ARTIQ

ARTIQ - AI Centres of Excellence

Application for a Host Institution

Institution	National Centre for Research and Development, National Science Centre
Project Joint National Project:	ARTIQ – AI Centres of Excellence
Deadline for the submission of applications	8th of April-11th of May 2021

I. HOST INSTITUTION DATA

Identification data of the Host Institution

Name (full)	***Hirszfeld Institute of Immunology and Experimental Therpy Polish Academy of Sceince***
Name (short)	***IITD PAN***
Name of the main organisational unit (where applicable)	
Address of the registered office	
Street	Ul. Rudolfa Weigla
Building No.	*12*
Office No.	
Postal code	*53-114*
City/district	*Wrocław*
Post office	*Wrocław*
Municipality	*Wrocław*
County	*Poland*

Province	*dolnoslaskie*
Correspondence address (if different than the address of the registered office)	
Street	
Building No.	
Office No.	
Postal code	
City/district	
Post office	
Municipality	
County	
Province	
EPUAP [Electronic Platform for Public Administration Services] mailbox	*/PAN/SkrytkaESP*
Legal form	*Basic (podstawowa): 1* *Specific (szczególna): 428*
The person appointed for contact with NCBR and with the potential Leader/Project Manager	
First name	*Lukasz*
Last name	*Łaczmański*
Position	*associate professor, Head of the Laboratory of Genomics & Bioinformatics*
Phone number	*+48 605 314 478*
E-mail address	*Lukasz.laczmanski@hirszfeld.pl*
The person authorised to represent the applicant	
First name	*Andrzej*
Last name	*Gamian*
Function/Position	*Head of the Institute*
First name	*Jolanta*
Last name	*Łukasiewicz*
Function/Position	*V-ce head of the Institute*

II. CAPACITY OF THE HOST INSTITUTION TO PERFORM THE PROJECT

1. Description of major research achievements in the scope of implementation of R&D projects, as well as the commercialisation of deliverables of such projects regarding artificial intelligence for the last 5 years prior to or in the year of the application along with a list of the most important publications and patents of the applicant.

At IITD PAN, many research projects are carried out in which machine learning algorithms are used to analyze biological data such as genomic sequence data, transcriptomic data, image data or other data characterizing biological molecules, cells or whole organisms. A research project, "Molecular characterization of cancer cells - analysis of transcriptomes and epigenetic data", is carried out at the GiB Laboratory. The main goal is to use data from high-throughput sequencing to develop a molecular model of cancer cells. For this purpose, we use supervised machine learning algorithms.

The Laboratory of Instrumental Analysis and Preparation, basing on the data obtained from the JEOL F200 transmission microscope and the Cryo-EM instrumentation, implements the project of analyzing protein structures by crayomicroscopy. In this project, structures are generated from image data using software based on machine learning algorithms.

The Laboratory of Microbial Immunochemistry and Vaccines uses AI procedures to analyze data describing pathogen glycan antigens. Data is obtained from the NMR 600 spectroscope. The IITD also obtains other biological data from the devices mentioned in point II 3.

The main publications:

1) Konopka B.M., Lwow F., Owczarz M., Łaczmański Ł. PLOS ONE. Doi: 10.1371/journal.pone.0201950
2) Łaczmański Ł., Łaczmańska I., Lwow F. Sci Rep. 2019 Nov 5;9(1):16026
3) Rami A, Łaczmański Ł, Jacków-Nowicka J, Jacków J. Int J Mol Sci. 2020 Dec 29;22(1):245. doi: 10.3390/ijms22010245. PMID: 33383666; PMCID: PMC7795642.
4) Stembalska A., Klapecki J., Pławski A., Karpinski P.: Eur J Obstet Gynecol Reprod Biol, 2019, 238:54-57,
5) Laczmanska I., Sasiadek M., Laczmanski L.: Anticancer Res., 2018, 38(3):14151418
6) Bińkowska A., Biernat M.M., Łaczmański Ł., Gościniak G.: 2018, 9:3154,
7) Wanecka A., Król J., Twardoń J., Mrowiec J., Korzeniowska-Kowal A., Wzorek A.: J Vet Diagn Invest, 2019:1040638719845423
8) Świętnicki W., Brzozowska E.: J Mol Graph Model, 2019, 92:8-16,

