DOI https://doi.org/10.15407/csc.2023.04.039
UDC 519.85

YU.H. TARASICH, PhD Inform. Technologies, Doctoral Student,
V.M.Glushkov Institute of Cybernetics of the NAS of Ukraine,
ORCID: https://orcid.org/0000-0002-6201-4569, Scopus Author ID 56436890300;
Glushkov ave., 40, Kyiv, 03187, Ukraine,
yutarasich@gmail.com

H.O. SOLOSHENKO,PhD Student, Kherson State University, Ukraine,
ORCID: https://orcid.org/0000-0001-9622-310X; Scopus Author ID 57878437800,
Universytets`ka st, 27, Kherson, 73000, Ukraine,
hannasoloshenko@gmail.com

NEUROSYMBOLIC APPROACH IN BIOLOGICAL RESEARCH

Modelling and studying the processes and methods of intercellular and intracellular signalling cascades regulation involved in the process of programmed cell death and searching for substances capable of influencing the activation or inhibition of the process of cell apoptosis and the methods of their transportation to a given cell, is one of the numerous actual and open issues in biological research. A safe and fast method for this that does not require research on living organisms is computer molecular modelling. Many approaches and tools have been proposed and developed in the last decade. In particular, today, we observe a wide use of analytical methods for drug creation and a search for effective treatment methods. Such methods include modern methods of artificial intelligence (AI) based on neural network technology and methods of modelling interactions in biological and chemical processes at different levels of abstraction. Neural networks are used to obtain the ligand representation, protein compounds, and others and to build predictive models of the molecular compound properties widely used in drug discovery research. Modelling methods for both continuous and discrete models are applied using various approaches: statistical, probabilistic, simulation, and visual. The most well-known and used molecular modelling methods include the docking method, the molecular dynamics method, and the Monte Carlo method. To date, many software tools that support these methods have been developed. However, the considered modelling approaches and tools have a number of disadvantages, which can be of critical importance for conducting experiments.

This article presents a new approach to modelling biochemical processes and biological systems based on the formalism of the behaviour algebra and algebraic modelling language APLAN and its combination with neural network methods, the so-called Neurosymbolic approach. In particular, the possibility of multilevel modelling (from the level of the atomic structure of substances and quantum–mechanical interactions to the level of interaction of biological objects) and modelling of biological systems as complex hybrid systems that combine discrete and continuous processes is considered. A brief review of the current research on using neural network methods in biological research was also presented.

Keywords*: Molecular Modelling, Algebraic Modelling, Neural Network Methods, Artificial Intelligence, Modelling of Biological Experiments, Cell Apoptosis Modelling.*

Introduction

The main challenges of current research in biology are the engineering development of drugs, the intensive im-plementation of cell and gene therapy, and the development of neuroengineering [1, 2]. In addition, the presence and importance of solving the problem of single-cell analysis [2–4] is noted, as it will allow us to obtain a clearer understanding of the dynamics of the development of tissues and organisms and also of structures in cell populations.

Mathematical methods play a significant role in studying and analysing biological processes and systems. There are the presentation and statistical processing of data from various experiments, the implementation of relevant calculations, the ana-

lysis and construction of mathematical models of complex biological systems, and so on. Mathematical models are successfully used to study oxygen transport, protein interactions, tumour angiogenesis, and various cancer treatment methods [5–7].

