_bin_orig	10 000	73.75	0.73	0.76	0.77	0.71
15_7_5_sigma_1062in_HERG_EMF7_bin_orig	50 000	73.75	0.71	0.77	0.78	0.70
5_3_fsr_1062in_HERG_EMF7_bin_orig_scale	300 000	73.73	0.73	0.75	0.77	0.71
...						
BayesNet_QK2_P1_SBAYES_ESimpleEstimator_A05_1062in_HERG_EMF7_bin_orig		69.31	0.63	0.76	0.75	0.65
RandomForest_I50_K0_S1_1062in_HERG_EMF7_bin_orig		68.69	0.60	0.79	0.76	0.63
NaiveBayes_K_1062in_HERG_EMF7_bin_orig		68.69	0.64	0.74	0.73	0.65

Najlepsze wyniki uzyskano dla algorytmów trenowanych z wykorzystaniem zbiorów przykładów opisanych 38-elementowym zestawem deskryptorów niezależnie od typu testowanego algorytmu. Rezultaty przedstawione w tabelach 19-21 wskazują, że wyższą poprawność klasyfikacji, niezależnie od wykorzystanego zbioru uczącego, uzyskują modele neuronowe, przy czym różnica pomiędzy najlepszą siecią neuronową a

najlepszymi algorytmami programu WEKA, oceniana testem McNemara, nie była istotna statystycznie.

Złożoność obliczeniowa jest funkcją nie tylko liczebności zbioru, ale również rozmiaru wektora uczącego co oznacza, że czas komputera niezbędny do wygenerowania modelu rośnie wraz z rozszerzaniem opisu prezentowanych przykładów. Ponadto zdolności dyskryminacyjne nie uległy poprawie w konsekwencji zastosowania poszerzonego opisu struktury chemicznej cząsteczek znajdujących się w zbiorze treningowym, o czym świadczą wyniki testowania modeli w procedurze 10-CV (tabele 19-21). Z tego względu do dalszych eksperymentów wybrano modele uczone na zbiorach opisanych za pomocą 38 deskryptorów.

b. walidacja zewnętrzna modeli

W celu dodatkowej weryfikacji zdolności generalizacyjnych najlepszych modeli przeprowadzono testowanie na zewnętrznym zbiorze testowym, którego wyniki przedstawiono w tabeli 22.

**Tabela 22.** Wyniki testowania wybranych modeli na zewnętrznym zbiorze testowym.

Architektura	iteracje	AG% (*)	SE	SP	PPV	NPV
15_7_5_fsr_38in_HERG_EM_bin_rand_scale	50 000	89 (49/55)	0.97	0.81	0.85	0.95
15_7_5_tanh_38in_HERG_EM_bin_rand_scale	1 000 000	87 (48/55)	0.93	0.81	0.84	0.91
RandomForest_I10_K0_S1_38in_HERG_EM_bin_orig		87 (48/55)	0.90	0.85	0.87	0.88
3_2_tanh_38in_HERG_EM_bin_orig_scale	1 000 000	86 (47/55)	0.90	0.81	0.84	0.88
NaiveBayes_K_38in_HERG_EM_bin_orig		86 (47/55)	0.93	0.77	0.82	0.91
5_3_2_tanh_38in_HERG_EM_bin_rand_scale	500 000	82 (45/55)	0.86	0.77	0.81	0.83
15_7_5_tanh_38in_HERG_EM_bin_orig_scale	100 000	75 (41/55)	0.79	0.69	0.74	0.75
5_3_tanh_38in_HERG_EM_bin_orig_scale	1 500 000	71 (39/55)	0.66	0.77	0.76	0.67
5_3_2_sigma_38in_HERG_EM_bin_rand	200 000	53 (29/55)	1.00	0.00	0.53	-
15_7_5_sigma_38in_HERG_EM_bin_orig	200 000	53 (29/55)	1.00	0.00	0.53	-
3_2_sigma_38in_HERG_EM_bin_orig	300 000	53 (29/55)	1.00	0.00	0.53	-
3_2_tanh_38in_HERG_EM_bin_orig	100 000	53 (29/55)	1.00	0.00	0.53	-

(\*) - w nawiasach podano liczbę poprawnie sklasyfikowanych rekordów w stosunku do liczebności zbioru testowego



Cztery najwyżej sklasyfikowane na podstawie wyników testowania w procedurze 10-CV architektury wykazywały satysfakcjonującą zdolność dyskryminacji także podczas testowania z wykorzystaniem zbioru zewnętrznego. Również dwa algorytmy programu WEKA, o najlepszych zdolnościach klasyfikacyjnych, osiągnęły zadowalające wyniki testowania na zbiorze zewnętrznym.

c. komitety ekspertów

W celu dalszej poprawy jakości modelu zaprojektowano 3 komitety ekspertów, w których zastosowano przetwarzanie wyjść sieci uczonych na danych oryginalnych polegające na ich uśrednianiu. Komitety ekspertów tworzone były z wykorzystaniem sieci neuronowych uczonych w oparciu o 38-elementowy zestaw deskryptorów charakteryzujących poszczególne rekordy. Pierwszy komitet ekspertów (A) złożony był z 6 sieci, które uzyskały najlepsze wyniki w procedurze 10-krotnego wzajemnego sprawdzania. Jako kryterium przyjęto wartość błędu generalizacji na poziomie niższym od 22% (Tabela 19). Drugi komitet utworzyły algorytmy, które uzyskały najlepsze wyniki dla zewnętrznego zbioru testowego (błąd generalizacji <15%; Tabela 22). W jego skład weszły modele utworzone w oparciu o sieci neuronowe oraz las losowy i naiwny algorytm Bayesa. Trzeci komitet (C) został zbudowany w oparciu o najlepszy (najwyższa wartość AG% w procedurze 10-CV) z algorytmów testowanych w programie WEKA oraz 5 sieci neuronowych. Procedura wyboru sieci obejmowała:

- wybór 10 sieci o najlepszej zdolności identyfikacji substancji potencjalnie bezpiecznych,
- wybór 10 sieci o najlepszej zdolności identyfikacji substancji potencjalnie niebezpiecznych,
- wybór najlepszych sieci wg kryterium błędu generalizacji (próg odcięcia = 75% poprawnej klasyfikacji).

W składzie tego komitetu ostatecznie znalazły się sieci zebrane w tabeli 23:

**Tabela 23.** Wyniki testowania (procedura 10-CV) algorytmów wybranych do utworzenia komitetu ekspertów C.

Architektura	iteracje	AG	SE	SP	PPV	NPV
5_3_fsr_38in_HERG_EM_bin_rand_scale	50 000	76.44	0.73	0.81	0.81	0.73
5_3_fsr_38in_HERG_EM_bin_rand	10 000	75.87	0.89	0.62	0.72	0.83
7_5_3_fsr_38in_HERG_EM_bin_rand_scale	1 500 000	75.68	0.71	0.81	0.81	0.71
20_lin_fuzzy_38in_HERG_EM_bin_rand	200 000	75.68	0.92	0.58	0.71	0.86
3_2_fsr_38in_HERG_EM_bin_rand_scale	2 000 000	74.66	0.69	0.81	0.80	0.70
RandomForest_I10_K0_S1_38in_HERG_EM_bin_orig		75.96	0.74	0.79	0.79	0.73

Tabela 24 prezentuje wyniki testowania przygotowanych komitetów ekspertów w procedurze 10-CV. Najlepszą zdolność dyskryminacji substancji blokujących kanał hERG i nieposiadających takiego potencjału wykazywał komitet C. Błąd generalizacji komitetu C

jest również mniejszy niż najlepszej pojedynczej sieci 15\_7\_5\_tanh\_38in\_HERG\_EM\_bin\_rand\_scale (Tabela 19). Zastosowanie komitetu ekspertów wpłynęło na zwiększenie liczby substancji poprawnie sklasyfikowanych, poprawie uległa także specyficzność modelu oraz wartości predyktywne dodatnia i ujemna, pogorszyła się natomiast czułość modelu. Różnica poprawności klasyfikacji komitetu C i pozostałych komitetów a także najlepszej pojedynczej sieci neuronowej była statystycznie istotna (test McNemara).

**Tabela 24.** Wyniki testowania komitetów ekspertów - procedura 10-CV.

	Liczba modułów	AG%	SE	SP	PPV	NPV
Komitet A	6	80%	0.82	0.78	0.80	0.79
Komitet B	5	79%	0.82	0.76	0.80	0.79
<b>Komitet C</b>	<b>6</b>	<b>82%</b>	<b>0.82</b>	<b>0.82</b>	<b>0.83</b>	<b>0.81</b>

Wyniki testowania wybranego komitetu ekspertów na zbiorze zewnętrznym, przeprowadzonego w celu dodatkowej weryfikacji jego zdolności generalizacyjnych, przedstawiono w tabeli 25.

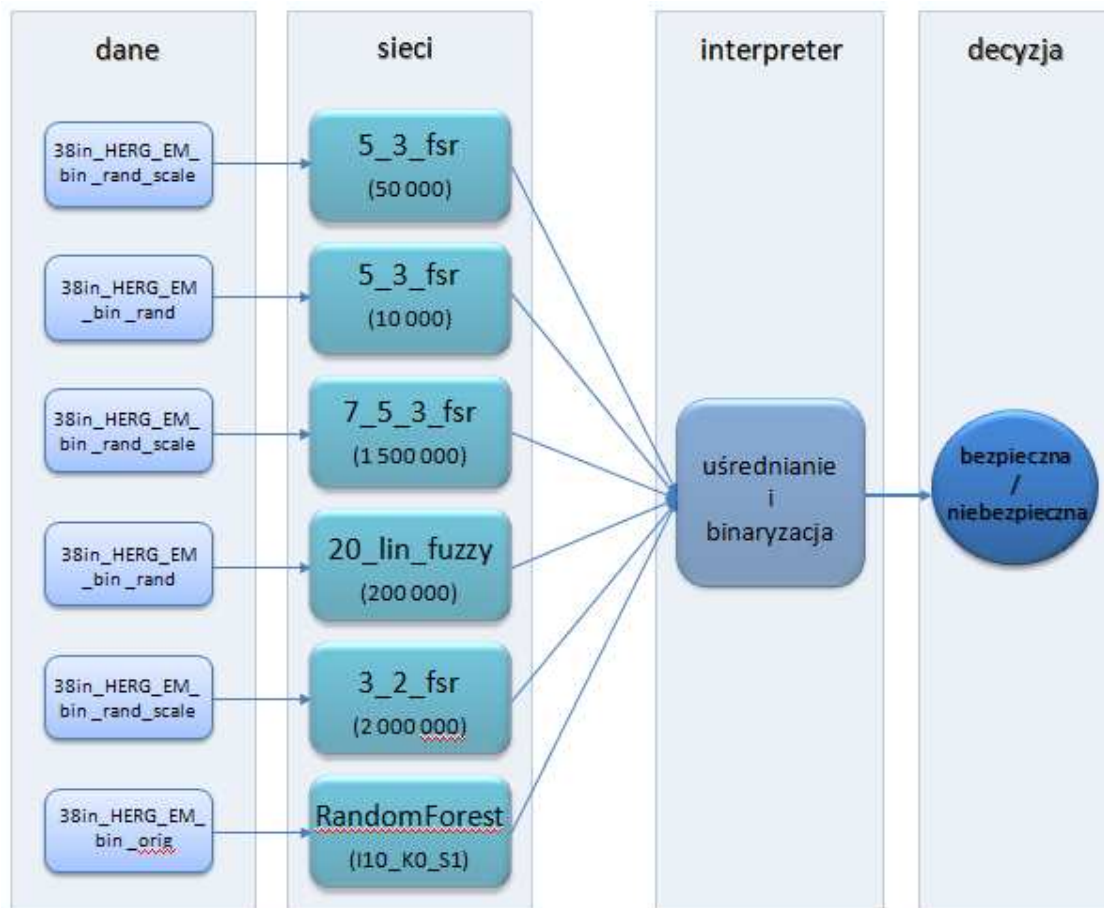
**Tabela 25.** Wyniki testowania komitetu ekspertów C na zewnętrznym zbiorze testowym.

	Liczba modułów	AG%(* )	SE	SP	PPV	NPV
<b>Komitet C</b>	<b>6</b>	<b>87% (48/55)</b>	<b>0.97</b>	<b>0.77</b>	<b>0.82</b>	<b>0.95</b>

(\* ) - w nawiasie podano liczbę poprawnie sklasyfikowanych rekordów w stosunku do liczebności zbioru testowego

Ogólna efektywność klasyfikacji modelu podczas testowania na zewnętrznym zbiorze danych była wyższa niż w procedurze 10-CV. Zmniejszyła się specyficzność i wartość predyktywna dodatnia, równocześnie jednak poprawie uległa czułość modelu oraz wartość predyktywna ujemna generowanych przez niego wyników.

Celem pracy było przygotowanie modelu działania proarytmicznego potencjalnych substancji leczniczych, którego zadaniem będzie klasyfikacja cząsteczek jako bezpieczne (nie blokujące kanału potasowego hERG) lub niebezpieczne, a więc mające potencjał hamowania prądu potasowego  $I_{Kr}$ , a w konsekwencji wydłużania odstępu QT w zapisie EKG i tworzenia warunków sprzyjających zagrażającej życiu arytmii torsade de pointes. Zadawalające wyniki komitetu ekspertów C zarówno w procedurze 10-krotnego wzajemnego sprawdzania jak i podczas weryfikacji za pomocą zewnętrznego zbioru danych zdecydowały, że będzie on stanowił bazę tworzonego modelu. Rycina 30 przedstawia strukturę i działanie ostatecznego modelu działania proarytmicznego.



Rycina 30. Budowa i działanie ostatecznego modelu (komitet ekspertów C).

## IV. PODSUMOWANIE WYNIKÓW I DYSKUSJA

### 1. BAZA DANYCH

Rezultatem pierwszego etapu pracy było stworzenie zbioru eksperymentalnych wartości  $IC_{50}$  dla kanału potasowego hERG wraz z podstawowymi informacjami o czynnikach, które mogą wpływać na wyniki oznaczeń. Przygotowana baza danych zawierała 599 rekordów opisujących 236 substancji, dla których oznaczenia wartości stężenia  $IC_{50}$  wykonane zostały z wykorzystaniem komórek XO, CHO lub HEK.

Przedstawiona w pracy baza danych powstała w oparciu o dane literaturowe pochodzące z artykułów naukowych dostępnych autorowi. Należy założyć, że czasopisma prenumerowane przez Bibliotekę Medyczną Collegium Medicum Uniwersytetu Jagiellońskiego oraz dostępne bezpłatnie za pośrednictwem sieci Internet (*OpenAccess*) nie obejmują wszystkich możliwych publikacji dotyczących badań substancji w kierunku blokowania kanałów potasowych hERG. Niemniej jednak, liczba rekordów oraz substancji uwzględnionych w zbiorze danych jest istotnie wyższa niż w wielu zbiorach informacji stanowiących podstawę modeli (opartych na publicznie dostępnych informacjach) opisywanych w literaturze (Tabela 26; [33,43,62,72,73,258,401]). Większą liczebność w stosunku do przygotowanej bazy mają zbiory danych generowane i używane przez firmy farmaceutyczne [113,283,305], jednak nie są one udostępniane w publikacjach przedstawiających zbudowane na ich podstawie modele. Zakres aktywności biologicznej zgromadzonych związków reprezentowany w zbiorze danych jest szeroki (od  $10^{-4}$  do  $10^4$   $\mu$ M). Można więc przyjąć założenie, że zgromadzona baza danych jest reprezentatywną próbą wszystkich przeprowadzanych i publikowanych wyników doświadczeń dotyczących badań potencjału blokowania hERG.

Kolejną kwestią mogącą limitować przydatność bazy danych jest ocena jakości i wiarygodności opisywanych w pozyskanych publikacjach wyników badań eksperymentalnych. Artykuły naukowe będące źródłem informacji pochodzą z recenzowanych czasopism (rozdział IX. Piśmiennictwo). Dodatkowo weryfikowano zgodność danych opisywanych w tekście z danymi przedstawianymi w sposób graficzny, a wszelkie ewentualne wątpliwości wyjaśniano z autorami. Dane niepewne nie były włączane do zbioru. Taki sposób postępowania wydaje się ograniczać postawiony problem do minimum.

Liczba zgromadzonych rekordów danych, ich jakość oraz uwzględniony zakres przestrzeni chemicznej jest wystarczająca do zastosowania ich jako zbioru uczącego podczas budowy modeli predykcyjnych dla działania proarytmicznego potencjalnych substancji leczniczych. Obecność porównywalnej liczby informacji o substancjach silnie blokujących kanał potasowy jak i o słabym powinowactwie do kanału również jest korzystna z punktu widzenia procesu tworzenia modelu, którego zadaniem jest klasyfikacja. Charakter

informacji zawartych w bazie danych daje również możliwość budowy modeli przewidujących wartości o charakterze ciągłym ( $IC_{50}$ ). Pozwoli to na tworzenie rankingu cząsteczek według potencjału blokowania kanału hERG i wybór najlepszych struktur do kolejnych faz badań.

Opracowana baza danych wraz z metodyką jej budowy została opublikowana w czasopiśmie *Journal of Applied Toxicology* [266]. Jej na bieżąco aktualizowana wersja jest dostępna w portalu *tox-portal.net*, dzięki czemu jest wykorzystywana przez jej użytkowników, m.in. toksykologów oraz specjalistów rozwoju leku zajmujących się oceną ryzyka substancji leczniczych.

## 2. WSPÓŁCZYNNIKI EKSTRAPOLACJI

Podczas analizy zgromadzonego zbioru danych zwrócono uwagę na duże zróżnicowanie wartości  $IC_{50}$  warunkowane zastosowaniem określonego modelu komórkowego oraz innych warunków doświadczalnych. Implikuje to konieczność uwzględnienia różnorodnych czynników podczas budowy modeli, a także skłania do poszukiwań możliwości standaryzacji tego rodzaju danych. Konstrukcja wydajnego, wysokiej jakości modelu matematycznego o pożądanych właściwościach predykcyjnych i zdolności ekstrapolacji w znaczącym stopniu zależy od spójności i jednorodności danych wejściowych [160]. Z tego względu zostały opracowane i zastosowane współczynniki ekstrapolacji pozwalające na ujednoczenie zgromadzonych wyników badań doświadczalnych oraz umożliwiające bezpośrednie porównanie danych o różnej charakterystyce, a także wiarygodną ekstrapolację wyników pozwalającą na wnioskowanie o efektach *in vivo*.

Opracowanie współczynników ekstrapolacji na podstawie ograniczonego zbioru danych, a także nie rozpatrywanie zróżnicowania innych, poza modelem i temperaturą, warunków doświadczeń laboratoryjnych może skutkować brakiem dokładności ich oszacowania. Należy jednak zwrócić uwagę na brak jakichkolwiek unormowań w zakresie standardów prowadzenia eksperymentów, a także prezentacji wyników jednostkom oceniającym bezpieczeństwo potencjalnych leków. Uwzględniając ten fakt, oraz przedstawione wyniki walidacji proponowanych współczynników ekstrapolacji świadczące o ich dobrej wartości predykcyjnej, można stwierdzić, iż ich zastosowanie podczas oceny potencjału proarytmicznego potencjalnych leków jest uzasadnione. Zaproponowane współczynniki mogą być przydatnym narzędziem wspomagającym ocenę ryzyka działania kardiotoksycznego na podstawie wartości  $IC_{50}$  uzyskanych z zastosowaniem różnych systemów *in vitro*. O ich użyteczności świadczą wyniki przedstawione na rycinach 26 i 27, zwłaszcza jeśli uwzględniony zostanie fakt, że w zbiorze danych znajdowały się również cząsteczki problematyczne. Wykorzystanie współczynników w procesie wstępnej oceny toksyczności potencjalnych leków oferuje możliwość większej elastyczności w doborze modelu do doświadczeń elektrofizjologicznych, zgodnie z preferencjami badacza, a także na ich przeprowadzenie w dogodniejszej, pokojowej temperaturze. Ponadto wydaje się, że ocena potencjalnego działania niepożądanego w warunkach klinicznych na podstawie interakcji lek-kanal hERG jest obciążona mniejszym błędem, jeśli znana jest wartość  $IC_{50}$  wyznaczona z wykorzystaniem komórek HEK w temperaturze fizjologicznej. Ocena względnego potencjału proarytmicznego dwóch różnych związków, będzie również bardziej wiarygodna, a ryzyko wyciągnięcia błędnych wniosków mniejsze jeśli porównywane będą wartości  $IC_{50}$  otrzymane w jednakowych warunkach [384].