The Institute of Immunology and Experimental Therapy of the Polish Academy of Science also carries out industrial research completed with implementations. Below we present a patent application related to the topic:

1. Zgłoszenie P.404229 - patent przyznany dnia 09.09.2017 (brak numeru prawa wyłącznego) Goździewicz Tomasz, Łukasiewicz Jolanta, Ługowski Czesław: Wyizolowany immunogenny antygen bakteryjny, kompozycja farmaceutyczna, szczepionka, glikokoniugat zawierające antygen, przeciwciało wiążące się z antygenem, zastosowanie antygenu w profilaktyce i leczeniu zakażeń wywoływanych przez Gram-ujemne bakterie oraz sposób otrzymywania antygenu. Właściciel patentu: Wrocławskie Centrum Badań EIT+ Sp. z o.o.

2. A list of 5 research and development projects within national and international competitions in the area of artificial intelligence and implemented within the last 5 years prior to or in the year of the application (title, manager, source of financing, amount of financing) (max. 1 A4 page).

1. Database for Scientific Information Supporting Innovative Therapies - BINWIT. Project manager: dr Krzysztof Pawlik. Granted budget: 18 867 895.65 PLN. Funded under: Operational program, Digital Poland. Priority axis: II e-administration and open government. Action: 2.3 Digital accessibility and usability of public information. Secondary action: 2.3.1. Digital availability of public information from administrative and scientific sources. Contest number: POPC.02.03.01-IP.01-00-006/17. The project covers digitization, database construction and digital access to selected scientific resources of the Institute of Immunology and Experimental Therapy (IITD). The project digitizes the genetic data of bacteriophages and their hosts, imaging data of mesenchymal stem cells and other biological data. Genetic data of bacteriophages are analyzed with the use of specially prepared algorithms using machine learning.
2. "Implementation of the immunogenetic database of the Polish population MultiGenBank". Project manager: mgr Dariusz Wójcik Funded under POIG 2007-2013. The project included the creation of a genetic database with analytical tools. Granted budget 8 661 084 PLN.
3. Implementation of an interoperable electronic scientific platform of the Polish Collection of Microorganisms. Project manager: dr inż. Agnieszka Korzeniowska-Kowal. Projects related to the development of information resources Project no. POIG.02.03.02-00-005 / 10. Granted budget 2,480,000 PLN.

The total amount of investments in the above-mentioned projects is PLN 30,008,979.56. The results of the projects constantly contribute to the scientific development of IITD.

3. Available research equipment, apparatus/infrastructure and intangible assets held in the context of implementation of a project regarding artificial intelligence (max. 1 A4 page).

IITD PAN is equipped with both a computing center and devices that enable the acquisition of data describing biological processes and specialized units supporting the acquisition of material for research.

Data Center (computing center) equipped with modern solutions based on technologies from IBM, Dell and HP:

1. Centralized storage in the form of the V5010 disk array based on Fiber Channel technology;
2. The Institute has a GPU / CUDA unit for calculations and structural modeling equipped with two PNY P5000 graphics cards, each of them has the computing power of 8.9 TFLOPS
3. The Institute has a WiFi network consisting of WIPS, 48 AP, providing coverage of the main building and managed by the controller;
4. The network is secured by modern UTM hardware devices, by the ESET central anti-virus protection system, the existence of a backup server room located in a building other than the main server room and daily backups.

Devices for acquiring biological data (only those whose value at the time of purchase exceeded PLN 500,000 were listed):

High-throughput DNA sequencers: NextSEQ 500/550 (Illumina), MiSEQ (Illumina), MinION (ONT) - nanopore sequencing (long reads);

Electron microscopes: Scanning electron microscope AURIGA 60 Zeiss, Transmission electron microscope JEOL JEM-F-200, with equipment for Cryo-EM procedures.,

Other imaging data acquisition devices: 3D HT-2 holotomographic microscope, Fluorescence visualization chamber, Digital system for vital imaging with Doppler system, Multifunctional fluorescence scanner TYPHOON FLA,

NMR spectrometer and MAS spectrometers: NMR spectrometer 600 with HR-MAS probe, MS / MS U-HPLC mass spectrometer, TOF-TOF mass spectrometer with MALDI ion source, Amazon SL spectrometer with HPLC.