There are a lot of approaches, methods (Monte Carlo, M.D. of molecular dynamics, M.D. of molecular mechanics, Docking, etc. [8–10]) and tools (AutoDock, ChemModLab, FlexAID, OCHEM, Open3DGrid, QSAR-tools, HORTON, DataWarrior, Bio-PEPA and many others [11–14]) to modelling biological processes and systems. However, despite the possibilities of these tools and methods, the discoveries in the fields of biology, chemistry, physics, and medicine, the complexity of biological systems, the need to process a large amount of data, and, unfortunately, the presence of some disadvantages (low accuracy, limitations of the framework of biological experiments, the need for a responsible selection of research methods and tools, the presence of errors in the structures of molecules with which software tools work, and so on), leave the issue of finding new approaches and tools as open. In particular, the open issues are to solve technical and computational problems of existing systems and to create such approaches to the mathematical representation of models of biological and biochemical processes, which would provide the possibility of the most complete reproduction of experiments in vivo in a computer environment. That is, the main task is to develop a system that will be able to study, explain, predict and indicate possible ways of controlling processes that occur in biological systems.

Neural networks are one of the powerful tools widely used in biological research today [15–23]. Artificial Intelligence (AI) systems built based on neural networks allow us to solve many urgent issues, such as:

• Molecule generation: the ability to generate new molecules that meet given criteria.

• Optimization of synthesis: finding the best synthesis path for a certain molecule, considering the availability of substances, efficiency of reactions, purity of products and environmental safety.

• Prediction of activity and affinity: prediction of desirable or undesirable biological activity or affinity of a molecule to a certain target based on its structure or functional groups.

• Study of the interaction of molecules: understanding the processes of interaction of molecules with other molecules, proteins, nucleic acids, and so on.

It is important to have the possibility of implementing existing molecular modelling methods (molecular docking, molecular mechanics, hybrid quantum-mechanical/molecular-mechanical simulations) and deep learning models for predicting the structure, energy, kinetics, and thermodynamics of molecular interaction and also methods for processing large data (Big Data) into such systems. The Big Data in biological research are molecular data (information about the structure of molecules, proteins, genes, etc.), clinical information, research and experimental data (for example, structural files containing information on the formation of connections between substances and characteristics of experiments), organism models, microscopic/X-ray images, etc. In particular, in the field of pharmacology and medicine, these are data from clinical trials, rare diseases, indications, side effects, protein interactions, and so on.

On the other hand, considering a number of previously described shortcomings of the current methods, we cannot claim that the results obtained by the neural network/artificial intelligence system are not erroneous, although they will indeed be such as to narrow the search. Accordingly, there is a need for additional experiments and verification and confirmation of the obtained results. In some cases, these may be relevant laboratory experiments, but, for example, taking into account such processes as the manufacture of drugs, we are talking about the high cost and labour-intensiveness of their implementation. We can see the solution to this problem by applying the algebraic approach and the corresponding formal methods [24]. Thus, the main idea of our research is to use the technology of algebraic modelling and quantum-chemical apparatus for modelling and verification of experiments in the field of medicine and pharmacology, in particular, modelling and verification of different approaches to the issue of investigating the properties of intercellular and intracellular interac-

tions in different environments. We understand the environment as a set of cellular and extracellular structures and influencing factors. In particular, the possibility of using the algebraic modelling system to study the nature of the interaction of cells with other agents under different conditions of the environment (change in temperature, concentration and structure of substances, acidity, etc.) is considered. The role of the neural network is to analyze the current state of the environment and determine the most effective action that will lead to the desired property (for example, activation and achievement of the cell's programmed death under the influence of drugs or irradiation, and so on).

Unlike traditional methods of modelling or simulation, which work with one specific scenario of behaviour, the application of algebraic modelling makes it possible to consider multiple scenarios of the behaviour of cellular compartments and structures. Moreover, algebraic modelling is multi-level, which allows the modelling of the cell structure and cellular processes at a higher level of abstraction based on intermolecular, quantum interactions, and, in particular, at the level of electron-electron interactions.

The article presents an overview of the results and prospects of using neural networks and artificial intelligence systems built on their basis in biological research, considers the general scheme of the approach proposed by the authors, and gives an example of a possible combination of neural network methods and algebraic modelling for modelling biochemical processes in cells (using the example of the process of programmable dead cells – apoptosis).