### 3. MODELE

#### 3.1. Analiza wrażliwościowa

Analiza wrażliwościowa - procedura matematyczna pozwalająca wyznaczyć wpływ danych wejściowych sieci na podawaną wartość wyjściową, jest jednym z narzędzi pozwalających zredukować rozmiar wektora wejściowego. Zmniejszenie stopnia komplikacji sieci poprzez wybór zmiennych kluczowych i ograniczenie liczby deskryptorów powinno poprawić zdolności generalizacyjne [253].

Pierwotnie modele działania proarytmicznego zbudowane zostały w oparciu o zbiór danych opisanych za pomocą 108 zmiennych niezależnych. Wyniki analizy wrażliwościowej pozwoliły na ograniczenie pierwotnego wektora wejściowego do 38, 21 oraz 12 składowych (Rycina 29, Tabela 18).

Rezultaty testowania sieci neuronowych, trenowanych z wykorzystaniem przygotowanych zgodnie z wynikami tej analizy zbiorów uczących, w procedurze 10-krotnego wzajemnego sprawdzania wykazały najwyższą użyteczność 38-elementowego opisu przykładów uczących. Wybrane deskryptory (Tabela 18) można określić jako zmienne kluczowe. W zbiorze deskryptorów wskazanych w procesie analizy wrażliwości znalazły się zarówno elementy opisujące warunki doświadczeń laboratoryjnych, jak i właściwości fizyko-chemiczne oraz przestrzenne cząsteczek wskazywane w literaturze jako istotne dla wiązania molekuł substancji z kanałem potasowym hERG [m.in. 11,12,43,238]. Obecność parametrów opisujących warunki doświadczenia potwierdza kluczową rolę ich doboru dla otrzymywanych wyników pomiaru stężenia  $IC_{50}$ , a ich uwzględnienie wśród parametrów modelu pozwala na przeprowadzanie wirtualnych eksperymentów w konfiguracji dowolnie zdefiniowanej przez użytkownika.

Deskryptory właściwości cząsteczek wybrane w procesie analizy wrażliwościowej pozostają w zgodności z wynikami badań strukturalnych kanału hERG (m.in. ukierunkowana mutageneza, skaning alaninowy, modelowanie struktur białkowych, dokowanie) oraz dostępnymi modelami farmakoforowymi. Badania te wskazują na kluczową rolę fragmentów niosących ładunki dodatnie oraz właściwości hydrofobowe [m.in. 252,261,336]. Ze 108 deskryptorów właściwości fizyko-chemicznych wygenerowanych przy pomocy pakietu Marvin (program Cxcalc), 30 wskazanych zostało jako zmienne kluczowe dla poprawnej dyskryminacji substancji. Wśród nich znalazły się m.in. deskryptory opisujące właściwości hydrofobowe ( $\log P$ , ASA\_H), związane z wielkością i kształtem cząsteczki (Maximal i minimal projection area/radius, Smallest/largest ring size, Rotatable bond count, ) czy ładunkiem (ayy).

Wyniki testowania modeli uzyskanych w oparciu o zredukowany wektor wejść (liczba składowych wektora wejściowego - 38) wskazują także na zasadność wstępnego przetwarzania danych z wykorzystaniem procesu analizy wrażliwościowej. Zdolności

generalizacyjne zredukowanych sieci poprawiły się w porównaniu z sieciami uczonymi przy zastosowaniu danych z pełnowymiarowym wektorem wejść, jednocześnie możliwe było znaczne uproszczenie architektury. Ograniczono więc jeden z elementów wpływających na złożoność obliczeniową problemu. Pomimo istotnego zredukowania liczby wejść zdolności klasyfikacyjne modelu nie uległy pogorszeniu, sieci nadal potrafiły efektywnie różnicować przedstawiane przykłady. Nadmierne ograniczenie opisu przykładów może spowodować, że usunięte zostaną informacje umożliwiające skonstruowanie wydajnego modelu. Weryfikacja modeli uzyskanych w oparciu o zredukowany wektor wejść z zastosowaniem zewnętrznego zbioru danych pozwoliła stwierdzić, że zastosowany opis (38 parametrów) jest wystarczający, potwierdzono zatem zasadność zmniejszenia wymiarowości wektora wejść [253].

## 3.2. Modelowanie

### 3.2.1. Odciski cząsteczek

Binarne odciski cząsteczek (*ang. fingerprints*) są wykorzystywane jako element narzędzi oceny podobieństwa struktur chemicznych, stąd podjęto próby oceny przydatności dodatkowego opisu substancji dla problemu klasteryzacji. W przypadku drzew decyzyjnych i sieci Bayesa wykazano brak istotnej zależności pomiędzy zdolnością generalizacji algorytmów uczonych z zastosowaniem zbiorów treningowych opisanych wektorem wejść wyłonionych w trakcie analizy wrażliwościowej, a tych opisanych dodatkowo za pomocą odcisków cząsteczkowych. Dla sieci neuronowych natomiast sprawność klasyfikacji malała wraz z rozbudową opisu rekordów uczących. Obserwowana zależność może być wynikiem niedostatecznej pojemności sieci neuronowych w stosunku do rozmiaru wektora wejść. Brak narzuconej a priori złożoności drzew decyzyjnych i algorytmów bayesowskich ogranicza wpływ ilości deskryptorów na zdolność generalizacji. Jednak brak istotnej poprawy zdolności generalizacyjnych po rozbudowie opisu zbioru treningowego może sugerować, że charakteryzacja rekordów za pomocą 30 deskryptorów struktury jest wystarczająca. Sugeruje to także, że uzyskanie dalszej poprawy zdolności generalizacyjnych modelu jest możliwe poprzez rozbudowanie zbioru przykładów (zwiększenie liczebności zbioru treningowego), a nie w wyniku rozszerzania opisu struktury chemicznej cząsteczek. Dodatkowo zwiększenie wymiarowości wektora wejściowego wpływa na wzrost złożoności obliczeniowej procesu trenowania i testowania algorytmów budowanych na jego podstawie. Nieuzasadnione jest więc wykorzystanie zbioru 58in oraz 1062in do budowy ostatecznego modelu.

### 3.2.2. Algorytmy

W wyniku uczenia i testowania prawie 700 różnych architektur sieci neuronowych oraz ponad 100 konfiguracji algorytmów dostępnych w programie WEKA trenowanych na zbiorze 38in\_HERG\_EM\_bin otrzymano zestaw parametrów charakteryzujących poszczególne modele pod względem ich zdolności generalizacji (Tabela 19). Poszukiwanie



optymalnej architektury/algorytmu polegało na porównaniu efektywności generalizacji różnych architektur i algorytmów (liczba rekordów zbioru testowego sklasyfikowanych poprawnie - AG) w procedurze 10-krotnego wzajemnego sprawdzania.

Optymalna sieć neuronowa znaleziona w toku eksperymentów numerycznych, posiadała trójwarstwową architekturę 15\_7\_5\_tanh. Poprawność klasyfikacji tej sieci wynosiła 79%. Najlepszym z algorytmów programu WEKA (poprawność klasyfikacji AG=76%) okazał się RandomForest\_I10\_K0\_S1, las losowy zbudowany z 10 drzew z losowo wybieranymi 6 deskryptorami w każdym węźle. Liczba podziałów drzewa nie była limitowana, co pozwoliło na elastyczną rozbudowę drzewa stosownie do zbioru przykładów uczących. Wartość ziarna (*ang. seed*) dla generatora liczb pseudolosowych algorytmu wynosiła 1.

Testowanie na zewnętrznym zbiorze danych, nie wykorzystanych w procesie trenowania sieci, wykonano w celu dodatkowej weryfikacji zdolności generalizacyjnych najlepszych modeli. Sześć spośród dziesięciu sieci neuronowych, które uzyskały najlepsze wyniki w procedurze 10-CV wykazywało zadowalające zdolności klasyfikacyjne również w przypadku testowania na zbiorze zewnętrznym. Cztery sieci nie potrafiły uogólnić zgromadzonej wiedzy na nowe przykłady. Sieci te nauczyły się odwzorowywać zbiór treningowy, jednak wygenerowana architektura nie pozwala na ekstrapolację zgromadzonej wiedzy poza ten zbiór, czego dowodem są jednakowe odpowiedzi w przypadku wszystkich rekordów zewnętrznego zbioru testowego. Może to być efektem zbyt małej liczby iteracji i niedostatecznej liczby prezentacji zbioru uczącego, która nie pozwala na wytworzenie ogólnych reguł.

### 3.2.3. Komitety ekspertów

W celu dalszej poprawy jakości modelu neuronowego zaprojektowano i przetestowano trzy komitety ekspertów. Najwyższym odsetkiem poprawnie sklasyfikowanych substancji blokujących kanał hERG i nieposiadających takiego potencjału charakteryzował się komitet C. Wytypowany komitet ekspertów został utworzony w oparciu o dwie sieci najlepiej przewidujące substancje niebezpieczne, trzy o najwyższym odsetku poprawnie przewidzianych substancji bezpiecznych oraz najlepszy algorytm uzyskany przy wykorzystaniu programu WEKA (las losowy). Wynik testowania komitetu C (Tabela 24) potwierdza przyjęte założenie o korzystnym wpływie łączenia różnych algorytmów i sieci neuronowych o odmiennej strukturze na dokładność klasyfikacji w wyniku wzajemnego uzupełniania się poszczególnych modułów komitetu. Błąd generalizacji komitetu C jest również mniejszy niż najlepszej pojedynczej sieci 15\_7\_5\_tanh\_38in\_HERG\_EM\_bin\_rand\_scale (Tabela 19).

Ogólna efektywność klasyfikacji tego modelu podczas testowania na zewnętrznym zbiorze danych była wyższa niż w procedurze 10-CV. Zmniejszyła się specyficzność i wartość predykcyjna dodatnia, równocześnie jednak poprawie uległa czułość modelu

oraz wartość predykcyjna ujemna generowanych przez niego wyników. Z punktu widzenia zadanej funkcji projektowanego systemu korzystne jest, aby model generował jak najmniej odpowiedzi fałszywie negatywnych (możliwie najwyższa wartość NPV), tak aby prawdopodobieństwo, że substancja niebezpieczna zostanie wytypowana do dalszych etapów badań było jak najmniejsze.

Opisany komitet ekspertów został wybrany jako podstawa projektowanego modelu predykcyjnego, którego zadaniem jest skринing oraz klasyfikacja potencjalnych substancji leczniczych na bezpieczne lub blokujące kanał hERG we wczesnych etapach badań nad lekiem.

Modelowanie 2D-QSAR uznawane jest za uzasadnioną i racjonalną metodę tworzenia predykcyjnych narzędzi skринingowych wspomagających ocenę toksyczności leku związaną z blokowaniem kanału hERG [90,113,172,283,345].

Do konstrukcji systemów klasyfikacyjnych stosuje się różnorodne techniki m.in. metody regresyjne PLS [106,113,172], HQSAR [43,172], sztuczne sieci neuronowe (ANN) [250,283,399], algorytmy ewolucyjne [16], metody wektorów nośnych (SVM) [149,208,359,399], metodę k-najbliższych sąsiadów (kNN) [383,399], klasyfikatory Bayesa [333,345,346], metody podziałów rekursywnych (drzewa klasyfikacyjne i regresyjne) [76,115,399], samoorganizujące się mapy (SOM) [90,283].

W ostatnich latach zaproponowanych zostało kilka modeli komputerowych prognozujących potencjał blokowania kanału potasowego hERG. Większość z nich bazuje na strukturach znanych ligandów (*ang. ligandbased approach*) – modele farmakoforowe. Bazy danych eksperymentalnych wykorzystanych do budowy modeli zwykle zawierają substancje podobne do siebie, należące do jednej grupy terapeutycznej, ponadto liczba cząsteczek w zbiorze uczącym i testowym przeważnie jest niewielka. Ogranicza to możliwość ich zastosowania dla zbiorów zawierających substancje różnorodne strukturalnie. Cavalli [33] zaproponował 4-elementowy model farmakoforowy (zasadowy atom azotu otoczony 3 strukturami aromatycznymi) w oparciu o strukturę 31 leków powodujących przedłużenie odstępu QT. Model, testowany na 6 kolejnych lekach, wykazywał zadowalającą poprawność predykcji. Przedstawiony przez Ekinsa [91] model wygenerowano z wykorzystaniem 15 struktur zawierających 4 fragmenty hydrofobowe i jeden z pozytywnym ładunkiem elektrycznym. Walidację przeprowadzono na cząsteczkach w większości stosowanych jako leki antypsychotyczne. Podobnie model 3D-QSAR Pearlsteina [258] z wykorzystaniem analizy porównawczej (CoMSiA – comparative molecular similarity analysis) zawierał 22 analogi sertindolu oraz 10 inhibitorów hERG o innej strukturze. Z kolei Roche [283] opisuje zastosowanie różnorodnych technik do modelowania właściwości blokowania kanału hERG dla 472 substancji. Najlepszy model wykorzystujący sieci neuronowe poprawnie klasyfikował 76% rekordów 95-elementowego zbioru testowego. Korzystając z danych zgromadzonych przez Roche oraz

listy Fenichela, Keseru z współpracownikami [172] opracowali klasyczne oraz holograficzne modele QSAR (HQSAR), które walidowano za pomocą trzech zbiorów danych. Modele te charakteryzują się dobrą zdolnością predykcyjną, zarówno dla IC<sub>50</sub> jak i problemów klasyfikacyjnych. W Tabeli 26 zamieszczono przegląd opisanych powyżej i innych opublikowanych modeli klasyfikacyjnych oraz porównanie z modelem zaproponowanym w niniejszej pracy.

**Tabela 26.** Charakterystyka opublikowanych modeli klasyfikacyjnych blokowania kanału hERG (opracowane na podstawie [345]).

Autor	Metoda	Próg dyskryminacji	zbiór uczący			zbiór testowy	
			liczebność	AG%	Metoda walidacji wew.	liczebność	AG%
<b>Aronov [11]</b>	filtr podobieństwa topologicznego 2D + 3D farmakofofor	40 µM	414	82	50-CV (80/20)	-	-
<b>Bains [16]</b>	algorytmy ewolucyjne	1 µM	100	88	10-CV (90/10)	24	85-90
<b>Chekmarev [38]</b>	SVM	<1 µM >10 µM	83	77	10-CV (90/10)	-	-
				69	5-CV (80/20)		
	UFS-SVM		83	78	10-CV (90/10)	-	-
				74	5-CV (80/20)		
	kNN (k=7)		83	68	10-CV (90/10)	-	-
				66	5-CV (80/20)		
kNN (k=3)	83	69	10-CV (90/10)	-	-		
		67	5-CV (80/20)				
<b>Dubus [76]**</b>	DT	1 µM	160	96-98	5-CV (80/20)	43	74-81
		<1 µM (1-10 µM)	203	86-90	5-CV (80/20)	-	-
		>10 µM	100	96-97	5-CV (80/20)	55	93-96
<b>Ekins [90]</b>	SOM i SNM	<1 µM >10 µM	95	83-86		35	81-95
<b>Fioravanzo [106]</b>	5 modeli QSAR (PLS)	10 µM	29	86A 100		38	82

				NA				
	5 modeli QSAR (PLS) + PASS + QikProp	10 $\mu$ M	-	-		38	87	
<b>Gepp [115]</b>	DT	TdP+ TdP-	246	92-93		75	76-80	
<b>Ivanciuc [146]</b>	AIRS	TdP+ TdP-	349	86	10-CV (90/10)	-	-	
<b>Jia [149]**</b>	SVM	30 $\mu$ M	997	85	10-CV (90/10)	66	94	
			997	86	5-CV (80/20)			
			997	74	2-CV (50/50)			
<b>Keseru [172]</b>	QSAR (PLS)	1 $\mu$ M	55	83A 87NA		13	85	
		'ECG changer' w World Drug Index (WDI)	55	83A 87NA		82	81	
		'APPROVED' w WDI	55	83A 87NA		4043	18	
<b>Li [208]</b>	SVM	40 $\mu$ M	495	74	LOO	66	72	
			495	74		1948 (Pubchem)	73	
<b>O'Brein [250]**</b>	sieci Bayesa	wypieranie dofetilidu +/-	46967	77		11996	82	
	ANN		46967	77			85	
	konsensus Bayes+ANN		46967	77		10437	89	
<b>Roche [283]</b>	ANN	<1 $\mu$ M >10 $\mu$ M	244	93	10-CV (80/20)	95	77	
<b>Sun [333]**</b>	naiwny klasyfikator Bayesa	30 $\mu$ M	1979	ROC 0.87		66	88	
<b>Thai [346]</b>	QSAR (Bayes)	1 $\mu$ M	223	78-81	LOO	64	91-94	
			223	84-85				
		10 $\mu$ M	223	78-82	LOO	64	73-75	
			223	80-83				
<1 $\mu$ M >10 $\mu$ M	150	84-85	LOO	40	88-93			
	150	87						
<b>Thai [347]</b>	ANN		285	64-82	5-CV (80/20)	-	-	
			230	89-95		55	84-85	
			285	74	5-CV (50/50)	-	-	
			143	94-98		142	72	
<b>Thai [345]</b>	QSAR (Bayes)	1 $\mu$ M	194	72-89		49	78-92	
			194	71-86	LOO			

			194	77-88	5-CV (80/20)	49	67-85
			243	-		58	93
		10 $\mu$ M	194	78-88		49	61-90
			194	76-86	LOO		
			194	75-88	5-CV (80/20)	49	69-85
			243	-		58	53
		<1 $\mu$ M	194	82-96		49	52-78
	ANN	1-10 $\mu$ M	194	83-94	5-CV (80/20)	49	44-74
		$\geq$ 10 $\mu$ M	243	-		58	95-100
<b>Tobita [359]</b>	SVM		73	90	10-CV (90/10)	414 Drugdex	82
		40 $\mu$ M	73	95	10-CV (90/10)	13	85
	DT C4.5			67	5-CV (80/20)	-	-
	ANN			78	5-CV (80/20)	-	-
	RBF	3.16 $\mu$ M	134	77	5-CV (80/20)	-	-
	LR			82	5-CV (80/20)	-	-
	SVM			85	5-CV (80/20)	-	-
<b>Wang [378]</b>	DT C4.5			77	5-CV (80/20)	47	72
	ANN			93	5-CV (80/20)	47	77
	RBF	<1 $\mu$ M >10 $\mu$ M	91	81	5-CV (80/20)	47	81
	LR			95	5-CV (80/20)	47	77
	SVM			95	5-CV (80/20)	47	77
	SVM			81	LOO	78	91
<b>Yap [399]</b>	ANN	TdP+	271	<81	LOO	78	78
	kNN	TdP-		<81	LOO	78	89
	DT C4.5			<81	LOO	78	65
<b>MODEL I</b>	ANN		202 (494)	79	10-CV* (~90/~10)	55	87
<b>MODEL II</b>	RF	1 $\mu$ M	202 (494)	76	10-CV* (~90/~10)	55	87
<b>MODEL III</b>	konsensus ANN+RF		202 (494)	82	10-CV* (~90/~10)	55	87

\*\* - dane firm farmaceutycznych – niedostępne publicznie

**10-CV\*** - procedura walidacji polegająca na wyłączeniu 10% rekordów zbioru treningowego z bardziej rygorystycznymi kryteriami niż stosowane standardowo, losowanie przynależności do zbioru dotyczy struktur a nie pojedynczych rekordów

**A/NA** – substancje aktywne/nieaktywne

**AIRS** - sztuczny system immunologiczny (ang. Artificial Immune Recognition System)

**ANN** - sztuczne sieci neuronowe (ang. Artificial Neural Network)

**DT** - drzewo decyzyjne (ang. Decision Tree)

**LOO** – procedura walidacji polegająca na wyłączeniu pojedynczych rekordów (ang. Leave-One-Out)

**LR** - regresja logistyczna (ang. Logistic Regression)

**n-CV** - procedura walidacji polegająca na wyłączeniu n% rekordów zbioru treningowego (ang. n-fold Cross Validation)

**RBF** – sieci neuronowe o radialnych funkcjach bazowych (ang. Radial Basis Functions)

**RF** – las losowy (ang. Random Forest)

**ROC** – krzywa charakterystyki odbiorcy (ang. Receiver Operating Characteristic)

**SNM** – mapy nieliniowe (ang. Sommon Nonlinear Mapping)

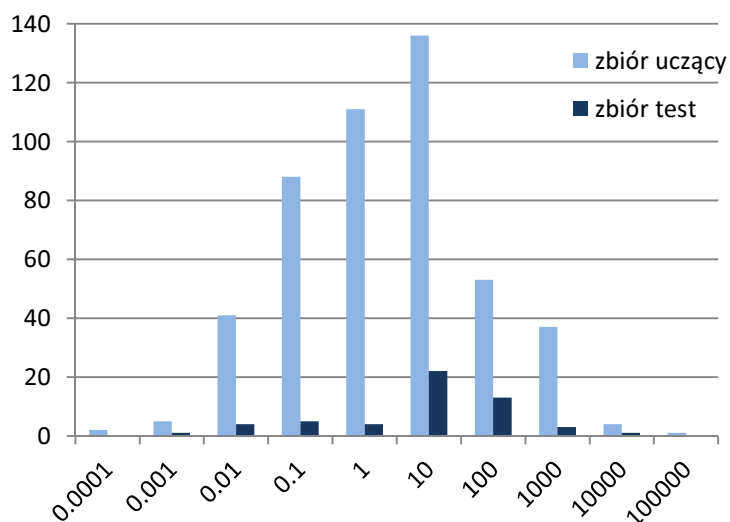
**SOM** – samoorganizująca się mapa (ang. Self-Organizing Map)

