

**Zgłoszenie zagadnienia badawczego realizowanego
w Krakowskiej Interdyscyplinarnej Szkole Doktorskiej
w dyscyplinie: nauki farmaceutyczne**

w Jednostce: Instytut Farmakologii im. J. Maja Polskiej Akademii Nauk

1	Nazwisko i imię promotora, tytuł/stopień naukowy, jednostka, adres e-mail	Dr hab. Rafał Kurczab, prof. IF PAN Instytut Farmakologii im. Jerzego Maja Polskiej Akademii Nauk kurczab@if-pan.krakow.pl
2	Nazwisko i imię promotora pomocniczego (opcjonalnie), jednostka, adres e-mail	Dr Wojciech Pietruś, Instytut Farmakologii im. Jerzego Maja Polskiej Akademii Nauk pietrus@if-pan.krakow.pl
3	Temat zagadnienia badawczego + krótki (do 250 słów) opis tematyki badawczej	Nowa reprezentacja molekularna oparta na strukturze 2D i widmach oscylacyjnych (FT-IR/RS) związków chemicznych: innowacyjne narzędzia in silico dla projektowania nowych leków. Celem projektu doktorskiego jest opracowanie nowej hybrydowej reprezentacji molekularnej łączącej informacje strukturalne 2D z sygnaturami drgań pochodzącymi z widm FT-IR i Ramanowskich. Klasyczne reprezentacje (fingerprinty 2D/3D, deskryptory, GNN) osiągają granice predykcji, gdyż nie uwzględniają lokalnych właściwości wiązań, polaryzowalności ani zmian konformacyjnych. Widma oscylacyjne zawierają unikalne informacje o tych cechach, co czyni je obiecującym źródłem danych do wzbogacenia modeli AI stosowanych w chemoinformatyce i projektowaniu leków. Projekt obejmuje cztery komplementarne etapy. W pierwszym zostanie zoptymalizowana metodologia pozyskiwania widm FT-IR/RS, obejmująca standaryzację próbek, wstępne przetwarzanie i redukcję szumu. W drugim etapie przeprowadzona zostanie systematyczna ewaluacja dostępnych generatorów widm (m.in. Chemprop-IR, Graphormer-IR, Mol2Raman), aby określić ich przydatność w modelowaniu właściwości związków oraz zidentyfikować ograniczenia wymagające ulepszeń. Trzeci etap obejmuje konstrukcję słownika istotnych cech widmowych i integrację ich ze strukturą 2D w nową, spójną reprezentację numeryczną. W etapie czwartym zostaną opracowane i

		przetestowane modele uczenia maszynowego oraz głębokiego (np. transformery 1D, autoenkodery wariacyjne) do analizy, ekstrakcji i symulacji widm na podstawie struktury cząsteczki. Skuteczność opracowanej reprezentacji zostanie oceniona w predykcji właściwości fizykochemicznych (logP, logD) oraz wybranych aktywności biologicznych (np. wybrane receptory serotoninowe, dopaminowe i/lub opioidowe), w porównaniu do konwencjonalnych fingerprintów i nowoczesnych modeli GNN. Należy zaznaczyć, iż aktualne badania Promotora w zakresie użycia danych pochodzących z widm ¹H i ¹³C NMR (rezonansu jądrowego) pokazują istotną poprawę predykcji parametru logD (10.1021/acs.jcim.3c02039; 10.1021/acs.jcim.5c01791; 10.1021/acs.jcim.4c02145), a pierwsze proste transformacje widm FT-IR okazały się istotne w predykcji parametru logP (10.55225/sti.492, 10.1021/acs.jcim.5c01574).
4	Wymagania w stosunku do kandydata (wykształcenie, umiejętności/kursy)	Dyplom ukończenia studiów wyższych w dziedzinie chemii lub bioinformatyki. Znajomość języka angielskiego na poziomie umożliwiającym swobodną komunikację tak w mowie jak i w piśmie, znajomość spektroskopii molekularnej, metod modelowania molekularnego, chemii kwantowej, chemo/bioinformatycznych oraz podstaw uczenia maszynowego. Dodatkowym atutem będzie umiejętność programowania w języku R i/lub Python
5	Wskazanie możliwych źródeł i zakresu finansowania spoza subwencji, np. stypendium naukowego, kosztów badań, wyjazdów itp.	działalność statutowa Zakładu Chemii Leków IF PAN, grant PRELUDIUM (będzie złożony przez doktoranta w 2026 r.),