Neural Networks in Biological Research

The use of neural networks and artificial intelligence systems built on their basis is gaining more and more popularity and bringing significant results in such fields as medicine, pharmacology, genetics, biochemistry, etc. Considerable results have been achieved in the diagnosis of diseases of the organs of vision and recommendations for their treatment, for the diagnosis of diseases of the cardiovascular system, oncological diseases, and many dangerous infections. Available algorithms and tools allow for quick and relatively accurate analysis of medical images (CT and MRI) and, as a result, to identify patterns and anomalies that are invisible to the human eye. One example of such an application of neural networks is the Deep Gestalt network, capable of detecting a significant number of rare hereditary diseases by analyzing the facial features of the patient in the photo [25]. In addition, today, we are discussing the possibility of using neural networks to predict the structure of proteins, discover molecules, and so on. So, in 2020, it became known about the use of artificial intelligence for drug development by the Japanese pharmaceutical company Sumitomo Dainippon Pharma and the British startup Exscientia. As a result of research, a drug molecule was created, which entered the first phase of clinical trials [26]. Insilico Medicine, Evotec, and Schr dinger have also started testing several other molecules invented by Exscientia. Today, in addition to Exscientia, such companies as Schr dinger, Insitro, AbCellera, Relay Therapeutics, Atomwise, Recursion Pharmaceuticals, Cellarity, and so on are working on the application of neural networks and artificial intelligence systems for drug discovery.

The principles of operation of neural networks and features of their application in a wide range of biological research are considered in articles [15–23]. Thus, in the article [15], the use of deep neural networks (DNN), namely the "Chemi-Net" program, is considered for predicting the properties of molecular compounds (the processes of absorption, distribution, metabolism and elimination of drugs (ADME)). The base of Chemi-Net is a molecular graph convolutional network and a multitask neural network (MT-DNN) method used to improve prediction accuracy. The input data are molecules with given 3D coordinates of each atom, previously determined by processing SMILES molecular structure files. The output is the ADME properties predicted for the input molecule. Experiments on multitasking prediction were performed on datasets of solubility and rate of inhibition of the PXR (master protein regulator cytochrome P450-3A, or CYP3A) molecule.

The article [16] proposes the use of the geometric deep learning (GDL) method to predict the resistance of the human immunodeficiency virus to drugs and the interaction of HIV with drugs. The input data is a set of drug data in the SMILES format. Output data – prediction of resistance of the HIV virus to the drug under consideration. The prediction result is determined by the QED index (quantitative evaluation of drug similarity), which is an index that shows the interaction of the drug with the virus, using available information in the range (0–1).

The article [17] discusses the application of artificial neural networks to the study of potential drug interactions (DDI). Data from the DrugBank database were used as input data (contains bioinformatics and chemoinformatics resources that combine details of drugs – there are files with information on the molecular structure of drugs (formats – MOL, SMILES, PDB, SDF, etc.), information about reactions – a table "substrate/enzyme/product, where you can view information about their molecular structure and characteristics). Output data is the prediction of drug interactions.

The article's authors [18] describe the implementation approach and analyze the performance of multitasking deep networks and corresponding deep models using the DeepChem open-source platform as an example and the possibility of its application for drug discovery. Inputs are datasets from the Kaggle (enzymatic inhibitors), Factors (PRSS12 (serine 12) inhibition compounds), Kinase (protein kinases), and UV (over 10,000 compounds) collections.

The article [19] describes the software tool MAGIK, which provides a graph neural network (GNN) framework for estimating the dynamic properties of moving objects based on time-lapse experiments. MAGIK models object movement and physical interaction using a graphical representation. Input data - microscopic images – a sequence of images illustrating the evolution of a group of cells over time corresponding to the frame numbers. Output data – linking coordinates in the trajectory; definition of local and global dynamic properties.