**SVM** – metody wektorów nośnych (ang. Support Vector Machine)

**UFS-SVM** – SVM z modyfikacją pozwalającą na redukcję wymiarowości problemu (unsupervised forward selection)

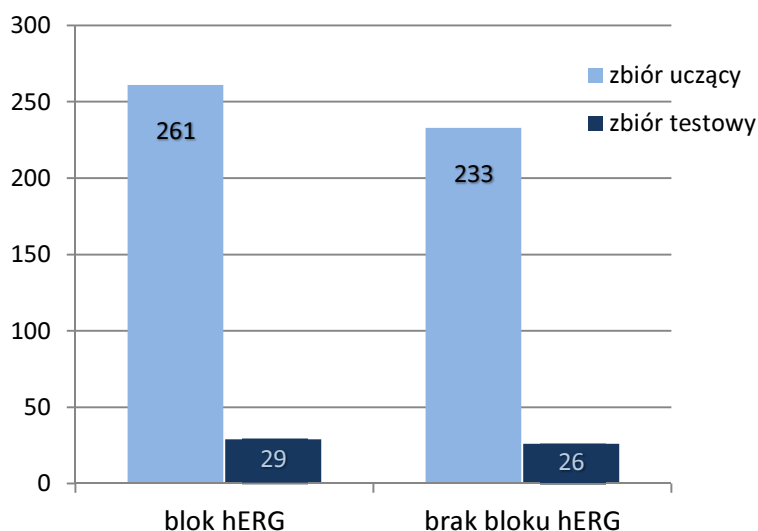
Jakość każdego modelu uzależniona jest w znacznym stopniu od danych doświadczalnych, przede wszystkim ich jakości i różnorodności oraz technik zastosowanych do jego utworzenia. Liczba rekordów i substancji uwzględniona w zbiorze stanowiącym podstawę opracowanego modelu jest wyższa niż w przypadku większości modeli tworzonych w oparciu o publicznie dostępne dane dotyczące blokowania kanału potasowego hERG (Tabela 26). Podkreślić należy niezwykle restrykcyjne kryteria włączania danych do zbioru uczącego i treningowego. W niektórych opisywanych w literaturze modelach spotyka się przypadki wykorzystania wcześniej opublikowanych baz danych bez krytycznej oceny zawartych w nich informacji. Ponadto klasyfikacja substancji wg obranego progu dyskryminacji oparta jest na wartościach  $IC_{50}$  wyznaczonych w różnorodnych warunkach laboratoryjnych i na różnych komórkach, co może prowadzić do błędów w ocenie bezpieczeństwa testowanych cząsteczek (rozdział III.2 – „Współczynniki ekstrapolacji, str. 86). Niekiedy także do generowania opisywanych w publikacjach modeli stosuje się dane niereprezentatywne dla sytuacji występujących w warunkach doświadczalnych, czego przykładem jest wykluczanie ze zbioru treningowego substancji, dla których wartości stężenia  $IC_{50}$  mieszczą się w granicach 1-10  $\mu$ M.

Wykorzystany w niniejszej pracy zbiór danych obejmuje szeroki zakres wartości stężeń  $IC_{50}$  - 10 rzędów wielkości (Rycina 31). Zastosowano także opracowaną i opisaną wcześniej metodę standaryzacji wyników pochodzących z doświadczeń wykonanych w różnych warunkach (rozdział III.2 – „Współczynniki ekstrapolacji, str. 86).



**Rycina 31.** Zakres wartości IC<sub>50</sub> wykorzystanych do wygenerowania modelu - zbiór uczący i testowy.

Ponadto zarówno zbiór treningowy, jak i użyty do zewnętrznej walidacji zaproponowanego modelu zawierają porównywalną liczbę substancji aktywnych i nie (Rycina 32). Takie zbalansowanie rekordów zbioru uczącego pozwala na nauczenie modelu rozpoznawania nie tylko cech warunkujących wiązanie z kanałem, ale także tych, które ograniczają tego rodzaju interakcje. Natomiast obecność zbliżonej liczby substancji obu kategorii w zbiorze testowym, pozwala wiarygodnie ocenić zdolności dyskryminacyjne wygenerowanego modelu.

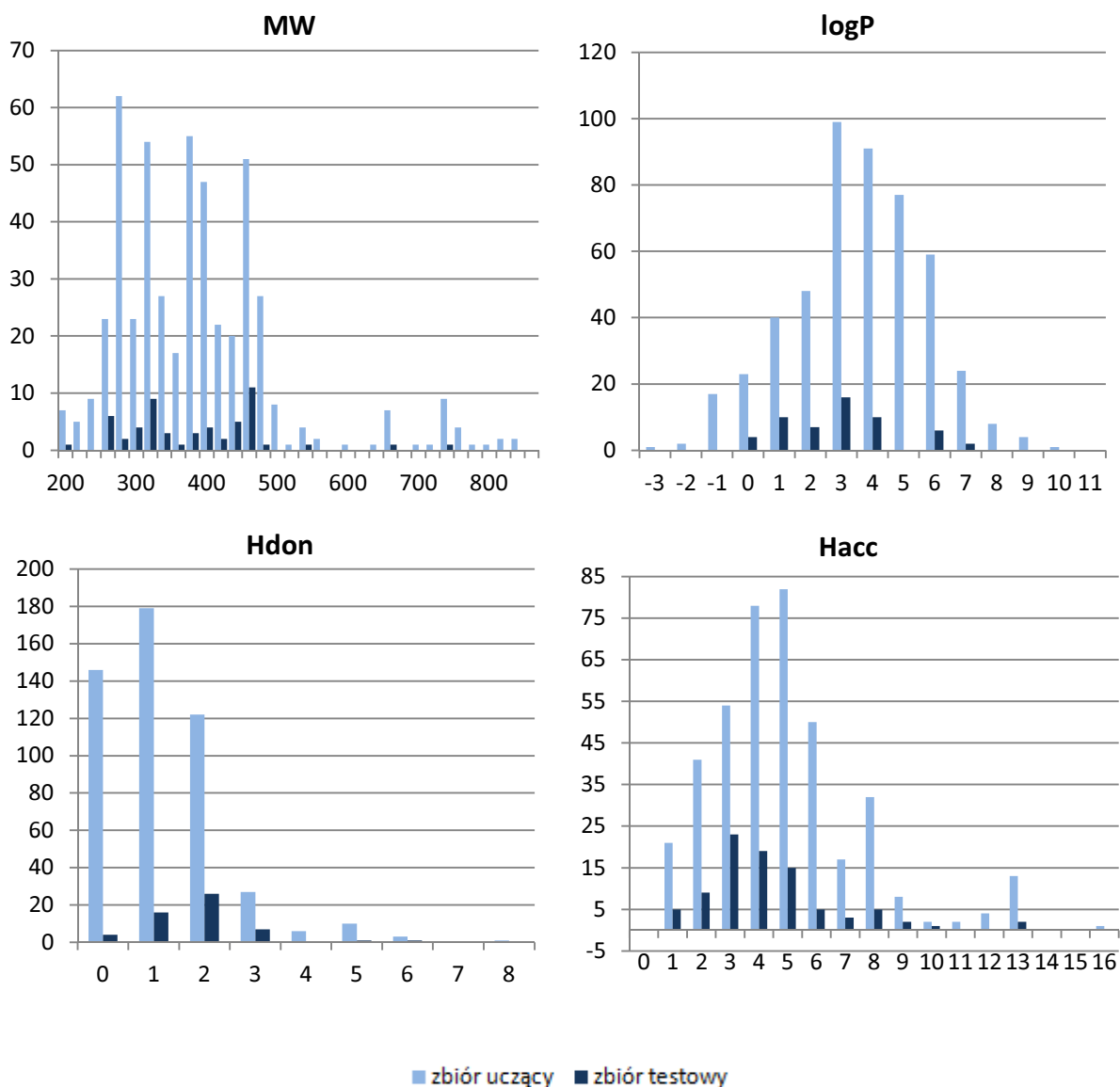


**Rycina 32.** Porównanie liczebności klas zbioru uczącego i testowego.

Zbiór uczący zapewnia heterogeniczność przykładów reprezentując jednocześnie przestrzeń chemiczną obejmującą swoim zakresem właściwości fizyko-chemiczne charakteryzujące substancje lecznicze. Ryciny 33a-d przedstawiają obszar przestrzeni chemicznej uwzględniony w danych uczących i testowych wg czterech parametrów: masy molowej (MW), współczynnika podziału olej/woda (logP), liczby donorów (Hdon) i

akceptorów wiązań wodorowych (Hacc). Zgodnie z regułą pięciu Lipińskiego są to parametry, których wartość pozwala oceniać i wybierać cząsteczki do dalszych badań we wczesnych fazach procesu rozwoju leku. Podobny rozkład tych właściwości obserwować można w zewnętrznym zbiorze testowym, co świadczy o tym, że jest on wystarczająco różnorodny, aby zweryfikować wygenerowane modele.

Poprawność dyskryminacji, oceniana na niezależnym zbiorze testowym, zaproponowanego w prezentowanej pracy modelu wyniosła 87%, co jest jednym z najlepszych uzyskanych i opublikowanych dotychczas rezultatów (Tabela 26). Potwierdza to między innymi odpowiednią jakość i różnorodność zgromadzonej w trakcie pracy bazy danych. Niemniej jednak proces gromadzenia danych dotyczących oceny potencjału blokowania kanałów potasowych hERG jest kontynuowany i być może pozwoli na aktualizację, dalszą rozbudowę i poprawę zaproponowanego modelu.



Rycina 38a-d. Właściwości fizyko-chemiczne substancji zbioru uczącego i testowego.

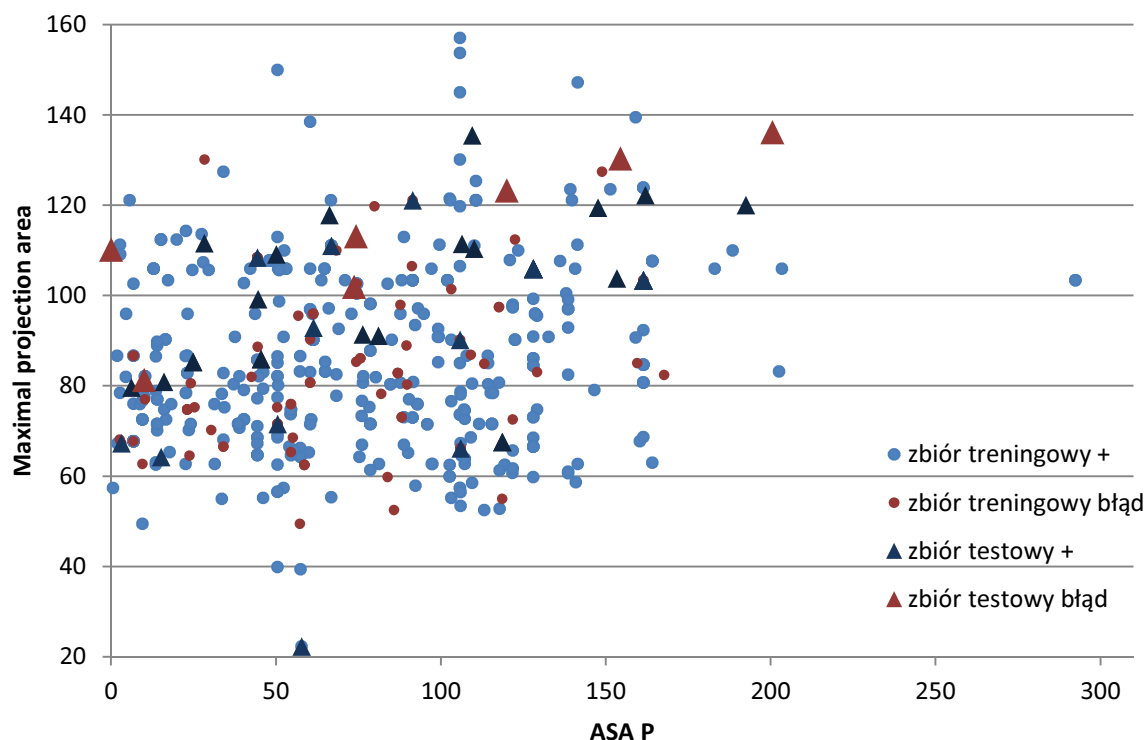


Ogólna poprawność klasyfikacji utworzonego modelu oceniana w zmodyfikowanym (wyłączenie 10% rekordów wg struktury – rozdział III.3 „Modele”, str. 114) procesie 10-krotnego wzajemnego sprawdzania wyniosła 82%. Walidacja modelu za pomocą zewnętrznego, niezależnego zbioru danych potwierdziła dobre zdolności dyskryminacyjne (87% poprawnie sklasyfikowanych rekordów).

Większość opisywanych w literaturze naukowej modeli wykazuje akceptowalne zdolności klasyfikacyjne dla silnych inhibitorów hERG ( $IC_{50} < 1 \mu M$ ) oraz ze słabym powinowactwem do kanału ( $IC_{50} > 10 \mu M$ ), nie sprawdza się jednak w przypadku substancji o umiarkowanym potencjale blokowania tego kanału ( $IC_{50} 1-10 \mu M$ ) [38,76,90,345]. Analiza błędnie zaklasyfikowanych przykładów ze zbioru uczącego wykazała, że jedna trzecia z nich to substancje o wartości  $IC_{50}$  z zakresu 1-10  $\mu M$ . Można spodziewać się, że prawdopodobieństwo błędnej klasyfikacji jest największe w przypadku cząsteczek, dla których oceniana właściwość bliska jest wartości progowej -  $IC_{50} \sim 1 \mu M$ . Powszechnie przyjmowane kryterium poprawności działania modelu zakłada, że model jest poprawny, jeśli podawane przez niego wyniki nie różnią się od rzeczywistych więcej niż dwa rzędy wielkości [333,343]. Kryterium to uwzględnia błąd oznaczeń eksperymentalnych. Spośród 89 błędnie sklasyfikowanych rekordów 47 dotyczyło substancji o zmierzonej wartości  $IC_{50}$  z przedziału 0.1-1 (różnica jednego rzędu wielkości w stosunku do przyjętego proggu odcięcia), a 72 z przedziału 0.01-10 (różnica dwóch rzędów wielkości), w tym kontekście zdolności klasyfikacyjne modelu ocenić należy jako bardzo dobre. Podobnie, w zbiorze testowym 4 z 7 nieprawidłowych decyzji dotyczyły substancji o  $IC_{50}$  w przedziale 1-10  $\mu M$ . Największa liczba błędów w tym zakresie stężeń może także sugerować konieczność zastosowania dodatkowych deskryptorów, innych niż uwzględnione właściwości fizykochemiczne w celu zwiększenia poprawności dyskryminacyjnych modelu dla słabych blokerów hERG.

Inną przyczyną błędów może być odmienna od reprezentowanej w zbiorze uczącym charakterystyka niektórych substancji. Masa molowa jednej z pozostałych błędnie sklasyfikowanych substancji wynosiła 543, a więc znajdowała się w obszarze słabo reprezentowanym w zbiorze uczącym.

Rycina 34 przedstawia przestrzeń chemiczną zbioru uczącego i testowego według dwóch najważniejszych (wskazanych w analizie wrażliwościowej) deskryptorów – ASA\_P (powierzchnia atomów polarnych dostępna dla rozpuszczalnika) oraz maksymalnej powierzchni projekcji konformeru. Niepoprawnie zaklasyfikowane substancje znajdują się na obrzeżach reprezentowanej przez przykłady uczące przestrzeni cech, co może utrudniać systemowi właściwą ocenę. W rozdziale III.3 – „Modele” (str. 114) zaznaczono, że na zewnętrzny zbiór walidacyjny składały się zarówno przykłady opisujące doświadczenia wykonane dla znanych systemowi substancji jednak przeprowadzone w odmiennej konfiguracji (inne komórki i/lub warunki oznaczenia) oraz dotyczące nowych, nieuwzględnionych w zbiorze uczącym struktur.



**Rycina 34.** Przestrzeń chemiczna zbioru uczącego i testowego wg dwóch najważniejszych deskryptorów – poprawność klasyfikacji.

Analiza błędnych decyzji modelu pod tym kątem pokazuje, że wszystkie dotyczyły nieznanymi systemowi struktur. Świadczy to o bardzo dobrych zdolnościach generalizacyjnych modelu w zakresie warunków eksperymentalnych i słabszej zdolności ekstrapolacji struktury chemicznej cząsteczek (Tabela 27). Należy zwrócić uwagę, że sześć z siedmiu błędnych decyzji to wyniki fałszywie pozytywne, a więc dotyczące zaklasyfikowania substancji nieblokującej kanału hERG jako niebezpiecznej. Z punktu widzenia funkcji jaką spełniać ma projektowany system jest to sytuacja korzystniejsza niż decyzje fałszywie negatywne, ponieważ ryzyko pojawienia się substancji blokującej kanał potasowy w dalszych etapach badań jest niewielkie (wartość predykcyjna ujemna 0.93). Minimalizacja tego ryzyka odbywa się tu kosztem odrzucenia podczas skringingu potencjalnie wartościowych substancji. Jednak obserwowane współcześnie zmiany metodyki poszukiwania substancji aktywnych (m.in. metody chemii kombinatorycznej), a także waga potencjalnych efektów stosowania leku z potencjałem zaburzania prądu potasowego w populacji czy starty finansowe związane z odrzuceniem takiej substancji na późnych etapach badań sprawiają, że poprawa wartości NPV kosztem niesłusznego wyeliminowania substancji bezpiecznych jest uzasadniona i akceptowalna.

**Tabela 27.** Wyniki testowania modelu na zewnętrznym zbiorze danych z podziałem wg występowania bądź braku struktury chemicznej w zbiorze uczącym.

zbiór danych	AG (*)	SE	SP	PPV	NPV
NOWE	77 (23/30)	0.91	0.68	0.63	0.93
ZNANE	100 (25/25)	1.00	1.00	1.00	1.00

NOWE - nieznaną strukturą i warunki doświadczenia

ZNANE - znana struktura i nieznanymi warunki doświadczenia

(\*) - w nawiasach podano liczbę poprawnie sklasyfikowanych rekordów w stosunku do liczebności poszczególnych zbiorów

Potwierdzony potencjał blokowania kanału potasowego hERG różnorodnych strukturalnie substancji, należących do różnych grup terapeutycznych, sugeruje, że problem ten może ujawniać się na drodze kilku mechanizmów. Wieloczynnikowe podłoże zjawiska inhibicji kanału hERG sprawia, że opracowanie globalnego modelu jest problematyczne i stanowi duże wyzwanie. Jednak wykorzystanie modelowania QSAR opartego na danych upraszcza nieco to zadanie poprzez możliwość opracowania problemu bez konieczności odwoływania się do szczegółów mechanizmów leżących u jego podstaw. Wykorzystywanie modeli *in silico* uzasadnia także fakt, iż do tej pory struktura krystalograficzna kanału hERG nie została zdefiniowana. Zastosowanie technik modelowania komputerowego w procesie badań i rozwoju leku oferuje możliwość zbudowania wydajnych kosztowo narzędzi skringingu i identyfikacji potencjalnie toksycznych substancji. W niniejszej pracy przedstawiono model efektywnie klasyfikujący substancje blokujące szybką komponentę opóźnionego prądu potasowego w komórkach serca. Podstawę modelu stanowią deskryptory struktury chemicznej cząsteczek, które mogą być łatwo i szybko obliczone (w niektórych przypadkach także wyznaczone doświadczalnie) zarówno dla rzeczywistych jak i wirtualnych bibliotek. Pozyskanie danych koniecznych i wystarczających do oceny potencjału kardiotoksycznego badanych substancji wymaga więcej czasu oraz odpowiednich ilości substancji, co znacznie spowalnia proces badań i rozwoju. Modele realizowane w warunkach *in silico*, bazujące na danych jakościowych wykazują ponadto większą tolerancję błędu laboratoryjnego obecnego w danych uczących, a jakościowy opis powinowactwa badanej substancji do kanału hERG, podobnie jak w przypadku wielu innych toksykologicznych punktów końcowych, jest często wystarczający do podjęcia stosownych decyzji we wczesnych etapach badań. Uzyskane wyniki modelowania ilustrują możliwości zastosowania zaproponowanego systemu do identyfikacji substancji niebezpiecznych i skringingu ukierunkowanego na eliminację potencjalnie problematycznych cząsteczek we wczesnych fazach badań przed wdrożeniem skringingu *in vitro*. Należy jednak podkreślić, że zaproponowany system predykcji, tak jak każdy model, jest rozwiązaniem suboptymalnym oraz zgodnie z zasadami modelowania powinien podlegać aktualizacji i testowaniu na nowych danych. Ponadto cechą immanentną modeli wywodzących się z klasycznych metod QSAR jest ograniczona możliwość ekstrapolowania poza przestrzeń zbioru treningowego. Odnalezione przez model zależności pomiędzy deskryptorami

opisującymi właściwość fizyko-chemiczne oraz parametrami doświadczeń, a klasą przypisywaną badanej substancji (bezpieczna/blokująca kanał hERG) są złożone i nie umożliwiają przeprowadzenia wnioskowania na temat pożądanych lub niepożądanych cech struktury cząsteczek. Wada ta jest jednak kompensowana przez fakt, że do predykcji model wymaga znajomości jedynie struktur 2D i łatwych do obliczenia na ich podstawie deskryptorów, a co za tym idzie możliwe jest przetestowanie dużej liczby substancji w krótkim czasie.