Other specialist devices: BIACORE T200 device for measuring intermolecular interactions, BECTON DICKINSON multi-parameter flow cytometer.

Other research support units:

Polish Collection of Microorganisms - has a rich collection of bacterial strains and bacteriophages. It is registered with the World Federation of Microbial Collections (WFCC, No. 106) and the European Organization of Microbial Collections (ECCO). The purpose of the collection is to provide reference bacterial strains to recipients from all over Poland, as well as from abroad, for research, comparative and educational purposes.

The IITD PAN Medical Center is a unit established, supervised and run by IITD PAN in Wrocław. It includes the Phage Therapy Center, the Tissue Immunology Laboratory and the SARS-CoV-2 Infection Diagnostics Laboratory. Cooperation with the medical center enables obtaining samples from patients, which can be a source of data in the conducted research.

Animal House - enables the acquisition of data in experiments conducted on animals.

4. Facilities or incentives to establish an AI Centre of Excellence in the entity.

IITD PAN has been investing in IT infrastructure for several years, expanding its computing center. As part of these efforts, the Genomics and Bioinformatics Laboratory was established in 2019, whose task is to develop bioinformatic analysis solutions based on artificial intelligence (AI) and machine learning (ML) algorithms. The Institute has equipment for omics analyzes (genetic material sequencers, electron microscopes, NMR, mas spectrometers), which is used to obtain data necessary for multivariate analyzes. The possibilities of obtaining biological data are supported by specialized laboratories, e.g. microbiological class II, GMM and GMO laboratories. Additional support is provided by such units as the Polish Collection of Microorganisms, IITD PAN Medical Center and the animal house facility.

In 2004, IITD PAN obtained the status of the IMMUNE Center of Excellence. Since then, IITD's research capabilities have grown steadily.

Currently, the Institute is planning a project on the use of artificial intelligence algorithms to fully characterize pathogens. It is planned to try to analyze omics data. By using high-throughput sequencers, it is possible to obtain a complete genomic and transcriptome map. Using MALDI TOF, we obtain data related to proteomics. The crayomicroscopy and NMR techniques enable the analysis of the protein or/and glycan structures.

Thanks to high-throughput sequencing, we can obtain the complete sequence of the genome of a given pathogen. Using the classification algorithms, we can determine the degree of similarity between the sequences of the genetic code of the tested microorganism and the reference sequences of genes. Thanks to such analysis, we can characterize the pathogen in terms of the presence of genes encoding various types of metabolites. Thanks to this information, it is possible to molecularly characterize the pathogen if it contains substances that determine antibiotic resistance or toxins. It should be emphasized that thanks to the use of clustering algorithms and cluster analysis, we are able to identify sequences with previously unknown properties.

NMR spectroscopy provides basic structural data on the structure of biological macromolecules at the atomic level. This data is quantitative, is obtained in a repeatable and non-destructive manner, and is rich in information on the complexity of the structure of the analyzed molecules, obtained as large data sets. NMR spectra represent unique profiles ("fingerprints") that allow the comparison, differentiation and classification of bacterial strains based on differences in glycan structures. The data obtained from one- and multivariate NMR experiments allow the use of mathematical methods of pattern recognition and multivariate analysis (PCA) for this purpose. These methods largely overlap with the techniques currently being developed using neural networks. Extending the scope of these techniques with analyzes using AI (machine learning, deep learning) algorithms in the processing of NMR data creates a starting point for the development of methods and automated procedures for verification and prediction of structures in correlation with real and simulated NMR data sets. In the case of bacterial glycans, AI techniques can be used in (1) the classification (chemotyping) and automatic recognition of the structural elements of bacterial glycans and (2) the reconstruction of NMR spectra based on a limited number of data (e.g. obtained by NUS technique, non-uniform sampling) by increasing efficiency of measurements and optimizing the time of use of NMR spectroscopes.

Thanks to the use of artificial intelligence algorithms, we can combine the analytical areas described above and create a general molecular model of pathogens.

5. Other information concerning internationalisation of the entity, foreign scientists employed in this institution, availability of seminars in English, etc. (max. 1 A4 page).