In the article [20], the authors build a model for predicting the simultaneous inhibition of the primary human cytochrome CYP450 isoforms by training a multitask deep autoencoder neural network (DNN). Input data is a dataset of compounds obtained from the PubChem BioAssay database. Each data set contained a compound activity score, potency, curve description, fitted log EC50, held R-squared and activity. Output data is a substance-inhibitor/probability that the selected substance is an inhibitor.

The article [21] discusses deep learning methods for predicting the affinity between a drug and a target protein. Input data are the representation of drugs and protein sequences in SMILES format/ based on the DAVIS dataset (data for drug-target binding affinity prediction experiments) and the KIBA dataset (drug target prediction dataset). Gene names and RefSeq accession numbers were derived from protein sequences found in the Davis databases retrieved from the UniProt database. Output data is the prediction of the binding affinity of drugs and their targets.

In [22], the authors describe the development of DrugCell. It is an interpreted model of deep learning of human cancer cells, trained on the reactions of 1235 tumour cell lines to 684 drugs. DrugCell uses a modular neural network that combines conventional artificial neural networks (ANN) with a visible neural network (VNN) to predict drug response. Binary encodings of individual genotypes are processed through VNNs with an architecture driven by a hierarchy of cellular subsystems, with multiple neurons assigned to each subsystem. Complex chemical structures are processed through ANNs using the Morgan fingerprint method as input characteristics. Input data are the drug molecules and samples of malignant tumours. Output data are survival curves for drug combinations predicted by DrugCell to be effective/ineffective.

The article [23] describes the Deep Docking (DD) platform, which uses artificial intelligence methods to improve the docking method, such as predicting the docking score based on two-dimensional molecular fingerprints, with automatic sorting of the “necessary” molecules. Input data – compound data from ZINC chemical library and "make-on-demand" collection from Enamine in SMILES format, 3D receptor structure in PDB

format. The output is a set of the most suitable molecules for docking evaluation, which existing docking tools can use.

This is only a brief overview of existing scientific works devoted to the issue of using neural network methods in biological research. There are many other, no less important works. However, despite the significant breakthrough in the field of drug discovery and treatment methods due to the use of neural network methods, scientists talk about a number of open questions and challenges, including:

Ability to predict interactions of more than two drugs or targets simultaneously. Determining which drug combinations are most successful for additional in vitro or in vivo testing in many types of preclinical models, such as higher-order combinations among new therapeutic compounds and doses.

• Identification of effective ways to reduce the need to generate large amounts of single-cell data to predict response to combination therapy and the impact of toxicity, as well as recommended dosing that optimizes both efficacy and safety.

• Needs to use patient data and clinical profiles to validate in silico predictions of therapy response.

From a technical point of view, the most common problems for the implementation and effective operation of neural networks and artificial intelligence systems are data limitations of biological experiments, severe discrepancies between the distributions of training and real test data, difficulties of analysis and interpretation, the impact on the effectiveness of training of the growth of the amount of data, etc.

Neural Networks and Algebraic Modelling in Studying Biochemical Processes and Systems

General scheme of the proposed approach

The diagram that describes the application of the algebraic modelling system and neural network methods to the modelling of experiments in the field of biology, as well as the main components and tools necessary for conducting the research, is presented in Fig.

The first stage is the preparation of modelling data, namely, the presentation of knowledge in the form of algebraic specifications. Accordingly, the input data are the formal representation of the modelled environment and the formalized properties that must be achieved in the modelling process. This can be, for example, a list of cell structures and a description of their interaction processes, and as a property, for example, the possibility of achieving programmed cell death. At this stage, to obtain the correct algebraic representation of the model, the interaction of algebraists with experts (biologists, chemists, physicists) is important.

The formalization of knowledge is multilevel. That is, it can be carried out at any level of abstraction – at the level of the atomic structure of substances and quantum-mechanical interaction; at the level of the molecular structure of substances taking into account their intermolecular interactions; at the level of interactions between substances and at the level of biological objects. Multilevel modelling is provided by the presence of a unique knowledge base, which contains a set of formalized knowledge and theories from quantum physics, chemistry, biology, and so on. The appropriate level of abstraction is chosen according to the purpose of the experiment and taking into account the impact on the result of the lowest level of abstraction.