Podsumowując, zdolności predykcyjne zbudowanego systemu wyróżniają się wśród modeli do tej pory zaproponowanych w literaturze, mimo zastosowania rygorystycznych metod testowania. Cechami wyróżniającymi system predykcji potencjału kardi toksycznego potencjalnych substancji leczniczych opisanego w pracy są ponadto różnorodność i liczba danych zastosowanych do jego budowy, wykorzystanie danych wystandaryzowanych oraz dostępność. Zarówno baza danych, opracowane współczynniki ekstrapolacji dla temperatury i modelu doświadczalnego (HEK, CHO, XO) jak i ostateczny model zostały udostępnione publicznie za pośrednictwem dedykowanego portalu [tox-portal.net](http://tox-portal.net), co pozwoli na ich ocenę przez naukowców z innych ośrodków akademickich. Opracowany model predykcyjny został także włączony do tworzonego w ramach programu Lider systemu wczesnej oceny kardi toksyczności substancji chemicznych, w oparciu o narzędzia modelowania matematycznego CompToxOne [140].

## V. WNIOSKI

- Liczba rekordów oraz różnorodność danych, a także zakres uwzględnionej przestrzeni chemicznej w przygotowanej bazie danych pozwalają na zbudowanie wydajnych modeli predykcyjnych działania proarytmicznego projektowanych leków związanego z inhibicją kanału potasowego hERG.
- Charakter informacji zgromadzonej w bazie danych umożliwia zarówno tworzenie modeli klasyfikacyjnych jak i rozbudowę zaproponowanego narzędzia w kierunku predykcji wartości ciągłych (IC<sub>50</sub>).
- Źródłem informacji niezbędnych w procesie tworzenia modeli inhibicji kanału hERG może być analiza danych literaturowych. W celu zapewnienia odpowiedniej jakości i efektywności modelu niezbędne jest jednak zastosowanie wstępnego przetwarzania (m.in. standaryzacji danych).
- Wyznaczone współczynniki ekstrapolacji pozwalają na standaryzację i w konsekwencji wiarygodną ocenę ryzyka opartą na danych pochodzących z różnych źródeł lub otrzymanych w różnych warunkach. Dobra wartość predykcyjna opracowanych współczynników ekstrapolacji potwierdza zasadność stosowania takiej metody przetwarzania danych.
- Zastosowana w pracy modyfikacja standardowej metody testowania sprawności modelu pozwala wiarygodnie ocenić zdolności generalizacyjne modelu.
- System modularny (tzw. komitet ekspertów) zbudowany z 5 sieci neuronowych oraz lasu losowego należy uznać za optymalny model, dobrze realizujący cel do jakiego został zaprojektowany.
- Opracowany model może być wykorzystany we wczesnych etapach badań nad lekiem jako efektywne narzędzie wspomagające skryning właściwości proarytmicznych projektowanych leków.

## VI. STRESZCZENIE

Określenie możliwości działania proarytmicznego stanowi jeden z kluczowych elementów profilu bezpieczeństwa wymaganego przez agencje rejestrujące i nadzorujące bezpieczeństwo stosowania leków. Niniejsza praca dotyczyła wykorzystania metod modelowania komputerowego do opracowania narzędzia wspomagającego i uzupełniającego konwencjonalne metody oceny potencjału arytmogennego substancji chemicznych.

Zadaniem zaprojektowanego narzędzia jest skrining cząsteczek na wczesnych etapach rozwoju leku poprzez klasyfikację na potencjalnie arytmogenne i potencjalnie bezpieczne. Podstawą przeprowadzanej klasyfikacji jest ocena możliwości blokowania przez badaną substancję kanału potasowego hERG odpowiedzialnego za generowanie prądu jonowego będącego elementem końcowej fazy repolaryzacji komórek mięśnia sercowego.

Pierwszy etap pracy obejmował analizę bibliograficzną dostępnych źródeł literaturowych oraz zgromadzenie danych opisujących w sposób ilościowy zjawisko blokowania kanału. Wynikiem tego etapu jest stworzony zbiór eksperymentalnych wartości  $IC_{50}$  (stężenie substancji wywołujące 50% blok prądu jonowego) dla kanału potasowego hERG wraz z podstawowymi informacjami o czynnikach, które mogą wpływać na uzyskiwane wyniki oznaczeń laboratoryjnych (warunki oznaczenia elektrofizjologicznego). Dane pochodzą z recenzowanych artykułów opublikowanych w czasopiśmie naukowych i uzyskane zostały w oparciu o standardową, uznaną za złoty standard technikę patch clamp, co gwarantuje jednorodność zbioru informacji w tym zakresie. Zgromadzona baza danych zawiera 599 rekordów opisujących 236 substancji, dla których oznaczenia wartości stężenia  $IC_{50}$  wykonane zostały z wykorzystaniem komórek XO, CHO lub HEK, transfekowanych ludzkim DNA kodującym kanał ERG. Liczba i jakość zgromadzonych rekordów danych oraz uwzględniony zakres przestrzeni chemicznej jest wystarczająca do zastosowania jako zbioru uczącego podczas budowy modeli predykcyjnych dla działania proarytmicznych potencjalnych substancji leczniczych. Baza danych jest na bieżąco aktualizowana i udostępniana w portalu tox-portal.net, dzięki czemu jest wykorzystywana przez m.in. toksykologów oraz specjalistów rozwoju leku zajmujących się oceną ryzyka substancji leczniczych.

Proces budowy bazy danych i przeprowadzonej oceny jakości zgromadzonych informacji obejmował szczegółową analizę zależności uzyskiwanych eksperymentalnie wyników pomiarów  $IC_{50}$  i warunków doświadczeń. Kolejnym elementem pracy było opracowanie metody standaryzacji wartości  $IC_{50}$  uzyskanych w warunkach laboratoryjnych z zastosowaniem różnych modeli komórkowych (XO, CHO, HEK) i warunków temperaturowych (temperatura pokojowa i fizjologiczna). Wyznaczone współczynniki ekstrapolacji pozwalają na ujednoczenie zgromadzonych wyników badań

doświadczalnych oraz umożliwiają bezpośrednie porównanie danych o różnej charakterystyce. W celu oszacowania użyteczności opracowanych współczynników przeprowadzono walidację wewnętrzną oceniając rozbieżność pomiędzy wartością szacowaną a rzeczywistą za pomocą pierwiastka błędu średniokwadratowego (RMSE) obliczonego dla poszczególnych przeliczeń ( $XO/HEK_{fizj}$ ;  $CHO_{r.t.}/HEK_{fizj}$ ;  $CHO_{fizj}/HEK_{fizj}$ ;  $HEK_{r.t.}/HEK_{fizj}$ ). Dodatkowo porównano wartości  $IC_{50}$  obliczone przy zastosowaniu najlepszych współczynników ekstrapolacji z wartościami natywnymi. Zdecydowana większość wartości  $IC_{50}$  obliczonych za pomocą współczynników ekstrapolacji nie różniła się od wartości zmierzonych doświadczalnie na komórkach HEK w temperaturze fizjologicznej więcej niż o jeden rząd wielkości. Współczynniki ekstrapolacji poddano również walidacji za pomocą zewnętrznego, niezależnego zbioru danych. Uzyskane w obu procedurach wyniki potwierdzają możliwość wykorzystania współczynników w procesie wstępnej oceny toksyczności potencjalnych leków w celu wiarygodnej ekstrapolacji wyników pozwalającej na wnioskowanie o efektach *in vivo*. Wykorzystanie opracowanych współczynników w procesie wstępnej oceny toksyczności potencjalnych leków oferuje również możliwość doboru modelu do doświadczeń elektrofizjologicznych, zgodnie z preferencjami badacza, a także na ich przeprowadzenie w dogodniejszej, pokojowej temperaturze.

Przygotowana baza danych wraz ze współczynnikami ekstrapolacji posłużyła do budowy modelu predykcyjnego. Proces konstruowania modelu obejmował dobór składowych wektora wejściowego oraz matematycznego algorytmu klasyfikacyjnego. Dla każdej substancji wygenerowano ponad 100 deskryptorów charakteryzujących właściwości fizyko-chemicznych oraz binarne odciski cząsteczek (o długości 20 i 1024 bitów), które dołączono do parametrów określających warunki eksperymentów laboratoryjnych. W celu zmniejszenia stopnia komplikacji sieci i oceny wpływu ograniczenia liczby deskryptorów na zdolność generalizacji modelu wykonano analizę wrażliwościową. W procesie tym wyłoniono zbiór 38 kluczowych dla modelu zmiennych. Wśród wybranych deskryptorów znajdują się zarówno elementy opisujące warunki doświadczeń laboratoryjnych, jak i właściwości fizyko-chemiczne oraz przestrzenne cząsteczek wskazywane w literaturze jako istotne dla wiązania cząsteczek substancji z kanałem potasowym hERG. Obecność parametrów opisujących warunki doświadczalne potwierdza kluczową rolę ich doboru dla otrzymywanych wyników pomiaru stężenia  $IC_{50}$ , a ich uwzględnienie wśród parametrów modelu pozwala na przeprowadzanie wirtualnych eksperymentów w konfiguracji dowolnie zdefiniowanej przez użytkownika. Zdolności generalizacyjne modeli uzyskanych w oparciu o zredukowany wektor wejść poprawiły się w porównaniu z modelami uczonymi przy zastosowaniu danych z pełnowymiarowym wektorem wejść, jednocześnie możliwe było znaczne uproszczenie architektury. Uzyskane wyniki weryfikacji pozwalają stwierdzić, że zastosowany opis (38 parametrów) jest wystarczający. Potwierdzono zatem zasadność zmniejszenia wymiarowości wektora wejść

ograniczając jednocześnie jeden z elementów wpływających na złożoność obliczeniową problemu.

Do opracowania modeli predykcyjnych wykorzystano sztuczne sieci neuronowe, algorytmy bayesowskie oraz metody podziałów rekursywnych (drzewa decyzyjne i ich zbiory). W procesie trenowania i testowania wszystkich stworzonych klasyfikatorów wykorzystano zbiory rekordów opisanych za pomocą 38 parametrów wyłonionych w procesie analizy wrażliwościowej. Około 800 modeli klasyfikacyjnych ocenianych było w zmodyfikowanej procedurze 10-krotnego wzajemnego sprawdzania (10-CV) oraz za pomocą zewnętrznego, niezależnego zbioru danych. Zastosowana modyfikacja metody 10-CV polega na wyłączeniu ze zbiorów testowych rekordów pogrupowanych według substancji. Procedura ta, w przeciwieństwie do standardowo stosowanego wyłączenia losowo wybranych 10% rekordów, gwarantuje, że w zbiorze testowym i uczącym nie znajdą się te same substancje. W podobnych badaniach opisywanych w literaturze, taka metoda nie jest domyślnie stosowana. Zaproponowany system opracowywania danych, stawiający modelowi największe wymagania stanowi element dorobku naukowego pracy.

Optymalny model znaleziony w toku eksperymentów numerycznych stanowiła sieć neuronowa zawierająca trzy warstwy ukryte, złożone kolejno z 15, 7 i 5 neuronów z tangensem hiperbolicznym jako funkcją aktywacji. Odsetek poprawnych klasyfikacji tej sieci wynosił 79% (procedura 10-CV). Sieć prawidłowo zaklasyfikowała 45 z 55 rekordów zewnętrznego zbioru walidacyjnego. Najlepszym z algorytmów programu WEKA (poprawność klasyfikacji w procedurze 10-CV - 76%, na zbiorze zewnętrznym - 87%) okazał się `RandomForest_I10_K0_S1`, las losowy zbudowany z 10 drzew z losowo wybieranymi 6 deskryptorami w każdym węźle.

W celu dalszej poprawy jakości modelu neuronowego zaprojektowano i przetestowano komitety ekspertów. Komitet ekspertów wybrany jako ostateczny model, klasyfikujący substancje na blokujące kanał hERG i potencjalnie bezpieczne, zbudowany został w oparciu o dwie sieci najlepiej przewidujące substancje niebezpieczne, trzy o najwyższym odsetku poprawnie przewidzianych substancji bezpiecznych oraz najlepszy algorytm uzyskany przy wykorzystaniu programu WEKA (las losowy). Ogólna efektywność klasyfikacji modelu podczas testowania na zewnętrznym zbiorze danych wynosi 87% (82% w procedurze 10-CV). Wynik testowania komitetu potwierdza przyjęte założenie o korzystnym wpływie łączenia różnych algorytmów i sieci neuronowych o odmiennej strukturze na dokładność klasyfikacji w wyniku wzajemnego uzupełniania się poszczególnych modułów komitetu.

W pracy wykazano, że techniki *in silico* pozwalają na automatyczną identyfikację wielowymiarowych zależności na podstawie danych doświadczalnych, umożliwiając modelowanie procesów wieloczynnikowych, jakim jest blokowanie kanału hERG, bez konieczności posiadania dogłębnej wiedzy na temat mechanizmu badanego zjawiska.



Model zaproponowany w niniejszej pracy charakteryzuje się 87% ogólną poprawnością przewidywania i wartością predyktywną ujemną (NPV – *ang. Negative Predictive Value*) na poziomie 93%. Z punktu widzenia funkcji systemu korzystna jest wysoka wartość NPV dla decyzji, co pozwala na wczesną eliminację struktur potencjalnie niebezpiecznych. Uzyskany model został wykorzystany w praktyce, jako element programu komputerowego CompToxOne wyspecjalizowanego w ocenie potencjału blokowania kanału hERG ([www.tox-comp.net](http://www.tox-comp.net)).

Zdolności predykcyjne zbudowanego systemu wyróżniają się wśród modeli do tej pory zaproponowanych w literaturze. Wyniki takie udało się uzyskać mimo zastosowania rygorystycznych metod testowania, które nie są powszechnie stosowane. Dodatkowymi atutami systemu predykcji potencjału kardiotoksycznego potencjalnych substancji leczniczych opisanego w pracy są ponadto różnorodność i liczba danych zastosowanych do jego budowy, wykorzystanie danych wystandaryzowanych oraz dostępność.

## VII. OZNACZENIA I SYMBOLE

- 10-CV** – metoda 10-krotnego wzajemnego sprawdzania
- 3R** – zasada Redukować, Doskonalić i Wymieniać (*ang. Refine, Reduce, Replace*)
- ADME/Tox** – wchłanianie, dystrybucja, metabolizm, wydalania i toksyczność (*ang. Absorption, Distribution, Metabolism, Excretion and Toxicity*)
- AG** lub **AG%** - trafność całkowita - odsetek poprawnie sklasyfikowanych rekordów względem całości zbioru testowego
- ANN** - sieci neuronowe (*ang. Artificial Neural Networks*)
- AP** – potencjał czynnościowy (*ang. action potential*)
- AUROC** – pole pod krzywą charakterystyki odbiorcy - ROC (*ang. the area under a receiver operating characteristic/curve*)
- BiR** – Badania i Rozwój (*ang. R&D – Reaserch and Development*)
- BP** – wsteczna propagacja (*ang. Back Propagation*)
- CDER** - Centrum do spraw Oceny i Badań nad Lekami (*ang. Center for Drug Evaluation and Research*)
- CHO** - komórki jajnika chomika chińskiego (*ang. Chinese Hamster Ovary*)
- EMA** – Europejska Agencja Leków (*ang. European Medicines Agency*)
- EPA** – Agencja Ochrony środowiska (*ang. Environmental Protection Agency*)
- FDA** - Amerykańska Agencja ds. Żywności i Leków (*ang. Food and Drug Administration*)
- HEK** – embrionalne komórki nerki ludzkiej (*ang. human embryonic kidney*)
- hERG** – gen kodujący kanał potasowy odpowiedzialny za szybką komponentę odkomórkowego prądu potasowego (*ang. human ether-a-go-go-related gene*)
- IC<sub>50</sub>** – stężenie powodujące połowiczny blok kanału
- ICH** – Międzynarodowa Konferencja ds. Harmonizacji (*ang. International Conference of Harmonization*)
- I<sub>Kr</sub>** – szybki odkomórkowy prąd potasowy
- IND** – badany produkt leczniczy (*ang. Investigational New Drug*)
- JRC** - Wspólnotowe Centrum Badawcze Komisji Europejskiej (*ang. Joint Research Centre*)
- k-NN** – algorytm k-najbliższych sąsiadów (*ang. k-Nearest Neighbours*)
- LQTS** – zespół wydłużonego odcinka QT (*ang. Long QT syndrome*)
- MLP** – perceptron wielowarstwowy (*ang. MultiLayer Perceptron*)
- MSE** i **RMSE** – błąd średniokwadratowy I pierwiastek błędu średniokwadratowego
- NDA** – nowy lek po III fazie badań klinicznych (*ang. New Drug Application*)
- NIH** - Narodowy Instytut Zdrowia (*ang. National Health Institute*)

**NPV** - wartość predykcyjna ujemna (*ang. Negative Predictive Value*)

**PLS** - metoda cząstkowych najmniejszych kwadratów (*ang. Partial Least Squares*)

**PPV** - wartość predykcyjna dodatnia (*ang. Positive Predictive Value*)

**QSAR** – ilościowa zależność struktura chemiczna-aktywność biologiczna (*ang. Quantitative Structure-Activity Relationship*)

**REACH** - rozporządzenie Parlamentu Europejskiego i Rady (WE) Nr 1907/2006 dotyczące bezpiecznego stosowania chemikaliów, poprzez ich rejestrację i ocenę, oraz w niektórych przypadkach udzielanie zezwoleń i ograniczenia handlu i stosowania niektórych chemikaliów (*ang. Registration, Evaluation, Authorisation and Restriction of Chemicals*)

**SE** - czułość (*ang. sensitivity*)

**SP** - specyficzność (*ang. specificity*)

**SVM** - metody wektorów nośnych (*ang. Support Vector Machines*)

**TdP** – tachyarytmia komorowa torsade de pointes

**XO** - oocyty żaby rodzaju *Xenopus* (*ang. Xenopus laevis oocytes*)

## VIII. PUBLIKACJE

Prezentowane w tej pracy zagadnienia i wyniki były częściowo opisane w poniższych publikacjach

1. Polak S, Wiśniowska B, Brandys J. 2009. Collation, assessment and analysis of literature *in vitro* data on hERG receptor blocking potency for subsequent modeling of drugs cardiotoxic properties. *J Appl Tox* 29(3): 183-206.
2. Wiśniowska B, Polak S. 2009. hERG *in vitro* interchange factors - development and verification. *Tox Mech Meth* 19(4): 278-284.
3. Wiśniowska B, Mendyk A, Polak M, Szlek J, Polak S. 2010. RandomForest based assessment of the hERG channel inhibition potential for the early drug cardiotoxicity testing, *Bio-Algorithms and Med-Systems* 6(12): 131-136.
4. Polak S, Wisniowska B, Ahamadi M, Mendyk A. 2011. Prediction of the hERG potassium channel inhibition potential with use of Artificial Neural Networks, *Appl Soft Comp* 11 (2): 2611-2617.
5. Polak S, Wiśniowska B, Ahamadi M, Mendyk A. 2010. Prediction of the hERG potassium channel inhibition potential with use of the artificial neural networks. *Soft Computing in Industrial Applications. Algorithms, Integration and Success Stories. In Advances in Intelligent and Soft Computing* vol. 75 Eds. Gao X-Z, Gaspar-Cunha A., Koeppen M., Schaefer G., Wang J. Springer, Berlin 2010.
6. Polak S, Fijorek K, Glinka A, Wiśniowska B, Mendyk A. 2012. Virtual population generator for human cardiomyocytes parameters. In silico drug cardiotoxicity assessment. *Tox Mech Meth*, 22(1):31-40.
7. Polak S, Wiśniowska B, Fijorek K, Glinka A, Polak M, Mendyk A. 2011. The open-access dataset for in silico cardiotoxicity prediction system. *Bioinformatics*, 6(6):244-245.

oraz były również przedstawiane na konferencjach

1. Wiśniowska B, Polak S, Brandys J. Extrapolating factors for different *in vitro* hERG systems. Central European Congress of Life Science EUROBIOTECH 2008, Kraków 17-19.10.2008; *Acta Biochimica Polonica* 2008, vol. 55 suppl. 4,
2. Polak S, Wiśniowska B, Brandys J. Modeling and simulation in the drug discovery, research and development. Toxicity prediction. 5<sup>th</sup> German-Polish symposium, "New Challenges for Pharmaceutical Sciences", Poznań, 15-16.05.2009.
3. Wiśniowska B, Polak S., Brandys, J. Bayesian classification methods for proarrhythmic liability prediction. 5<sup>th</sup> German-Polish symposium, "New Challenges for Pharmaceutical Sciences", Poznań, 15-16.05.2009.
4. Polak S, Wiśniowska B, Brandys J. Decision trees algorithms for the *in silico* prediction of hERG potassium channels inhibition (in polish). Polish Society of Toxicology Workshop "New trends in toxicology", Łódź 1-2.06.2009
5. Wiśniowska B, Polak S. Cardiotoxicity prediction algorithms. 1st International Workshop Computational Pharmacy. Krakow 19.06.2009 [1st International Workshop Computational Pharmacy](#), Kraków 19.06.2009
6. Polak S, Wiśniowska B, Ahamadi M, Mendyk A. Prediction of the hERG potassium channel inhibition potential with use of the artificial neural networks. [14th Online World Conference on Soft Computing in Industrial Applications \(WSC14\)](#), 17-29.11.2009
7. Wiśniowska B, Mendyk A, Polak M, Szlęk J, Polak S. Wykorzystanie w badaniach nad lekiem algorytmu RandomForest do wczesnej ocena potencjału blokowania kanału potasowego kodowanego przez gen hERG. Modelowanie Cybernetyczne Systemów Biologicznych, 21-22.05.2010 Kraków.