The Institute's international advisory board:

- Jean Francois Bach, Secretaire Perpetuel, French Academy of Sciences, Department of Immunology, Hopital Necker, Paris, France
- Michael Goldman, Department of Immunology, Hopital Erasme, Universite Libre de Bruxelles, Belgium
- J. Kupiec-Weglinski, Department of Surgery, David Geffen School of Medicine at UCLA, Los Angeles, USA
- Gerhard Opelz , Institute of Immunology, University of Heidelberg, Germany
- Emil Skamene, The Research Institute, McGill University Health Centre, Montreal, Quebec, Canada
- Hans-Dieter Volk, Institute of Medical Immunology, Medical University of Berlin, Charite Hospital, Berlin, Germany
- Claudine Kieda, Centre National de la Recherche Scientifique, Orleans, France

There are three Scientific Societies in IITD: the Polish Genetic Society, the Polish Immunogenetic Society and the Polish Society of Experimental and Clinical Immunology, which regularly organize meetings with invited guests from various, also international, research units. Weekly, Thursday's IITD scientific seminars are held, also conducted in English and with the participation of international guests.

IITD PAN runs the International and Interdisciplinary Doctoral School (Wrocław Doctoral School of Institutes of the Polish Academy of Sciences), where education is conducted in English.

It should also be emphasized that the Institute is the publisher of two international peer-reviewed journals (from the Philadelphia list): Archivum Immunologiae et Therapiae Experimentalis (IF = 3.2) and Advances in Hygiene and Experimental Medicine (IF = 0.8).

IITD PAN supports international cooperation, promotes the mobility of employees and doctoral students and creates opportunities for them to participate in projects in cooperation with foreign partners by participating in programs such as Erasmus Plus.

6. Other significant information confirming the experience and resources of the institution (max. 1 A4 page).

IITD PAN has the "A" category. The Institute employs 247 people, including 19 professors, 23 associate professors, 43 assistant professors, 7 assistants, 21 research and technical employees, 57 specialists, 22 lab technicians, 34 administration and documentation employees and others.

The facilities of large research equipment are supported by specialized personnel in several apparatus laboratories: NMR Spectroscopy Laboratory, Inter-Departmental Laboratory of Cytometry and Confocal Microscopy, Inter-Departmental Laboratory of Instrumental Analysis and Preparation, Polish Collection of Microorganisms, whose task is to contribute to the development of science by providing material necessary for biological and medical science. The scope of PCM activities covers the entire country. The Polish Collection of Microorganisms cooperates with other polish and foreign collections of microorganisms. The institute has laboratories enabling work on most of the current problems in the field of biology and medicine at the world level. The Institute has equipped laboratories for cell and tissue cultures, microbiological laboratories of pathogenicity classes 1 and 2, a virological laboratory, genetic laboratories allowing for work with genetically modified organisms (GMO and GMM classes 1 and 2), protein chemistry laboratories, a vaccine laboratory and others.

The Genomics and Bioinformatics Laboratory will be involved in the project. It posesses a Next Generation Sequencing Laboratory (NGS) with two high-throughput sequencers (MiSEQ and NextSEQ) and a nanopore sequencing technique (MinION). In addition, the laboratory has a bioinformatics laboratory employing four bioinformatics (two with doctoral degrees and two with masters).

The Institute includes a Medical Center, containing: Phage Therapy Center, Laboratory of Tissue Immunology and the Laboratory of Infection Diagnostics SARS-CoV-2.

The Institute participates in the international project to improve anti-terrorist security in Europe financed by the Norwegian Financial Fund entitled "Strengthening CBRNE safety and security - Coordination and Standardization". The following task is being carried out at the Institute: “Innovative solutions for the detection of a biological agent”. The main goal of this stage is to develop algorithms for the detection of biological substances of bioterrorist importance.

The abovementioned IITD PAN resources, allowing for research in a very wide range of problems in the field of medicine and biology, require integration at the level of data analysis. The solution that allows for an effective answer to the research questions is the use of artificial intelligence algorithms in the analysis of the mass data obtained, e.g. omics data. IITD PAN took steps to develop the application of AI algorithms in research in the field of medicine and biology.

IITD PAN is developing topics related to the use of artificial intelligence and machine learning in the development of a molecular model of the pathogen.

The following topics are proposed as part of the project:

1. The use of artificial intelligence algorithms to search for nucleotide sequences related to the production of metabolites.
2. The use of artificial intelligence algorithms for the analysis of data from the electron microscope and NMR spectra.