1	Supervisor: name/surname, degree, affiliation, e-mail address	Dr hab. Rafał Kurczab, prof. IF PAN Maj Institute of Pharmacology, Polish Academy of Sciences kurczab@if-pan.krakow.pl
2	Auxiliary supervisor (optional) affiliation, e-mail address	Dr Wojciech Pietruś, Maj Institute of Pharmacology Polish Academy of Sciences

		pietrus@if-pan.krakow.pl
3	Research subject title Short description, up to 250 words	Novel molecular representation based on 2D structure and vibrational spectra (FT-IR/RS) of chemical compounds: innovative in silico tools for new drug design. The aim of the PhD project is to develop a new hybrid molecular representation that combines 2D structural information with vibrational signatures derived from FT-IR and Raman spectra. Classical representations (2D/3D fingerprints, physicochemical descriptors, GNN-based encodings) have reached predictive limits because they do not capture local bond environments, polarizability effects, or conformational dynamics. Vibrational spectra contain unique information related to these properties, making them a promising data source for enriching AI models used in cheminformatics and drug design. The project consists of four complementary stages. In the first stage, the methodology for acquiring FT-IR and Raman spectra will be optimized, including sample standardization, preprocessing, and noise reduction. The second stage involves a systematic evaluation of state-of-the-art spectral generators (e.g., Chemprop-IR, Graphormer-IR, Mol2Raman) to assess their applicability in modeling molecular properties and to identify limitations requiring methodological improvements. The third stage focuses on constructing a dictionary of informative spectral features and integrating them with 2D structural data into a unified numerical representation. In the fourth stage, machine learning and deep learning models (e.g., 1D transformers, variational autoencoders) will be developed and tested for spectral analysis, feature extraction, and spectrum simulation based on molecular structure. The effectiveness of the developed representation will be evaluated in predicting physicochemical properties (e.g. logP, logD) and selected biological activities (e.g. selected serotonin, dopamine and/or opioid receptors), and compared with conventional fingerprints and modern GNN-based approaches. Preliminary results from the research team using spectroscopic data (NMR, FT-IR) demonstrate substantial improvements in predictive performance, supporting the relevance of the proposed research direction and its application potential in drug discovery. It should be noted that Supervisor research on the use

		of data derived from ^1H and ^{13}C NMR spectra (nuclear resonance) showed a significant improvement in the prediction of the logD parameter (10.1021/acs.jcim.3c02039; 10.1021/acs.jcim.5c01791; 10.1021/acs.jcim.4c02145), and the simple transformations of FT-IR spectra proved significant in the prediction of the logP parameter (10.55225/sti.492, 10.1021/acs.jcim.5c01574).
4	Additional requirements to the candidate (education, skills / courses)	M.Sc. in chemistry or bioinformatics. English language at a level enabling fluent communication both in speech and writing, knowledge of molecular spectroscopy, molecular modeling methods, quantum chemistry, chemo/bioinformatics, and the basics of machine learning. An additional advantage will be basic programming skills in R and/or Python.
5	Possible sources of financing, other than subsidy, e.g., scientific scholarship, research and travel costs, etc.	Statutory funds of the Department of Medicinal Chemistry, Maj Institute of Pharmacology, Polish Academy of Sciences, PRELUDIUM grant (will be submitted by a PhD student in 2026),