8. Mendyk A, Wiśniowska B, Polak M, Glinka A, Polak S. System wczesnej oceny potencjału blokowania prądu potasowego I<sub>Kr</sub> oparty na sztucznych sieciach neuronowych. XXI Naukowy Zjazd PTFarm, Gdańsk 2010.
9. Wiśniowska B, Mendyk A, Polak M, Glinka A, Polak S. System wczesnej oceny potencjału blokowania prądu potasowego I<sub>Kr</sub> oparty o algorytm SVM. XXI Naukowy Zjazd PTFarm, Gdańsk 2010.

## IX. PIŚMIENICTWO

1. Abbott GW i in. 1999. MiRP1 forms IKr potassium channels with HERG and is associated with cardiac arrhythmia. *Cell* 97(2): 175-187.
2. Agatonovic-Kusturin S, Beresford R. 2000. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *J Pharm Biomed Anal* 22: 717-727.
3. Albrecht C. 2004. Proarrhythmia with non-antiarrhythmics: A review. *Cardiology* 102(3): 122-139.
4. Alexandrou A i in. 2006. Mechanism of hERG K<sup>+</sup> channel blockade by the fluoroquinolone antibiotic moxifloxacin. *Br J Pharmacol* 147(8): 905-916.
5. Al-Khatib S i in. 2003. What Clinicians Should Know about the QT Interval. *J Am Med Assoc* 289(16): 2120-2127.
6. Antzelevitch C. 2005. Role of transmural dispersion of repolarization in the genesis of drug-induced torsades de pointes. *Heart Rhythm* 2: S9-S15.
7. Antzelevitch C. 2007. Ionic, molecular, and cellular bases of QT-interval prolongation and torsade de pointes. *Europace* 9: 4-15.
8. Arcangeli A i in. 1998. Long-term exposure to retinoic acid induces the expression of IRK1 channels in HERG channel-endowed neuroblastoma cells. *Biochem Biophys Res Commun* 244: 706-711.
9. Arias C i in. 2003. Effects of propafenone and its main metabolite, 5-hydroxypropafenone, on HERG channels. *Cardiovasc Res* 57(3): 660-669.
10. Arizona CERT. <http://www.qtdrugs.org/index.cfm> [30.05.2009]
11. Aronov AM, Goldman BB. 2004. A model for identifying HERG K<sup>+</sup> channel blockers. *Bioorg Med Chem* 12(9): 2307-2315
12. Aronov AM. 2005. Predictive *in silico* modeling for hERG channel blockers. *Drug Discov Today* 10(2): 149-155.
13. Arvidson KB i in. 2010. Regulatory use of computational toxicology tools and databases at the United States Food and Drug Administration's Office of Food Additive Safety. *Expet Opin Drug Metabol Toxicol* 6(7): 793-796.
14. Assessment of the QT Prolongation Potential of Non-Antiarrhythmic Drugs. Health Canada, Therapeutic Products Directorate (March 15 2001).
15. Bachmann A i in. 2002. Inhibition of cardiac potassium currents by pentobarbital. *Naunyn Schmiedebergs Arch. Pharmacol* 365(1): 29-37.
16. Bains W, Basman A, White C. 2004. HERG binding specificity and binding site structure: evidence from a fragment-based evolutionary computing SAR study. *Progress in Biophys Mol Biol* 86: 205-233.
17. Baker B i in. 1997. Electrocardiographic effects of fluoxetine and doxepin in patients with major depressive disorders. *J Clin Psychopharmacol* 17: 15-21.
18. Barnett AA. 1996. Safety concerns over antipsychotic drug, sertindole. *Lancet* 348: 256.
19. Bass AS i in. 2008. International Life Sciences Institute (Health and Environmental Sciences Institute HESI) initiative on moving towards better predictors of drug-induced torsades de pointes. *Br J Pharmacol* 154: 1491-1501.

20. Baxter DF i in. 2002. A Novel Membrane Potential-Sensitive Fluorescent Dye Improves Cell-Based Assays for Ion Channels. *J Biomol Screen* 7(1): 79-85.
21. Ben-Bassat M. 1982. Pattern Recognition and Reduction of Dimensionality. *Handbook of Statistics-II*, Krishnaiah PR, Kanal LN. (wyd.), s773-791, North Holland.
22. Bilski, J. 2003. Uczenie nadzorowane wielowarstwowych sieci neuronowych. Ewolucja czy rewolucja. Nowoczesne techniki informatyczne. W R. Cierniak (wyd.), s59–101. Częstochowa: Katedra Inżynierii Komputerowej Politechniki Częstochowskiej.
23. Bischoff U, Schmidt C, Netzer R, Pongs O. 2000. Effects of fluoroquinolones on HERG currents. *Eur. J. Pharmacol.* 406(3): 341-343.
24. Bonett, DG, Price RM. 2002. Statistical inference for a linear function of medians: Confidence intervals, hypothesis testing and sample size requirements. *Psychol meth* 7(3): 370-383.
25. Breiman L. 2001. Random forests. *Mach Learn* 45(1): 5–32.
26. Brown AM. 2004. Drugs, hERG and sudden death. *Cell Calcium* 35: 543–547.
27. Brown, AM. 2005. hERG block, QT liability and sudden death. The hERG cardiac potassium channel: structure, function and long QT syndrome. Derek J i in. (wyd.), Chichester, Novartis Foundation 2005, John Wiley & Sons.
28. Buciński A. 2003. Sztuczne sieci neuronowe – współczesne narzędzie przetwarzania informacji. *Farmacja Polska* 59(5): 196-203.
29. Burashnikov A, Antzelevitch C. 2002. Prominent IKs in epicardium and endocardium contributes to development of transmural dispersion of repolarization but protects against development of early afterdepolarizations. *J Cardiovasc Electrophysiol* 13: 172–177.
30. Busch A i in. 1998. Blockade of HERG channels by the class III antiarrhythmic azimilide: Mode of action. *Br J Pharmacol* 123(1): 23-30.
31. Caballero R i in. 2003. Spironolactone and its main metabolite, canrenoic acid, block human ether-a-go-go-related gene channels. *Circulation* 107(6): 889-895.
32. Cahill SA, Gross GJ. 2004. Propafenone and its metabolites preferentially inhibit IKr in rabbit ventricular myocytes. *J Pharmacol Exp Ther* 308(1): 59-65.
33. Cavalli A i in. 2002. Toward a pharmacophore for drugs inducing the long QT syndrome: insights from a CoMFA study of HERG K(+) channel blockers. *J Med Chem* 45(18): 3844-3853.
34. CDER Report to the Nation: Improving Public Health through Human Drugs. 2000-2007. <http://www.fda.gov/CDER/about/default.html>
35. CEPA. 1999. Canadian Environmental Protection Act. Opublikowany przez Ministra sprawiedliwości: <http://laws-lois.justice.gc.ca>.
36. Chachin M i in. 1999. Epinastine, a nonsedating histamine H1 receptor antagonist, has a negligible effect on HERG channel. *Eur J Pharmacol* 374(3): 457-460.
37. Chapman H, Pasternack M. 2007. The action of the novel gastrointestinal prokinetic prucalopride on the HERG K+ channel and the common T897 polymorph. *Eur J Pharmacol* 554(2-3): 98-105
38. Chekmarev DS. i in. 2008. Shape Signatures: New Descriptors for Predicting Cardiotoxicity *In silico*. *Chem Res Toxicol* 21: 1304–1314.

39. Chen S, Jiang M, Zhen Y. 2005. HERG K<sup>+</sup> channel expression-related chemosensitivity in cancer cells and its modulation by erythromycin. *Cancer Chemother Pharmacol* 56(2): 212-220.
40. Chin DN, Chuaqui CE, Singh J. 2004. Integration of virtual screening into the drug discovery process. *Mini Rev Med Chem* 4(10):1053–1065.
41. Choi S, Koh Y, Jo S. 2005. Inhibition of human ether-a-go-go-related gene K<sup>+</sup> channel and IKr of guinea pig cardiomyocytes by antipsychotic drug trifluoperazine. *J Pharmacol Exp Ther* 313(2): 888-895.
42. Christ T i in. 2008. Electrophysiological profile of propiverine - Relationship to cardiac risk. *Naunyn-Schmiedebergs Arch Pharmacol* 376(6): 431-440.
43. Cianchetta G i in. 2005. Predictive models for hERG potassium channel blockers. *Bioorg Med Chem Lett* 15: 3637-3642.
44. Cichosz P. 2000. Systemy uczące się. Indukcja drzew decyzyjnych, 138-191. Wydawnictwa naukowo techniczne, Warszawa.
45. Claassen S, Zünkler B. 2005. Comparison of the effects of metoclopramide and domperidone on HERG channels. *Pharmacology* 74(1): 31-36.
46. CPM. 2003. Guideline on safety pharmacology studies for assessing the potential for delayed ventricular repolarization (QT interval prolongation) by human pharmaceuticals. The European Agency for the Evaluation of Medicinal Products CPMP/ICH/423/02.
47. Collier R. 2009. Drug development cost estimates hard to swallow. *CMAJ* 180(3):279-280.
48. Committee on Safety of Medicines-Medicines Control Agency. 1995. Cardiac arrhythmias with pimozide (Orap). *Curr Probl Pharmacov* 21: 1.
49. Cordes JS. 2005. Novel potent human ether-a-go-go-related gene (hERG) potassium channel enhancers and their *in vitro* antiarrhythmic activity. *Mol Pharmacol* 68(3): 876-884.
50. Crumb Jr.W. 2000. Loratadine blockade of K<sup>+</sup> channels in human heart: Comparison with terfenadine under physiological conditions. *J Pharmacol Exp Ther* 292(1): 261-264.
51. Crumb W i in. 2006. Effects of cyamemazine on hERG, I<sub>Na</sub>, I<sub>Ca</sub>, I<sub>to</sub>, I<sub>sus</sub> and IK<sub>1</sub> channel currents, and on the QT<sub>c</sub> interval in guinea pigs. *Eur J Pharmacol* 532(3): 270-278.
52. Crumb W, Cavero I. 1999. QT interval prolongation by non-cardiovascular drugs: issues and solutions for novel drug development. *Pharm Sci Technolo Today* 2(7): 270-280.
53. Cuypers E i in. 2008. Gambierol, a toxin produced by the dinoflagellate *Gambierdiscus toxicus*, is a potent blocker of voltage-gated potassium channels. *Toxicon* 51: 974–983.
54. Danielsson B i in. 2003. Phenytoin and phenobarbital inhibit human HERG potassium channels. *Epilepsy Res* 55(1-2): 147-157.
55. Danielsson B i in. 2005. Effects of the antiepileptic drugs lamotrigine, topiramate and gabapentin on hERG potassium currents. *Epilepsy Res* 63(1): 17-25.
56. Darpö B, Nebout T, Sager P. 2006. Clinical evaluation of QT/QT<sub>c</sub> prolongation and proarrhythmic potential for nonantiarrhythmic drugs: The International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use E14 guideline. *J Clin Pharmacol* 46(5): 498-507.
57. Darpö B. 2001. Spectrum of drugs prolonging QT interval and the incidence of torsades de pointes. *Eur Heart J* 3(Suppl.K): K70-K80.



58. Davie C i in. 2004. Comparative pharmacology of guinea pig cardiac myocyte and cloned hERG (IKr) channel. *J. Cardiovasc Electrophysiol* 15(11): 1302-1309.
59. De Bruin ML i in. 2005. Anti-HERG activity and the risk of drug-induced arrhythmias and sudden death. *Eur Heart J* 26: 590–597.
60. Deisemann H i in. 2008. Effects of common antitussive drugs on the hERG potassium channel current. *J Cardiovasc Pharmacol* 52(6): 494-499.
61. Dennis A i in. 2007. hERG channel trafficking: novel targets in drug-induced long QT syndrome. *Biochem Soc Trans* 35(5): 1060-1063.
62. Diller D, Hobbs D. 2007. Understanding hERG inhibition with QSAR models based on a one-dimensional molecular representation. *J Comp-Aided Mol Des* 21(7): 379-393.
63. DiMasi JA, Hansen RW, Grabowski HG. 2003. The price of innovation: new estimates of drug development costs. *J Health Econ* 22: 151–185.
64. DiMasi JA. 2002. The value of improving the productivity of the drug development process: faster times and better decisions. *Pharmacoeconomics*. 20(3):1-10.
65. Dong D i in. 2004. Acidification alters antiarrhythmic drug blockade of the ether-a-go-go-related gene (HERG) channels. *Basic Clin Pharmacol Toxicol* 94(5): 209-212.
66. Dorian P, Newman D. 2000. Rate dependence of the effect of antiarrhythmic drugs delaying cardiac repolarization: an overview. *Europace* 2: 277–285.
67. Dorn A i in. 2005. Evaluation of a High-Throughput Fluorescence Assay Method for hERG Potassium Channel Inhibition. *J Biom Screen* 10(4): 339-347.
68. Drolet B i in. 1998. Block of the rapid component of the delayed rectifier potassium current by the prokinetic agent cisapride underlies drug-related lengthening of the QT interval. *Circulation* 97(2): 204-210.
69. Drolet B i in. 1999. Droperidol lengthens cardiac repolarization due to block of the rapid component of the delayed rectifier potassium current. *J Cardiovasc Electrophysiol* 10(12): 1597-1604.
70. Drolet B i in. 2000. Domperidone should not be considered a no-risk alternative to cisapride in the treatment of gastrointestinal motility disorders. *Circulation* 102(16): 1883-1885.
71. Drolet B i in. 2003. Risperidone prolongs cardiac repolarization by blocking the rapid component of the delayed rectifier potassium current. *J Cardiovasc Pharmacol* 41(6): 934-937.
72. Du L i in. 2004. The pharmacophore hypotheses of IKr potassium channel blockers: Novel class III antiarrhythmic agents. *Bioorg Med Chem Lett* 14(18): 4771-4777.
73. Du L i in. 2007. A novel structure-based virtual screening model for the hERG channel blockers. *Biochem Biophys Res Commun* 355(4): 889-894.
74. Du X i in. 2006. Dimethyl sulfoxide effects on hERG channels expressed in HEK293 cells. *J Pharmacol Toxicol Meth* 54(2): 164-172.
75. Duan J i in. 2007. Verapamil blocks HERG channel by the helix residue Y652 and F656 in the S6 transmembrane domain. *Acta Pharmacol Sin* 28(7): 959-967.
76. Dubus E i in. 2006. *In silico* Classification of hERG Channel Blockers: a Knowledge-Based Strategy. *Chem Med Chem* 1(6): 622 – 630.

77. Duch W i in. red. 2000. Biocybernetyka i Inżynieria Biomedyczna. Tom 6 Sieci Neuronowe. Akademicka Oficyna Wydawnicza EXIT, Warszawa.
78. Ducroq J i in. 2007. Action potential experiments complete hERG assay and QT-interval measurements in cardiac preclinical studies. *J Pharmacol Toxicol Meth* 56(2): 159-170.
79. Ducroq J i in. 2010. Dexrazoxane protects the heart from acute doxorubicin-induced QT prolongation: a key role for IKs. *Br J Pharmacol* 159: 93-101.
80. Ducroq J, Printemps R, Le Grand M. 2005. Additive effects of ziprasidone and D,L-sotalol on the action potential in rabbit Purkinje fibres and on the hERG potassium current. *J Pharmacol Toxicol Meth* 52(1): 115-122.
81. Dumaine R, Cordeiro JM. 2007. Comparison of K<sup>+</sup> currents in cardiac Purkinje cells isolated from rabbit and dog. *J Mol Cell Cardiol* 42: 378-389.
82. Dumaine R, Roy M, Brown A. 1998. Blockade of HERG and Kv1.5 by ketoconazole. *J Pharmacol Exp Ther* 286(2): 727-735.
83. Dumotier BM, Georgieva AV. 2007. Preclinical cardio-safety assessment of torsadogenic risk and alternative methods to animal experimentation: The inseparable twins. *Cell Biol Toxicol* 23: 293-302.
84. Duncan R i in. 2006. Erythromycin block of the HERG K<sup>+</sup> channel: Accessibility to F656 and Y652. *Biochem Biophys Res Commun* 341(2): 500-506.
85. Duncan R i in. 2007. Inhibition of the HERG potassium channel by the tricyclic antidepressant doxepin. *Biochem. Pharmacol.* 74(3): 425-437.
86. Dupuis D, Klaerke D, Olesen S. 2005. Effect of  $\beta$ -adrenoceptor blockers on Human Ether-a-go-go-Related Gene (HERG) potassium channels. *Basic Clin Pharmacol Toxicol* 96(2): 123-130.
87. Dustan Sarazan R i in. 2004. Absence of clinically important HERG channel blockade by three compounds that inhibit phosphodiesterase 5 - Sildenafil, tadalafil, and vardenafil. *Eur J Pharmacol* 502(3): 163-167.
88. Eap C i in. 2007. Stereoselective block of hERG channel by (S)-methadone and QT interval prolongation in CYP2B6 slow metabolizers. *Clin Pharmacol Ther* 81(5): 719-728.
89. Ekins S i in. 2002. Towards a new age of virtual ADME/TOX and multidimensional drug discovery. *Mol Divers* 5: 255-275.
90. Ekins S, Balakin K, Savchuk N, Ivanenkov Y. 2006a. Insights for human ether-a-go-go-related gene potassium channel inhibition using recursive partitioning and Kohonen and Sammon mapping techniques. *J Med Chem* 49(17): 5059-5071.
91. Ekins S i in. 2002a. Three-dimensional quantitative structure-activity relationship for inhibition of human ether-a-go-go-related gene potassium channel. *J Pharmacol Exp Ther* 301(2): 427-434.
92. Ekins S, Rose J. 2002b. Meeting report. *In silico* ADME/Tox: the state of the art. *J Mol Graph Model* 20: 305-309.
93. EMEA (European Agency for the Evaluation of Medicinal Products). Committee for Proprietary Medicinal Products (CPMP). 1997. Points to consider: The assessment of the potential for QT interval prolongation by non-cardiovascular medicinal products.
94. EU. 2006. Regulation (EC) no. 1907/2006 of the European Parliament and of the council concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European chemicals agency, amending directive 1999/45/ec and repealing

- council regulation (EEC) no. 793/93 and commission regulation (EC) no. 1488/94 as well as council directive 76/769/EEC and commission directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. *Offic. J. Eur. Union* L396, 1–843.
95. Fanoë S i in. 2009. Oxycodone is associated with dose-dependent QTc prolongation in patients and low-affinity inhibiting of hERG activity *in vitro*. *Br J Clin Pharmacol* 67(2): 172–179.
  96. FDA. 2003. The clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs. Preliminary concept paper. FDA (January 28 2003).
  97. Federal Trade Commission. 2007. Pharma 2020: Virtual R&D. Which path will you take? PriceWaterhouseCoopers. <http://www.pwc.com/pharma>
  98. Fenichel RR i in. 2004. Drug-Induced Torsades de Pointes and Implications for Drug Development. *J Cardiovasc Electrophysiol* 15(4): 476-495.
  99. Fenichel RR. <http://www.fenichel.net/pages/Professional/subpages/QT/Tables/pbydrug.htm> [30.05.2009]
  100. Fermini B, Fossa AA. 2003. The impact of drug-induced QT interval prolongation on drug discovery and development. *Nat Rev. Drug Discov* 2(6): 439–447.
  101. Ferreira S i in. 2001. Effects of cocaine and its major metabolites on the HERG-encoded potassium channel. *J Pharmacol Exp Ther* 299(1): 220-226.
  102. Ferrer-Villada T i in. 2006. Inhibition of cardiac HERG potassium channels by antidepressant maprotiline. *Eur J Pharmacol* 531(1-3): 1-8.
  103. Ficker E i in. 1998. Molecular determinants of dofetilide block of HERG K<sup>+</sup> channels. *Circ Res* 82(3): 386-395.
  104. Finlayson K i in. 2001a. [<sup>3</sup>H]dofetilide binding to HERG transfected membranes: a potential high throughput preclinical screen. *Eur J Pharmacol* 430(1): 147– 148.
  105. Finlayson K i in. 2004. Acquired QT interval prolongation and HERG: implications for drug discovery and development. *E J Pharm* 500: 129–142.
  106. Fioravanzo E i in. 2004. General and independent approaches to predict HERG affinity values. *Internet Electron J Mol Des* 4(9): 625–646.
  107. Fossa A i in. 2004. Differential effect of HERG blocking agents on cardiac electrical alternans in the guinea pig. *Eur J Pharmacol* 486(2): 209-221.
  108. Friederich P, Solth A, Pongs O. 2001. Ropivacaine Inhibits HERG Potassium Channels Cloned from Human Heart. *Anesthesiology* 95: A904.
  109. Friederich P, Solth A, Schillemeit S, Isbrandt D. 2004. Local anaesthetic sensitivities of cloned HERG channels from human heart: Comparison with HERG/MiRP1 and HERG/MiRP1T8A. *Br J Anaesth* 92(1): 93-101.
  110. Fung M i in. 2001. Evaluation of the characteristics of safety withdrawal of prescription drugs from worldwide pharmaceutical markets – 1960 to 1999. *Drug Inform J* 35: 293–317.
  111. Furuta S i in. 2004. Drug–drug interactions of Z-338, a novel gastroprokinetic agent, with terfenadine, comparison with cisapride, and involvement of UGT1A9 and 1A8 in the human metabolism of Z-338. *Eur J Pharmacol* 497: 223–231.