For modelling, we use the Insertion Model Creator system [27–29], which formalizes the subject area and creates an experiment's model that will be solved in terms of formal theories. The model is then transferred to the Algebraic Server, which combines formalized mathematical theories and corresponding methods that work with models and solve problems. The main theory is the theory of agents and environments, initiated by the Ukrainian academician O.A. Letychevskyi and the British scientist D. Gilbert [28]. Accordingly, we consider biological systems, or some structural elements of biological systems, as agents interacting with each other in some environment.

The environment may also be an agent that interacts with similar agents in a higher-level environment and so on. Each agent has its own type, which is determined by the attributes of the agent. Each attribute is typed and belongs to a certain theory in which predicates and operations are defined. We can define arithmetic, symbolic, bit, and byte

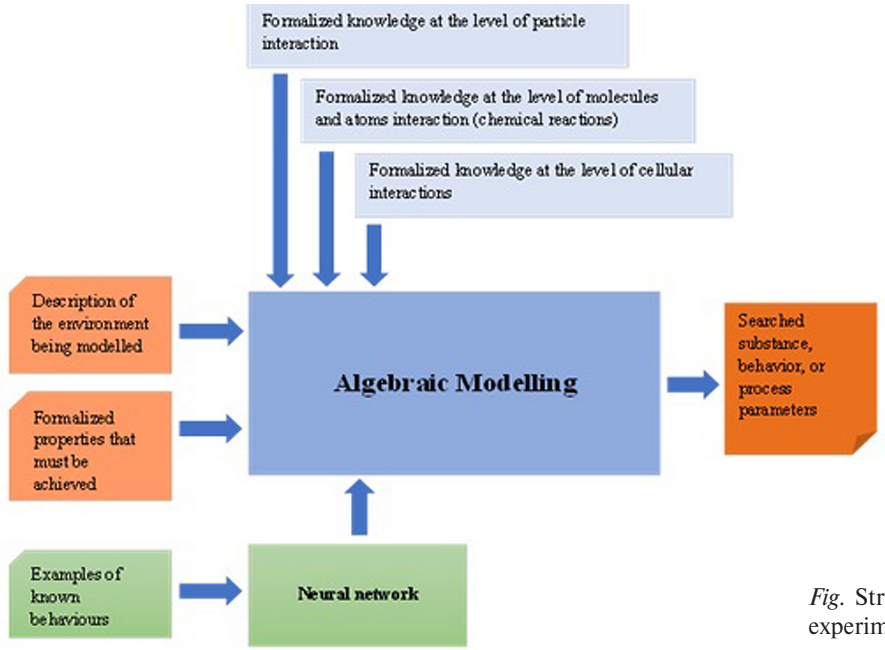

Fig. Structural diagram of biological experiments modelling

attributes. Attributes also can be parameterized, i.e., they can define some functional symbol. Due to the presence of a large number of different types of attributes in different theories, it is important to define the problem of executability of the formula in the chosen theory, i.e. to solve the problem of finding the values of attributes in the formula with which the formula is true. This problem is solved by so-called solvers, or solver systems and systems for automatic proof of theorems, which are the basis of the algebraic approach in modelling.

The process of interaction of agents in some environment is described by a behavioural equation, which is an equality in which the name of the behaviour is on the right, and the expression in the algebra of behaviours over actions and other behaviours is on the left. An action is analogous to the Hoare triple and contains a precondition that determines the possibility to perform the action and then a postcondition that determines the transformation or change of the attributes of the agent. The precondition is determined by the predicate, which is a Boolean expression over the formulas of the corresponding theories, for example, – equality or inequality in linear arithmetic. Postcondition changes the environment and also uses predicates, assignment operators, and operations in the chosen theory [24, 29, 30]. Algebraic modelling is the basis of solving behavioural equations. The solution of the behavioural equation is a set of scenarios or traces that consist of a sequence of actions and behaviours and lead to the reachability of a given property.