112. Gautier P i in. 2004. *In vivo* and *In vitro* Characterization of the Novel Antiarrhythmic Agent SSR149744C. Electrophysiological, Anti-Adrenergic, and Anti-Angiotensin II Effects. *J Cardiovasc Pharmacol* 44(2): 244-257.
113. Gavaghan CL i in. 2007. Development, interpretation and temporal evaluation of a global QSAR of hERG electrophysiology screening data. *J Comput Aided Mol Des* 21:189–206.
114. Geelen P i in. 2000. Sildenafil (viagra) prolongs cardiac repolarization by blocking the rapid component of the delayed rectifier potassium current. *Circulation* 102(3): 275-277.
115. Gepp MM, Hutter MC. 2006. Determination of hERG channel blockers using a decision tree. *Bioorg Med Chem* 14: 5325-5332.
116. Gintant A i in. 2006. Utility of hERG Assays as Surrogate Markers of Delayed Cardiac Repolarization and QT Safety. *Toxicologic Path* 34: 81-90.
117. Gintant G. 1998. Azimilide causes reverse rate-dependent block while reducing both components of delayed-rectifier current in canine ventricular myocytes. *J Cardiovasc Pharmacol* 31(6): 945-953.
118. Gögelein H i in. 2004. Effects of the atrial antiarrhythmic drug AVE0118 on cardiac ion channels. *Naunyn-Schmiedebergs Arch Pharmacol* 370(3): 183-192.
119. González T i in. 2002. Effects of levobupivacaine, ropivacaine and bupivacaine on HERG channels: Stereoselective bupivacaine block. *Br J Pharmacol* 137(8): 1269-1279.
120. Greenberg M, Lowrie K. 2009. Toxicity Testing in the 21st Century. *Risk Analysis* 29(4): 471-473.
121. Gu D i in. 2009. Blockade of HERG K<sup>+</sup> channel by isoquinoline alkaloid neferine in the stable transfected HEK293 cells. *Naunyn-Schmied Arch Pharmacol* 380:143–151.
122. Guo J i in. 2010. The Action of a Novel Fluoroquinolone Antibiotic Agent Antofloxacin Hydrochloride on Human-Ether-go-go-Related Gene Potassium Channel. *Basic Clin Pharmacol Toxicol* 107: 643–649.
123. Guo J, Gang H, Zhang S. 2006. Molecular determinants of cocaine block of human Ether-a-go-go-related gene potassium channels. *J Pharmacol Exp Ther* 317(2): 865-874.
124. Guo L, Guthrie H. 2005. Automated electrophysiology in the preclinical evaluation of drugs for potential QT prolongation. *J Pharmacol Toxicol Meth* 52(1): 123-135.
125. Gupta A i in. 2007. Current concepts in the mechanisms and management of drug-induced QT prolongation and torsade de pointes. *Am Heart J* 153(6): 891-899.
126. Guth B, Germeyer S, Kolb W, Markert M. 2004. Developing a strategy for the nonclinical assessment of proarrhythmic risk of pharmaceuticals due to prolonged ventricular repolarization. *J Pharmacol Toxicol Meth* 49(3): 159-169.
127. Guth B. 2007. Preclinical cardiovascular risk assessment in modern drug development. *Toxicol Sci* 97(1): 4-20.
128. Haddad P, Anderson I. 2002. Antipsychotic-related QTc prolongation, torsade de pointes and sudden death. *Drugs* 62(11): 1649-1671.
129. Han S. 2011. Fluconazole inhibits hERG K<sup>+</sup> channel by direct block and disruption of protein trafficking. *Eur J Pharmacol* 650(1): 138-14.
130. Hancox JC i in. 2008. The hERG potassium channel and hERG screening for drug-induced torsades de pointes. *Pharmacol Therap* 119: 118–132.

131. Haverkamp W i in. 2000. The potential for QT prolongation and pro-arrhythmia by non-anti-arrhythmic drugs: clinical and regulatory implications Report on a Policy Conference of the European Society of Cardiology. *Cardiovasc Res* 47: 219–233 i *Eur Heart J* 21: 1216–1231.
132. Hiramatsu M i in. 2004. Block of HERG current expressed in HEK293 cells by the Na<sup>+</sup>-channel blocker cibandoline. *Heart Vessels* 19(3): 137-143.
133. Hoffmann P, Warner B. 2006. Are hERG channel inhibition and QT interval prolongation all there is in drug-induced torsadogenesis? a review of emerging trends. *J Pharm Toxicol Meth* 53: 87-105.
134. Hong H-K i in. 2010. Block of the human ether-a-go-go-related gene (hERG) K<sup>+</sup> channel by the antidepressant desipramine. *Biochem Biophys Res Comm* 394: 536–541.
135. Hong H-K, Jo S-H. 2009. Block of hERG K<sup>+</sup> Channel by Classic Histamine H1 Receptor Antagonist Chlorpheniramine. *Korean J Physiol Pharmacol* 13: 215-220.
136. Honig PK i in. 1993. Terfenadine-ketoconazole interaction. Pharmacokinetic and electrocardiographic consequences. *JAMA* 269: 1513-1518.
137. <http://www.chemaxon.com/>
138. <http://www.cs.waikato.ac.nz/~ml/weka/>
139. <http://www.ncbi.nlm.nih.gov/pubmed/>
140. <http://www.tox-comp.net>
141. Huang X i in. 2008. The Effects of a Novel Anti-arrhythmic Drug, Acehytisine Hydrochloride, on the Human Ether-a-go-go Related Gene K<sup>+</sup> Channel and Its Trafficking. *Basic Clin Pharmacol Toxicol* 104: 145-154.
142. Hunt BR, Kubler O. 1984. Karhunen–Loeve multispectral image restoration, Part 1: theory, *IEEE Trans. Acoust. Speech Signal Process* 32(3): 592–600.
143. Hurst R i in. 2003. Dopamine receptor agonists differ in their actions on cardiac ion channels. *Eur J Pharmacol* 482(1-3): 31-37.
144. ICH. 2005. Guidance for Industry S7B Nonclinical Evaluation of the Potential for Delayed Ventricular Repolarization (QT Interval Prolongation) by Human Pharmaceuticals.
145. Ishii K i in. 2003. Dissociation of E-4031 from the HERG channel caused by mutations of an amino acid results in greater block at high stimulation frequency. *Cardiovasc Res* 57: 651-659.
146. Ivanciuc O. 2006. Artificial Immune System Classification of Drug–induced Torsade de Pointes with AIRS (Artificial Immune Recognition System). *Internet Electron J Mol Des* 5: 488–502.
147. Jacobson ED. 2004. AdvaMed’s Written Testimony on The Critical Path White Paper at Science Board to the Food and Drug Administration. Food and Drug Administration, Rockville.
148. Jeong I, Choi BH, Hahn SJ. 2010. Effects of lobeline, a nicotinic receptor ligand, on the cloned Kv1.5. *Pflugers Arch - Eur J Physiol* 460:851–862.
149. Jia L, Sun H. 2008. Support vector machines classification of hERG liabilities based on atom types. *Bioorg Med Chem* 16: 6252–6260.
150. Jo S i in. 2007. Maprotiline block of the human ether-a-go-go-related gene (HERG) K<sup>+</sup> channel. *Arch Pharmacol Res* 30(4): 453-460.

151. Jo S i in. 2000. Blockade of the HERG human cardiac K<sup>+</sup> channel by the antidepressant drug amitriptyline. *Br J Pharmacol* 129: 1474-1480.
152. Jo S-H i in. 2008. Clomipramine block of the hERG K<sup>+</sup> channel: Accessibility to F656 and Y652. *Eur J Pharmacol* 592: 19-25.
153. Jo S-H i in. 2008a. Protriptyline block of the human ether-à-go-go-related gene (HERG) K<sup>+</sup> channel. *Life Sci* 82: 331-340.
154. Jo S-H i in. 2009. H1 antihistamine drug promethazine directly blocks hERG K<sup>+</sup> channel. *Pharmacol Res* 60: 429-437.
155. Jurkiewicz N i in. 1996. Mechanism of action potential prolongation by RP 58866 and its active enantiomer, terikalant: Block of the rapidly activating delayed rectified K<sup>+</sup> current, I(Kr). *Circulation* 94(11): 2938-2946.
156. Kakusaka S i in. 2009. A Receptor-Independent Effect of Estrone Sulfate on the hERG Channel. *J Pharmacol Sci* 109; 152-156.1
157. Kamiya K i in. 2001. Open channel block of HERG K<sup>+</sup> channels by vesnarinone. *Mol Pharmacol* 60(2): 244-253.
158. Kamiya K i in. 2001a. Short- and long-term effects of amiodarone on the two components of cardiac delayed rectifier K<sup>+</sup> current. *Circulation* 103(9): 1317-1324.
159. Kamiya K, Niwa R, Mitcheson J, Sanguinetti M. 2006. Molecular determinants of hERG channel block. *Mol Pharmacol* 69(5): 1709-1716.
160. Kan SH. 2002. Metrics and Models in Software Quality Engineering. Metrics and Models in Software Quality Engineering 2<sup>nd</sup> Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA.
161. Kang J i in. 2001a. Interactions of a series of fluoroquinolone antibacterial drugs with the human cardiac K<sup>+</sup> channel HERG. *Mol Pharmacol* 59(1): 122-126.
162. Kang J in. 2004. Cardiac Ion Channel Effects of Tolterodine. *J Pharmacol Exp Ther* 308(3): 935-940.
163. Kang J, Chen X, Wang L, Rampe D. 2001. Interactions of the antimalarial drug mefloquine with the human cardiac potassium channels KvLQT1/minK and HERG. *J Pharmacol Exp Ther* 299(1): 290-296.
164. Kang J, Wang L, Cai F, Rampe D. 2000. High affinity blockade of the HERG cardiac K<sup>+</sup> channel by the neuroleptic pimozide. *Eur J Pharmacol* 392(3): 137-140.
165. Karczewski J i in. 2009. Analogs of MK-499 are differentially affected by a mutation in the S6 domain of the hERG K<sup>+</sup> channel. *Biochem Pharmacol* 77: 1602-1611.
166. Karle C i in. 2001. Antiarrhythmic drug carvedilol inhibits HERG potassium channels. *Cardiovasc Res* 49(2): 361-370.
167. Katayama Y i in. 2000. Inhibitory effects of vesnarinone on cloned cardiac delayed rectifier K(+) channels expressed in a mammalian cell line. *J Pharmacol Exp Ther* 294(1): 339-346.
168. Katchman A i in. 2002. Influence of opioid agonists on cardiac human ether-a-go-go-related gene K<sup>+</sup> currents. *J Pharmacol Exp Ther* 303(2): 688-694.
169. Katchman A i in. 2006. Comparative evaluation of HERG currents and QT intervals following challenge with suspected torsadogenic and nontorsadogenic drugs. *J Pharmacol Exp Ther* 316(3): 1098-1106.

170. Kawakami K i in. 2006. Comparison of HERG channel blocking effects of various beta-blockers- implication for clinical strategy. *Br J Pharmacol* 147(6): 642-652.
171. Kelemen K i in. 2007. Green tea flavonoid epigallocatechin-3-gallate (EGCG) inhibits cardiac hERG potassium channels. *Biochem Biophys Res Comm* 364: 429–435.
172. Keserü GM. 2003. Prediction of hERG potassium channel affinity by traditional and hologram qSAR methods. *Bioorg Med Chem Lett* 13(16): 2773-2775.
173. Kiehn J i in. 1999. Inhibitory effects of the class III antiarrhythmic drug amiodarone on cloned HERG potassium channels. *Naunyn-Schmiedeberg Arch Pharmacol* 359(3): 212-219.
174. Kiehn J, Lacerda AE, Wible B, Brown AM. 1996. Molecular physiology and pharmacology of HERG single-channel currents and block by dofetilide. *Circulation* 94: 2572-2579.
175. Kiesecker C i in. 2004. Class Ia anti-arrhythmic drug ajmaline blocks HERG potassium channels: Mode of action. *Naunyn-Schmiedeberg Arch Pharmacol* 370(6): 423-435.
176. Kiesecker C i in. 2006. Atypical tetracyclic antidepressant maprotiline is an antagonist at cardiac hERG potassium channels. *Naunyn-Schmiedeberg Arch Pharmacol* 373(3): 212-220.
177. Kijawornrat A. 2007. Potential mechanisms for drug-induced prolongation of QT interval and genesis of torsades de pointes evaluated in the failing rabbit heart. Doctoral thesis, Ohio State University.
178. Kikuchi K i in. 2005. Blockade of HERG cardiac K<sup>+</sup> current by antifungal drug miconazole. *Br J Pharmacol* 144(6): 840-848.
179. Kim C i in. 2007. Inhibitory effects of coronary vasodilator papaverine on heterologously-expressed HERG currents in *Xenopus* oocytes. *Acta Pharmacol Sin* 28(4): 503-510.
180. Kim E, Kim K, Shin W. 2005. Electrophysiological safety of DW-286a, a novel fluoroquinolone antibiotic agent. *Hum Exp Toxicol* 24(1): 19-25.
181. Kim K, Kim E. 2005a. The phenothiazine drugs inhibit hERG potassium channels. *Drug Chem Toxicol* 28(3): 303-313.
182. Kim K, Shin W, Park S, Kim E. 2007a. Effect of clebopride, antidopaminergic gastrointestinal prokinetics, on cardiac repolarization. *Int J Toxicol* 26(1): 25-31.
183. Kim K-S i in. 2008. Electrophysiological safety of sibutramine HCl. *Humm Expl Toxicol* 27: 553–558.
184. Kim K-S i in. 2010. Blockade of hERG K<sup>+</sup> Channel by Antimalarial Drug, Primaquine. *Arch Pharm Res* 33 (5): 769-773.
185. Kim K-S, Kim E-J, Lee H-A, Park S-J. 2009. Effect of sibutramine HCl on cardiac hERG K<sup>+</sup> channel. *Mol Cell Biochem* 320:125–131.
186. Kim M, Eun S, Jo S. 2006. Blockade of HERG human K<sup>+</sup> channel and IKr of guinea pig cardiomyocytes by prochlorperazine. *Eur J Pharmacol* 544(1-3): 82-90.
187. Kim YJ i in. 2008c. Papaverine, a vasodilator, blocks the pore of the HERG channel at submicromolar concentration. *J Cardiovasc Pharmacol* 52(6): 485-93.
188. Kirsch GE i in. 2004. Variability in the measurement of hERG potassium channel inhibition: Effects of temperature and stimulus pattern. *J. Pharmacol. Toxicol. Meth.* 50: 93-101.
189. Ko CM i in. 1996. Suppression of Mammalian K1 Channel Family by Ebastine. *J Pharmacol Exp Ther* 281(1): 233-281.

190. Kola I, Landis J. 2004. Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov* 2004;3 (8):711–715.
191. Kongsamut S i in. 2002. A comparison of the receptor binding and HERG channel affinities for a series of antipsychotic drugs. *Eur J Pharmacol* 450(1): 37-41.
192. Korbicz J, Obuchowski A. 1994. Uciński D. Sztuczne sieci neuronowe. Podstawy i zastosowania. Akademicka Oficyna Wydawnicza PLJ, Warszawa.
193. Kornick C i in. 2003. QTc interval prolongation associated with intravenous methadone. *Pain* 105(3): 499-506.
194. Kruhlak NL i in. 2007. Progress in QSAR toxicity screening of pharmaceutical impurities and other FDA regulated products. *Advanced Drug Delivery Reviews* 59: 43–55.
195. Kumar N i in. 2006. Applying computational modeling to drug discovery and development. *Drug Discov Today* 11(17–18): 806–811.
196. Kurokawa J i in. 2008. Acute effects of oestrogen on the guinea pig and human IKr channels and drug-induced prolongation of cardiac repolarization. *J Physiol* 586(12) 2961–2973.
197. Kuryshev Y i in. 2000. Interactions of the 5-hydroxytryptamine 3 antagonist class of antiemetic drugs with human cardiac ion channels. *J Pharmacol Exp Ther* 295(2): 614-620.
198. Kuryshev Y i in. 2005. Pentamidine-induced long QT syndrome and block of hERG trafficking. *J Pharmacol Exp Ther* 312(1): 316-323.
199. Kushida S, Ogura T, Komuro I, Nakaya, H. 2002. Inhibitory effect of the class III antiarrhythmic drug nifekalant on HERG channels: Mode of action. *Eur J Pharmacol* 457(1): 19-27.
200. Lacroix P i in. 2003. Prulifloxacin: *in vitro* (HERG current) and *in vivo* (conscious dog) assessment of cardiac risk. *Eur J Pharmacol* 477: 69– 72.
201. Lankipalli RS i in. 2005. Mechanisms underlying arrhythmogenesis in long QT syndrome. *J Electrocardiol* 38: 69–73.
202. Lasser KE i in. 2002. Timing of new black box warnings and withdrawals for prescription medications. *JAMA* 287: 2215–2220.
203. Lee S i in-. 2004. Block of HERG Human K<sup>+</sup> Channel and IKr of Guinea Pig Cardiomyocytes by Chlorpromazine. *J Cardiovasc Pharmacol* 43(5): 706-714.
204. Lee S i in. 2006. Blockade of HERG human K<sup>+</sup> channels and I(Kr) of guinea-pig cardiomyocytes by the antipsychotic drug clozapine. *Br J Pharmacol* 148(4): 499-509.
205. Lees-Miller J, Duan Y, Teng G, Duff H. 2000. Molecular determinant of high-affinity dofetilide binding to HERG1 expressed in *Xenopus* oocytes: Involvement of S6 sites. *Mol Pharmacol* 57(2): 367-374.
206. Lexchin J. 2005. Drug withdrawals from the Canadian market for safety reasons, 1963–2004. *Can Med Assoc J* 172 (6): 765-767.
207. Li B i in. 2001. Inhibitory effects of berberine on IK1, IK, and HERG channels of cardiac myocytes. *Acta Pharmacol Sin* 22(2): 125-131.
208. Li Q i in. 2008. hERG Classification Model Based on a Combination of Support Vector Machine Method and GRIND Descriptors. *Mol. Pharm.* 5(1): 117-127.
209. Li Y i in. 2010. Aconitine blocks HERG and Kv1.5 potassium channels. *J Ethnopharmacol* 131: 187–195.



210. Lin C i i. 2008b. The Effects of Chiral Isolates of Methadone on the Cardiac Potassium Channel IKr. *Cardiology* 113(1):59-65.
211. Lin C i, Ke X, Cvetanovic I, Ranade V, Somberg J. 2005. The Influence of Extracellular Acidosis on the Effect of IKr Blockers., *J Cardiovasc Pharmacol Therapeut.* 10(1): 67-76.
212. Lin C, Ke X, Cvetanovic I, Ranade V, Somberg J. 2007. The effect of high extracellular potassium on IKr inhibition by anti-arrhythmic agents. *Cardiology* 108(1): 18-27.
213. Lin C, Ke X, Ranade V, Somberg J. 2007a. Extracellular Acidification and Hyperkalemia Induce Changes in HERG Inhibition by Ibutilide. *Cardiology* 110(3): 209-216.
214. Lin C, Ke X, Ranade V, Somberg J. 2007b. The Additive Effects of the Active Component of Grapefruit Juice (Naringenin) and Antiarrhythmic Drugs on HERG Inhibition. *Cardiology* 110(3): 145-152.
215. Lin J i in. 2005x. Intracellular K<sub>+</sub> Is Required for the Inactivation-Induced High-Affinity Binding of Cisapride to HERG Channels. *Mol Pharmacol* 68(3): 855-865.
216. Lindquist M, Edwards IR. 1997. Risks of non-sedating antihistamines. *Lancet* 349: 1322.
217. Liu H i in. 2010. The selective estrogen receptor modulator raloxifene inhibits cardiac delayed rectifier potassium currents and voltage-gated sodium current without QTc interval prolongation. *Pharmacol Res* 62: 384–390.
218. Liu Z, Yan F. 1997. Fuzzy neural network in case-based diagnostic system, *IEEE Trans Fuzzy Syst* 5: 209–222.
219. Lu H i in. 2006. In-vitro experimental models for the risk assessment of antibiotic-induced QT prolongation. *Eur J Pharmacol* 553(1-3): 229-239.
220. Lu HR i in. 2008. Predicting drug-induced changes in QT interval and arrhythmias: QT-shortening drugs point to gaps in the ICHS7B Guidelines. *Br J Pharmacol* 154: 1427–1438.
221. Luo T i in. 2008. Inhibition of the HERG Channel by Droperidol Depends on Channel Gating and Involves the S6 Residue F656. *Anesth Analg* 106: 1161–1170.
222. Mancilla-Simbro C i in. 2008. Chlorthalidone inhibits the KvLQT1 potassium current in guinea-pig ventricular myocytes and oocytes from *Xenopus laevis*. *Br J Pharmacol* 153: 448–458.
223. Martin R i in. 2006. *In vitro* preclinical cardiac assessment of tolterodine and terodiline: Multiple factors predict the clinical experience. *J Cardiovasc Pharmacol* 48(5): 199-206.
224. Martin R i in. 2004. The Utility of hERG and Repolarization Assays in Evaluating Delayed Cardiac Repolarization: Influence of Multi-Channel Block. *J Cardiovasc Pharmacol* 43(3): 369-379.
225. Mbai M, Rajamani S, January C. 2002. The anti-malarial drug halofantrine and its metabolite N-desbutylhalofantrine block HERG potassium channels. *Cardiovasc Res* 55(4): 799-805.
226. McInnes C. 2007. Virtual screening strategies in drug discovery. *Curr Opin Chem Biol* 11: 494–502.
227. McPate M, Duncan R, Witchel H, Hancox J. 2006. Disopyramide is an effective inhibitor of mutant HERG K<sup>+</sup> channels involved in variant 1 short QT syndrome. *J Mol Cell Cardiol* 41(3): 563-566.
228. Meek ME, Armstrong VC. 2007. The assessment and management of industrial chemicals in Canada. *Risk Assessment of Chemicals: An Introduction.* van Leeuwen CJ i Vermeire TG (wyd.), s591–621.

229. Mendyk A. 2004. Sztuczne sieci neuronowe jako uniwersalne narzędzia do modelowania w technologii postaci leku i biofarmacji. Praca doktorska, UJCM, Katedra technologii Postaci Leku i Biofarmacji, Kraków.
230. Mergenthaler J i in. 2001. Blocking effects of the antiarrhythmic drug propafenone on the HERG potassium channel. *Naunyn-Schmiedebergs Arch Pharmacol* 363(4): 472-480.
231. Milberg P i in. 2002. Divergent proarrhythmic potential of macrolide antibiotics despite similar QT prolongation: Fast phase 3 repolarization prevents early afterdepolarizations and torsade de pointes. *J Pharmacol Exp Ther* 303(1): 218-225.
232. Milnes J i in. 2006. hERG K<sup>+</sup> channel blockade by the antipsychotic drug thioridazine: An obligatory role for the S6 helix residue F656. *Biochem Biophys Res Commun* 351(1): 273-280.
233. Milnes JT i in. 2003. Preferential closed channel blockade of HERG potassium currents by chemically synthesised BeKm-1 scorpion toxin. *FEBS Lett* 547 (1-3): 20-26.
234. Milnes JT i in. 2010. Investigating dynamic protocol-dependence of hERG potassium channel inhibition at 37 °C: Cisapride versus dofetilide. *J Pharmacol Toxicol Meth* 61 (2010) 178-191.
235. Modi S. 2003. Computational approaches to the understanding of ADMET properties and problems. *Drug Discov Today* 8(14): 621-623.
236. Mohammad S, Zhou Z, Gong Q, January C. 1997. Blockage of the HERG human cardiac K<sup>+</sup> channel by the gastrointestinal prokinetic agent cisapride. *Am. J Physiol Heart Circ. Physiol.* 273: H2534 - H2538.
237. Molleman A. 2003. Patch Clamping. An Introductory Guide to Patch Clamp Electrophysiology. John Willey&Sons, the Atrium, Southern Gate, Chichester, West Sussex, England.
238. Möller C, Netzer R. 2006. Effects of estradiol on cardiac ion channel currents. *Eur J Pharmacol* 532: 44-49.
239. Moreno I i in. E. 2003. Effects of irbesartan on cloned potassium channels involved in human cardiac repolarization. *J Pharmacol Exp Ther* 304(2): 862-873.
240. Muster W i in. 2008. Computational toxicology in drug development. *Drug Discov Today* 13: 303-310.
241. Nakamura Y, Takahara A, Sugiyama A. 2009. Famotidine neither affects action potential parameters nor inhibits human ether-a-go-go-related gene (hERG) K<sup>+</sup> current. *J Toxicol Sci* 34(5): 563-567.
242. Nałęcz M. (red.) 2000. Biocybernetyka i inżynieria biomedyczna 2000. Tom 6 Sieci neuronowe. Akademicka Oficyna Wydawnicza EXIT, Warszawa.
243. Nattel S. 1999. The molecular and ionic specificity of antiarrhythmic drug actions. *J Cardiovasc Electrophysiol* 10(2): 272-282.
244. Neher E, Sakmann B. 1992. The patch clamp technique. *Sci Am* 266(3): 44-51.
245. Netzer R i in. 2001. Screening lead compounds for QT interval prolongation. *Drug Discov Today* 6 (2): 78- 84.
246. Netzer R, Bischoff U, Ebnet A. 2003. HTS techniques to investigate the potential effects of compounds on cardiac ion channels at early-stages of drug discovery. *Curr Opin Drug Discov Devel* 6(4): 462-469.

247. Niwa R i in. 2003. Voltage-Dependent Effects of Amiodarone on D540K HERG Channels. *Environmental medicine: annual report of the Research Institute of Environmental Medicine, Nagoya University* 47: 51-52.
248. Nosten F i in. 1993. Cardiac effects of antimalarial treatment with halofantrine. *Lancet* 341: 1054-1056.
249. Obers S i in. 2010. Multiple mechanisms of hERG liability: K<sup>+</sup> current inhibition, disruption of protein trafficking, and apoptosis induced by amoxapine. *Naunyn-Schmied Arch Pharmacol* 381: 385–400.
250. O'Brien SE, de Groot MJ. 2005. Greater than the sum of its parts: combining models for useful ADMET prediction. *J Med Chem* 48: 1287–1291.
251. Olivier P, Montastruc JL. 2006. The nature of the scientific evidence leading to drug withdrawals for pharmacovigilance reasons in France. *Pharmacoepidemiology and Drug Safety* 15: 808–812.
252. Olson HI. i in. 2000. Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regul Toxicol Pharmacol* 32: 56–67.
253. Osowski S. 1996. Sieci neuronowe w ujęciu algorytmicznym. Wydawnictwo Naukowo-Techniczne, Warszawa.
254. Park S-J, Kim K-S, Kim E-J. 2008. Blockade of HERG K<sup>+</sup> channel by an antihistamine drug brompheniramine requires the channel binding within the S6 residue Y652 and F656. *J Appl Toxicol* 28: 104–111.
255. Paul A, Witchel H, Hancox J. 2001. Inhibition of HERG potassium channel current by the class 1a antiarrhythmic agent disopyramide. *Biochem Biophys Res Commun* 280(5): 1243-1250.
256. Paul A, Witchel H, Hancox J. 2002. Inhibition of the current of heterologously expressed HERG potassium channels by flecainide and comparison with quinidine, propafenone and lignocaine. *Br J Pharmacol* 136(5): 717-729.
257. Pearl GM, Livingston-Carr S, Durham SK. 2001. Integration of computational analysis as a sentinel tool in toxicological assessments. *Curr Top Med Chem* 1: 247–255.
258. Pearlstein RA i in. 2003. Characterization of HERG Potassium Channel Inhibition Using CoMSiA 3D QSAR and Homology Modeling Approaches. *Bioorg Med Chem Lett* 13: 1829-1835.
259. Perrin MJ i in. 2008. Drug Binding to the Inactivated State Is Necessary but Not Sufficient for High-Affinity Binding to Human Ether-a`-go-go-Related Gene Channels. *Mol Pharmacol* 74(5): 1443-1452.
260. Perry M i in. 2006. Drug binding interactions in the inner cavity of hERG channels: Molecular insights from structure-activity relationships of clofilium and ibutilide analogs. *Mol Pharmacol* 69(2): 509-519.
261. Perry M i in. 2004. Structural determinants of HERG channel block by clofilium and ibutilide. *Mol Pharmacol* 66(2): 240-249.
262. Persson F, Carlsson L, Duker G, Jacobson I. 2005. Blocking Characteristics of hERG, hNav1.5, and hKvLQT1/hminK after Administration of the Novel Anti-Arrhythmic Compound AZD7009. *J Cardiovasc Electrophys* 16(3):329-341.
263. Pharmaceutical Research and Manufacturers of America. 2008. Pharmaceutical Industry Profile. Washington, DC: PhRMA.

264. Philips Z i in. 2006. Good Practice Guidelines for Decision-Analytic Modelling in Health Technology Assessment. A Review and Consolidation of Quality Assessment. *Pharmacoeconomics* 24(4): 355-371.
265. Po S i in. 1999. Modulation of HERG potassium channels by extracellular magnesium and quinidine. *J Cardiovasc Pharmacol* 33(2): 181-185.
266. Polak S, Wiśniowska B, Brandys J. 2009. Collation, assessment and analysis of literature *in vitro* data on hERG receptor blocking potency for subsequent modeling of drugs cardiotoxic properties. *J Appl Tox* 29(3): 183-206.
267. Raehl C, Patel A, LeRoy M. 1985. Drug-induced torsade de pointes. *Clin. Pharm.* 4(6), 675-690.
268. Rajamani R, Tounge B, Li J, Reynolds C. 2005. A two-state homology model of the hERG K<sup>+</sup> channel: Application to ligand binding. *Bioorg Med Chem Lett* 15(6): 1737-1741.
269. Rajamani S. 2006. Drug-induced long QT syndrome: hERG K<sub>t</sub> channel block and disruption of protein trafficking by fluoxetine and norfluoxetine. *Br J Pharmacol* 149: 481–489.
270. Rajamani S, Anderson C, Anson B, January C. 2002. Pharmacological rescue of human K<sup>+</sup> channel long-QT2 mutations: Human ether-a-go-go-related gene rescue without block. *Circulation* 105(24): 2830-2835.
271. Rajamani S, Shryock J, Belardinelli L. 2008. Rapid Kinetic Interactions of Ranolazine With HERG K<sup>+</sup> Current. *J Cardiovasc Pharmacol* 51(6): 581-589.
272. Raporty Joint Research Centre – Institute for Prospective Technological Studies. European Commission. 2006-2010. <http://iri.jrc.ec.europa.eu/research/scoreboard.htm>
273. Ray WA i in. 2004. Oral erythromycin and the risk of sudden death from cardiac causes. *N Engl J Med* 351: 1089–1096.
274. Recanatini M i in. 2005. QT prolongation through hERG K<sup>+</sup> channel blockade: current knowledge and strategies for the early prediction during drug development. *Med Res Rev* 25: 133–166.
275. Redfern WS i in. 2003. Relationships between preclinical cardiac electrophysiology, clinical QT interval prolongation and torsade de pointes for a broad range of drugs: evidence for a provisional safety margin in drug development. *Cardiovasc Res* 58: 32–45.
276. Renganathan M, Sidach S, Blight AR. 2009. Effects of 4-Aminopyridine on Cloned hERG Channels Expressed in Mammalian Cells. *Arch Drug Info* 2: 51–57.
277. Ridley J i in. 2003. Characterisation of recombinant HERG K<sup>+</sup> channel blockade by the Class Ia antiarrhythmic drug procainamide. *Biochem Biophys Res Commun* 306(2): 388-393.
278. Ridley J I in. 2003a. Inhibition of HERG K<sup>+</sup> current and prolongation of the guinea-pig ventricular action potential by 4-aminopyridine. *J Physiol* 549(3): 667-672.
279. Ridley J i in. 2004. Lidoflazine is a high affinity blocker of the HERG K<sup>+</sup> channel. *J Mol Cell Cardiol* 36(5): 701-705.
280. Ridley J i in. 2006. Inhibition of the HERG K<sup>+</sup> channel by the antifungal drug ketoconazole depends on channel gating and involves the S6 residue F656. *FEBS Lett* 580(8): 1999-2005.
281. Ridley J, Milnes J, Hancox J, Witchel H. 2006a. Clemastine, a conventional antihistamine, is a high potency inhibitor of the HERG K<sup>+</sup> channel. *J Mol Cell Cardiol* 40(1): 107-118.

282. Ridley J, Milnes J, Witchel H, Hancox J. 2004a. High affinity HERG K<sup>+</sup> channel blockade by the antiarrhythmic agent dronedarone: Resistance to mutations of the S6 residues Y652 and F656. *Biochem Biophys Res Commun* 325(3): 883-891.
283. Roche O i in. 2002. A virtual screening method for prediction of the hERG potassium channel liability of compound libraries. *Chem Bio Chem* 3(5): 455-459.
284. Roden DM. 2004. Drug-induced prolongation of the QT interval. *N Engl J Med* 350: 1013–1022.
285. Rodriguez-Menchaca A i in. 2006. Block of hERG Channels by Berberine Mechanisms of Voltage- and State-Dependence Probed With Site-Directed Mutant Channels. *J Cardiovasc Pharmacol* 47(1): 21-29.
286. Rutkowska D, Piliński M, Rutkowski L. 1997. Sieci neuronowe, algorytmy genetyczne i systemy rozmyte. PWN, Warszawa-Łódź.
287. Saenen J i in. 2007. A single hERG mutation underlying a spectrum of acquired and congenital long QT syndrome phenotypes. *J Mol Cell Cardiol* 43(1): 63-72.
288. Sakaguchi T i in. 2008. Hydroxyzine, a First Generation H<sub>1</sub>-Receptor Antagonist, Inhibits Human Ether-a-go-go-Related Gene (HERG) Current and Causes Syncope in a Patient With the HERG Mutation. *J Pharmacol Sci* 108, 462 – 471 (2008)
289. Salata J i in. 1995. Cardiac electrophysiological actions of the histamine H<sub>1</sub>-receptor antagonists astemizole and terfenadine compared with chlorpheniramine and pyrillamine. *Circ Res* 76(1): 110-119.
290. Sánchez-Chapula J, Ferrer T, Navarro-Polanco R, Sanguinetti M. 2003. Voltage-dependent profile of human ether-a-go-go-related gene channel block is influenced by a single residue in the S6 transmembrane domain. *Mol Pharmacol* 63(5): 1051-1058.
291. Sánchez-Chapula J, Navarro-Polanco R, Sanguinetti M. 2004. Block of wild-type and inactivation-deficient human ether-a-go-go-related gene K<sup>+</sup> channels by halofantrine. *Naunyn-Schmiedebergs Arch Pharmacol* 370(6): 484-491.
292. Sánchez-Chapula JA i in. 2002. Molecular Determinants of Voltage-dependent Human Ether-a-Go-Go Related Gene (HERG) K<sup>+</sup> Channel Block. *J Biol Chem* 277(26, 28): 23587-23595.
293. Sanguinetti M, Jiang C, Curran M, Keating M. 1995. A mechanistic link between an inherited and an acquired cardiac arrhythmia: HERG encodes the I<sub>Kr</sub> potassium channel. *Cell* 81(2): 299-307.
294. Sanguinetti MC, Mitcheson JS. 2005. Predicting drug-hERG channel interactions that cause acquired long QT syndrome. *Trends Pharmacol Sci* 26: 119–124.
295. Sanguinetti MC, Tristani-Firouzi M. 2006. hERG channels and cardiac arrhythmia. *Nature* 440(7083): 463-469.
296. Sarazan RD i in. 2004. Absence of clinically important HERG channel blockade by three compounds that inhibit phosphodiesterase 5—sildenafil, tadalafil, and vardenafil. *Eur J Pharmacol* 502: 163– 167.
297. Scherer C i in. 2002. The antihistamine fexofenadine does not affect I<sub>Kr</sub> currents in a case report of drug-induced cardiac arrhythmia. *Br J Pharmacol* 137(6): 892-900.
298. Scherer D i in. 2008. Selective noradrenaline reuptake inhibitor atomoxetine directly blocks hERG currents. *Br J Pharmacol* 156: 226–236.

299. Scholz E i in. 2003. Drug binding to aromatic residues in the HERG channel pore cavity as possible explanation for acquired Long QT syndrome by antiparkinsonian drug budipine. *Naunyn-Schmiedebergs Arch Pharmacol* 368(5): 404-414.
300. Scholz E i in. 2005. Inhibition of cardiac HERG channels by grapefruit flavonoid naringenin: Implications for the influence of dietary compounds on cardiac repolarisation. *Naunyn-Schmiedebergs Arch Pharmacol* 371(6): 516-525.
301. Scholz E i in. 2006a. *In vitro* modulation of HERG channels by organochlorine solvent trichlormethane as potential explanation for proarrhythmic effects of chloroform. *Toxicol. Lett.* 165(2): 156-166.
302. Scholz E i in. 2007. Anticholinergic antiparkinson drug orphenadrine inhibits HERG channels: Block attenuation by mutations of the pore residues Y652 or F656. *Naunyn-Schmiedebergs Arch Pharmacol* 376(4): 275-284.
303. Schram G i in. 2004. Ranolazine: Ion-channel-blocking actions and *in vivo* electrophysiological effects. *Br J Pharmacol* 142(8): 1300-1308.
304. Scott CW i in. 2003. A mediumthroughput functional assay of KCNQ2 potassium channels using rubidium efflux and atomic absorption spectrometry. *Anal. Biochem.* 319 (2): 251–257.
305. Seierstad M, Agrafiotis D. 2006. A QSAR model of hERG binding using a large, diverse, and internally consistent training set. *Chem Biol Drug Des* 67(4): 284-296.
306. Seop Kim D i in. 2006. Electrophysiological safety of novel fluoroquinolone antibiotic agents gemifloxacin and balofloxacin. *Drug Chem Toxicol* 29(3): 303-312.
307. Shah RR. 2005. Drug-induced QT-interval prolongation – regulatory guidance and perspectives on hERG channel studies. W The hERG cardiac potassium channel: structure, function and long qt syndrome. Derek J i in. (wyd.), Chichester, Novartis Foundation, John Wiley & Sons.
308. Shah RR. 2006. Can pharmacogenetics help rescue drugs withdrawn from the market? *Pharmacogenomics* 7, 889–908.
309. Shankar R, Frapaise X, Brown B. 2006. LEAN drug development in R&D. *Drug Discov Development* 5: 57–60.
310. Shantsila E, Watson T, Lip GYH. 2007. Drug-induced QT-interval prolongation and proarrhythmic risk in the treatment of atrial arrhythmias. *Europace* 9: 37–44.
311. Sheiner LB, Steimer JL. 2000. Pharmacokinetic/pharmacodynamic modeling in drug development. *Annu Rev Pharmacol Toxicol* 40, 67–95.
312. Sheiner LB. 1997. Learning versus confirming in clinical drug development. *Clin Pharmacol Ther* 61: 275–291.
313. Shim W-S i in. 2010. An aqueous extract of *Poncirus fructus* activates the prokinetic activity of 5-HT receptor subtype 4 without hERG interaction. *J Ethnopharmacol* 132: 328–333.
314. Shimizu A i in. 2003. Effects of Dronedarone on HERG and KCNQ1/KCNE1 Channels. *Environ Med* 47: 48-50.
315. Shin W, Kim K, Kim E. 2006. Electrophysiological effects of brompheniramine on cardiac ion channels and action potential. *Pharmacol. Res.* 54(6): 414-420.
316. Shuba Y i in. 2001. Testosterone-mediated modulation of HERG blockade by proarrhythmic agents. *Biochem. Pharmacol.* 62(1): 41-49.

317. Siebrands C, Friederich P. 2007. Inhibition of HERG channels by the local anaesthetic articaine. *Eur. J. Anaesthesiol.* 24(2): 148-153.
318. Siebrands C, Friederich P. 2007a. Structural requirements of human ether-a-go-go-related gene channels for block by bupivacaine. *Anesthesiology* 106(3): 523-531.
319. Siebrands C, Schmitt N, Friederich P. 2005. Local anesthetic interaction with human ether-a-go-go-related gene (HERG) channels: Role of aromatic amino acids Y652 and F656. *Anesthesiology* 103(1): 102-112.
320. Simulation in Drug Development: Good Practices Draft Publication of the Center for Drug Development Science (CDDS) Draft version 1.0, July 23, 1999. Holford NHG i in. (wyd.). <http://bts.ucsf.edu/cdds/research/>
321. Singh B N. 1999. Overview of trends in the control of cardiac arrhythmia: past and future. *Am J Cardiol* 84: 3R-10R.
322. Slee DH i in. 2008. 2-Amino-N-pyrimidin-4-ylacetamides as A2A Receptor Antagonists: 2. Reduction of hERG Activity, Observed Species Selectivity, and Structure-Activity Relationships. *J Med Chem* 51: 1730-1739.
323. Snodin DJ. 2002. An EU perspective on the use of *in vitro* methods in regulatory pharmaceutical toxicology. *Toxicol Lett* 127: 161-168.
324. Snyders DJ, Chaudhary A. 1996. High affinity open channel block by dofetilide of HERG expressed in a human cell line. *Mol Pharmacol* 49: 949-955.
325. Stanat S i in. 2003. Characterization of the inhibitory effects of erythromycin and clarithromycin on the HERG potassium channel. *Mol. Cell. Biochem.* 254(1-2): 1-7.
326. Stansfeld PJ. i in. 2007. Drug Block of the hERG Potassium Channel: Insight From Modeling. *Protein Struct Funct Genet Bioinf* 68:568-580.
327. Stefanowski J. 2001. Algorytmy indukcji reguł decyzyjnych w odkrywaniu wiedzy. Rozprawa habilitacyjna. Wydawnictwo Politechniki Poznańskiej, Seria Rozprawy nr 361.
328. Stevens JL, Baker TK. 2009. The future of drug safety testing: expanding the view and narrowing the focus. *Drug Discov Today* 14(34): 162-167.
329. Su Z i in. 2004. Mesoridazine: an open-channel blocker of human ether-a-go-go-related gene K<sup>+</sup> channel. *J Mol Cell Cardiol* 36: 151-160.
330. Su Z i in. 2006. Block of hERG channel by ziprasidone: Biophysical properties and molecular determinants. *Biochem. Pharmacol.* 71(3): 278-286.
331. Suessbrich H i in. 1997. The inhibitory effect of the antipsychotic drug haloperidol on HERG potassium channels expressed in *Xenopus* oocytes. *Br J Pharmacol* 120(5): 968-974.
332. Suessbrich H, Waldegger S, Lang F, Busch AE. 1996. Blockade of HERG channels expressed in *Xenopus* oocytes by the histamine receptor antagonists terfenadine and astemizole. *FEBS Lett* 385(1-2): 77-80.
333. Sun H. 2006. An accurate and interpretable Bayesian classification model for prediction of hERG liability. *Chem Med Chem* 1(3): 315-322.
334. Syroka VJ, Leahy DE. 2008. Chemical Descriptors Library (CDL): A Generic, Open Source Software Library for Chemical Informatics. *J Chem Inf Model* 48(10): 1931-1942.
335. Tadeusiewicz R. 1993. Sieci neuronowe. Akademicka Oficyna Wydawnicza, Warszawa.

336. Taglialatela M i in. 1998. Molecular basis for the lack of HERG K<sup>+</sup> channel block-related cardiotoxicity by the H1 receptor blocker cetirizine compared with other second-generation antihistamines. *Mol Pharmacol* 54(1): 113-121.
337. Taglialatela M i in. 2000. Inhibition of HERG1 K<sup>+</sup> channels by the novel second-generation antihistamine mizolastine. *Br J Pharmacol* 131(6): 1081-1088.
338. Taglialatela M, Timmerman H, Annunziato L. 2000a. Cardiotoxic potential and CNS effects of first-generation antihistamines. *Trends Pharmacol. Sci.* 21(2): 52-56.
339. Takemasa H i in. 2008. Coexistence of hERG current block and disruption of protein trafficking in ketoconazole-induced long QT syndrome. *Br J Pharmacol* 153: 439-447.
340. Tang Q i in. 2008. The 5-HT<sub>2</sub> antagonist ketanserin is an open channel blocker of human cardiac ether-a`-go-go-related gene (hERG) potassium channels. *Br J Pharmacol* 155: 365-373.
341. Tang W i in. 2001. Development and evaluation of high throughput functional assay methods for hERG potassium channel. *J Biomol Screen* 6 (5): 325-331.
342. Tarantino P, Appleton N, Lansdell K. 2005. Effect of trazodone on hERG channel current and QT-interval. *Eur J Pharmacol* 510(1-2): 75-85.
343. Teitelbaum Z, Lave T, Freijer J, Cohen AF. 2010. Risk assessment in extrapolation of pharmacokinetics from preclinical data to humans. *Clin Pharmacokinet* 49: 619-32.
344. Teschemacher A, Seward E, Hancox J, Witchel H. 1999. Inhibition of the current of heterologously expressed HERG potassium channels by imipramine and amitriptyline. *Br J Pharmacol* 128(2): 479-485.
345. Thai K-M, Ecker GF. 2009. Similarity-based SIBAR descriptors for classification of chemically diverse hERG blockers. *Mol Divers*, DOI 10.1007/s11030-009-9117-0.
346. Thai KM, Ecker GF. 2008. A binary QSAR model for classification of hERG potassium channel blockers. *Bioorg Med Chem* 16: 4107-4119.
347. Thai K-M, Ecker GF. 2008. Classification Models for hERG Inhibitors by Counter-Propagation Neural Networks. *Chem Biol Drug Des* 72: 279-289.
348. Thomas D i in. 2001. High-affinity blockade of human ether-a-go-go-related gene human cardiac potassium channels by the novel antiarrhythmic drug BRL-32872. *J Pharmacol Exp Ther* 297(2): 753-761.
349. Thomas D i in. 2003. Inhibition of cloned HERG potassium channels by the antiestrogen tamoxifen. *Naunyn-Schmiedeberg's Arch Pharmacol* 368(1): 41-48.
350. Thomas D i in. 2003a. Acute effects of dronedarone on both components of the cardiac delayed rectifier K<sup>+</sup> current, HERG and KvLQT1/minK potassium channels. *Br J Pharmacol* 140(5): 996-1002.
351. Thomas D i in. 2003b. The antipsychotic drug chlorpromazine inhibits HERG potassium channels. *Br J Pharmacol* 139(3): 567-574.
352. Thomas D i in. 2004. Direct block of hERG potassium channels by the protein kinase C inhibitor bisindolylmaleimide I (GF109203X). *Cardiovasc Res* 64(3): 467-476.
353. Thomas D i in. 2004a. Inhibition of cardiac HERG currents by the DNA topoisomerase II inhibitor amsacrine: Mode of action. *Br J Pharmacol* 142(3): 485-494.



354. Thomas D i in. 2008. Doxazosin induces apoptosis of cells expressing hERG K<sup>+</sup> channels. *Eur J Pharmacol* 579(1-3): 98-103.
355. Thomas D i in. 2004b. Inhibition of human ether-a-go-go-related gene potassium channels by  $\alpha$ 1-adrenoceptor antagonists prazosin, doxazosin, and terazosin. *Naunyn-Schmiedeberg Arch Pharmacol* 369(5): 462-472.
356. Thomsen M i in. 2003. Electrophysiological Safety of Sertindole in Dogs with Normal and Remodeled Hearts. *J Pharmacol Exp Ther* 307(2): 776-784.
357. Tian M i in. 2006. Effects of the antifungal antibiotic clotrimazole on human cardiac repolarization potassium currents. *Br J Pharmacol* 147(3): 289-297.
358. Tie H. 2002. Cellular mechanisms of QT prolongation and proarrhythmia induced by non-antiarrhythmic drugs. University of New South Wales.
359. Tobita M, Nishikawa T, Nagashima R. 2005. A discriminant model constructed by the support vector machine method for HERG potassium channel inhibitors. *Bioorg Med Chem Lett* 15(11): 2886-2890.
360. Toga T, Kohmura Y, Kawatsu R. 2007. The 5-HT<sub>4</sub> Agonists Cisapride, Mosapride, and CJ-033466, a Novel Potent Compound, Exhibit Different Human Ether-a-go-go-Related Gene (hERG)-Blocking Activities. *J Pharmacol Sci* 105: 207 – 210.
361. Tollman P i in. 2001. A revolution in R&D. How genomics and genetics are transforming the biopharmaceutical industry. Boston Consulting Group report ([www.forskningsradet.no/CSStorage/Flex\\_attachment/BiotekBCGGenomics.pdf](http://www.forskningsradet.no/CSStorage/Flex_attachment/BiotekBCGGenomics.pdf))
362. Traebert M i in. 2004. Inhibition of hERG K<sup>+</sup> currents by antimalarial drugs in stably transfected HEK293 cells. *Eur J Pharmacol* 484(1): 41-48.
363. Trepakova E, Dech S, Salata J. 2006. Flunarizine is a highly potent inhibitor of cardiac hERG potassium current. *J Cardiovasc Pharmacol* 47(2): 211-220.
364. Tsujimae K, Suzuki S, Yamada M, Kurachi Y. 2004. Comparison of kinetic properties of quinidine and dofetilide block of HERG channels. *Eur J Pharmacol* 493(1-3): 29-40.
365. Tu D i in. 2008. Electropharmacological properties of telmisartan in blocking hKv1.5 and HERG potassium channels expressed on *Xenopus laevis* oocytes. *Acta Pharmacol Sin* 29(8): 913-922.
366. Ulens C, Daenens P, Tytgat J. 1999. Norpropoxyphene-induced cardiotoxicity is associated with changes in ion-selectivity and gating of HERG currents. *Cardiovasc Res* 44(3): 568-578.
367. Valerio LG. 2008. Tools for evidence-based toxicology: computational-based strategies as a viable modality for decision support in chemical safety evaluation and risk assessment. *Hum Exp Toxicol* 27: 757-760.
368. Valerio LG. 2009. *In silico* toxicology for the pharmaceutical sciences, *Toxicol. Appl. Pharmacol.* doi:10.1016/j.taap.2009.08.022
369. Van de Waterbeemd H, Gifford E. 2003. ADMET *in silico* modeling: towards prediction paradise? *Nat Rev Drug Discov* 2, 192-204.
370. Viskin S. 1999. Long QT syndromes and torsade de pointes. *Lancet* 354(9190): 1625-1633.
371. Volberg W i in. 2002. Blockade of human cardiac potassium channel Human Ether-a-go-go-Related Gene (HERG) by macrolide antibiotics. *J Pharmacol Exp Ther* 302(1): 320-327.

372. Volders PGA i in. 2000. Progress in the understanding of cardiac early afterdepolarizations and torsades de pointes: time to revise current concepts. *Cardiovasc Res* 46: 376–392.
373. Vormberge T, Hoffmann M, Himmel H. 2006. Safety pharmacology assessment of drug-induced QT-prolongation in dogs with reduced repolarization reserve. *J Pharmacol Toxicol Meth* 54(2): 130-140.
374. Walker B i in. 1999. Inhibition of the human ether-a-go-go-related gene (HERG) potassium channel by cisapride: Affinity for open and inactivated states. *Br J Pharmacol* 128(2): 444-450.
375. Walker B i in. 1999a. Inhibition of HERG channels stably expressed in a mammalian cell line by the antianginal agent perhexiline maleate. *Br J Pharmacol* 127(1): 243-251.
376. Walker B i in. 2000. Comparative effects of azimilide and ambasilide on the human ether-a-go-go-related gene (HERG) potassium channel. *Cardiovasc Res* 48(1): 44-58.
377. Wang H, Shi H, Liao S, Wang Z. 1999. Inactivation gating determines nicotine blockade of human HERG channels. *Am. J Physiol Heart Circ. Physiol.* 277: H1081-H1088.
378. Wang J i in. 2003. Functional and pharmacological properties of canine ERG potassium channels. *Am. J Physiol Heart Circ. Physiol.* 284(1): H256-H267.
379. Wang L, Wible B, Wan X, Ficker E. 2007. Cardiac glycosides as novel inhibitors of human Ether-a-gogo-related gene channel trafficking. *J Pharmacol Exp Ther* 320(2): 525-534.
380. Weerapura M i in. 2002. A comparison of currents carried by HERG, with and without coexpression of MiRP1, and the native rapid delayed rectifier current. Is MiRP1 the missing link? *J Physiol* 540(1): 15-27.
381. Wilke RA i in. 2007. Identifying genetic risk factors for serious adverse drug reactions: current progress and challenges. *Nature Rev. Drug Discov* 6: 904-916.
382. Willett P. 2006. Similarity-based virtual screening using 2D-fingerprints. *Drug Discov Today* 11(23/24): 1046-1053.
383. Willett P i in. 1986. Implementation of nearest-neighbor searching in an online chemical structure search system. *J Chem Inf Comput Sci* 26: 36–41.
384. Wiśniowska B, Polak S. 2009. hERG *in vitro* interchange factors - development and verification. *Tox Mech Meth* 19(4): 278-284.
385. Witchel H i in. 2002. Troubleshooting problems with *in vitro* screening of drugs for QT interval prolongation using HERG K<sup>+</sup> channels expressed in mammalian cell lines and *Xenopus* oocytes. *J Pharmacol Toxicol Meth* 48(2): 65-80.
386. Witchel H i in. 2002a. Inhibitory actions of the selective serotonin re-uptake inhibitor citalopram on HERG and ventricular L-type calcium currents. *FEBS Lett* 512(1-3): 59-66.
387. Wood C, Williams C, Waldron GJ. 2004. Patch clamping by numbers. *Drug Discov Today* 9: 434-441.
388. Woosley RL i in. 1993. Mechanism of the cardiotoxic actions terfenadine. *JAMA* 269: 1532-1536.
389. Woosley RL. 1996. Cardiac actions of antihistamines. *Annu Rev Pharmacol Toxicol* 36: 233–252.
390. Written IH, Frank E. 2005. *Data Mining: Practical machine learning tools and techniques*, 2nd Edition, Morgan Kaufmann, San Francisco.

391. Wu L i in. 2003. Effects of Na<sup>+</sup> channel blocker, pilsicainide, on HERG current expressed in HEK-293 cells. *J Cardiovasc Pharmacol* 42(3): 410-418.
392. Wu L i in. 2008. Augmentation of late sodium current unmasks the proarrhythmic effects of amiodarone. *Cardiovasc Res* 77: 481–488.
393. Wysowski DK, Bacsanyi J. 1996. Cisapride and fatal arrhythmia. *N Engl J Med* 35: 290-291.
394. Wysowski DK, Swartz L. 2005. Adverse Drug Event Surveillance and Drug Withdrawals in the United States, 1969-2002. The Importance of Reporting Suspected Reactions. *MBA Arch Intern Med* 165: 1363-1369.
395. Xing J i in. 2010. Block effect of capsaicin on hERG potassium currents is enhanced by S6 mutation at Y652. *Eur J Pharmacol* 630: 1–9.
396. Yamada M i in. 2006. The Effects of Sevoflurane and Propofol on QT Interval and Heterologously Expressed Human Ether-A-Go-Go Related Gene Currents in *Xenopus* Oocytes. *Anesth Analg* 102: 98–103.
397. Yao X i in. 2008. Predicting QT prolongation in humans during early drug development using hERG inhibition and an anaesthetized guinea-pig model. *Br J Pharmacol* 154: 1446–1456.
398. Yao X i in. 2005. Propranolol inhibits the human ether-a-go-go-related gene potassium channels. *Eur J Pharmacol* 519(3): 208-211.
399. Yap CW. i in. 2004. Prediction of Torsade-Causing Potential of Drugs by Support Vector Machine Approach *Toxicol Sci* 79: 170–177.
400. Yap Y, Camm A. 1999. The current cardiac safety situation with antihistamines. *Clin. Exp. Allergy Suppl.* 29(1): 15-24.
401. Yoshida K, Niwa T. 2006. Quantitative Structure-Activity Relationship Studies on Inhibition of HERG Potassium Channels. *J Chem Inf Model* 46: 1371-1378.
402. Young T. 2005. An Agenda for Healthcare and Information Simulation. *Health Care Manag Sci* 8: 189-196.
403. Yuill K i in. 2004. Potent inhibition of human cardiac potassium (HERG) channels by the anti-estrogen agent clomiphene - Without QT interval prolongation. *Biochem Biophys Res Commun* 318(2): 556-561.
404. Zhang G-H. 2009. Effect of dipfluzine on hERG potassium currents expressed in *xenopus* oocytes. *Chin Pharmacol Bull* 2009-01.
405. Zhang Q, Muegge I. 2006. Scaffold hopping through virtual screening using 2D and 3D similarity descriptors: ranking, voting, and consensus scoring. *J Med Chem* 49: 1536–1548.
406. Zhang S i in. 1999. Mechanism of block and identification of the verapamil binding domain to HERG potassium channels. *Circ Res* 84(9): 989–998.
407. Zhang S i in. 2001. Cocaine blocks HERG, but not KvLQT1+minK, potassium channels. *Mol Pharmacol* 59(5): 1069-1076.
408. Zhang X-H i in. 2010. The calmodulin inhibitor N-(6-aminohexyl)-5-chloro-1-naphthalene sulphonamide directly blocks human ether à-go-go-related gene potassium channels stably expressed in human embryonic kidney 293 cells. *Br J Pharmacol* 161: 872–884.
409. Zhao XL i in. 2009. Comparative effects of sophocarpine and sophoridine on hERG K<sup>+</sup> channel. *Eur J Pharmacol* 607: 15–22.

410. Zhou Z i in. 1998. Properties of HERG channels stably expressed in HEK 293 cells studied at physiological temperature. *Biophys J* 74(1): 230-241.
411. Zhou Z i in. 1999. Block of HERG potassium channels by the antihistamine astemizole and its metabolites desmethylastemizole and norastemizole. *J Cardiovasc Electrophysiol* 10(6): 836-843.
412. Zitron E i in. 2002. Bertosamil blocks HERG potassium channels in their open and inactivated states. *Br J Pharmacol* 137(2): 221-228.
413. Zitron E i in. 2004. Inhibition of cardiac HERG potassium channels by the atypical antidepressant trazodone. *Naunyn Schmiedebergs Arch Pharmacol* 370(2): 146-156.
414. Zitron E i in. 2005. QTc prolongation by grapefruit juice and its potential pharmacological basis: HERG channel blockade by flavonoids. *Circulation* 111(7): 835-838.
415. Zünkler B i in. 2006. Effects of fluoroquinolones on HERG channels and on pancreatic beta-cell ATP-sensitive K<sup>+</sup> channels. *Toxicology* 228(2-3): 239-248.
416. Zünkler BJ, Wos-Maganga M. 2010. Comparison of the Effects of Methadone and Heroin on Human ether-a-go-go-Related Gene Channels. *Cardiovasc Toxicol* 10: 161–165.
417. Żurada JM. 1992. *Introduction to Artificial Neural Systems*, West Publishing Company, ISBN-10: 053495460X, USA.