For the modelling of hybrid systems (systems that combine a discrete and a continuous component), the Insertion Model Creator system is extended with the possibility of an analytical solution of differential equations, the operators of which are executable algebraic specifications.

To avoid the possible phenomenon of combinatorial explosion due to the high complexity of the considered models, we use the neural network methods (AI methods) and Big Data processing methods, as they can indicate the correct modelling path. We create a neural system that analyzes the current state of the environment and determines the most effective action that will lead to the desired property. We train the neural network on the sequences of such actions that reach the required property. For example, you can set possible options for developing the process of programmed cell death (possible sequences of actions) depending on the initial state of the modelled environment. The classification model, in turn, indicates

the most effective trace or action to be taken or rejects traces that reach an undesirable outcome. In addition, neural networks can be used as one of the data sources for modelling. For example, we can use a neural network or a corresponding artificial intelligence system for the first stage of searching for substances with given properties and only then apply algebraic modelling to prove the properties of a given substance.

As a result of the modelling, we obtain a set of possible scenarios of the model's behaviour (forward algebraic modelling), or we can determine the substances that will correspond to the given parameters and the desired behaviour of the model (forward and/or backward algebraic modelling), determine the necessary initial state of the environment under which it is possible to achieve the desired scenario behaviour (backward algebraic modelling).

Formalization of Knowledge. An Example of Modelling the Process of Programmed Cell Death

The process of programmed cell death (apoptosis) is one of the main cellular processes that unites many biological studies. In particular, it is interesting to consider the possibility of activating and inhibiting the process of apoptosis as a method of combating ageing, cancer cells, human immunodeficiency disease, and so on. Accordingly, one of the directions of our research is the application of algebraic modelling methods to the modelling and research of processes and methods of regulation of intercellular and intracellular cascades of signal transmission, which participate in the process of cell apoptosis, taking into account their interactions with cell compartments and their structures. One of the experiments is the search for nanoparticles or nanoparticle composites with magnetic properties capable of activating the tumour cell apoptosis process and/or protecting a healthy cell from radiation, and, in particular, the possibility of modulating pharmacokinetic processes in a tumour cell. In the model at this level, we will take into account the qualitative and quantitative characteristics of such elements of the cell and the intercellular environment as proteins (TRADD, FADD, RIP, TRAF2, APAF-1, NF-kb, c-Jun, clAP (1, 2), some proteins of the Bcl-2 family), caspases (2, 3, 8, 9), cytochrome, reactive oxygen species, ions (Ca, Na), and so on. These elements and nanoparticles are considered as agents whose environment of interaction is the cell. In addition, we took into account the following parameters of the environment: temperature, acidity, the presence of a magnetic field and its characteristics (constant, alternating, pulsed, combined), and so on. As the action of the nanoparticle agent, we consider all possible interactions with key metal-containing proteins and enzymes.

Agents and Attributes

To formalize the knowledge of the lowest level of abstraction (quantum and interatomic interactions), we define such types of agents as ELECTRON (agents are electrons as individual particles), PROTON (agents are protons as individual particles) and ATOM (agents of the ATOM type can be ions, atoms or isotopes). For models of a higher level of abstraction, these agents can be present both as attributes of other agents (molecules, amino acids, proteins, etc.) and as agents that interact with higher-level agents in some environment (for example, a calcium ion in a cell; an ion, a proton or an electron as a particle injected into a substance during radiation therapy, etc.). As an example of the formalization of the agent type, let's consider an agent of the PROTEIN type, which presents proteins.

```
PROTEIN:obj(
polypeptideChainsNum:int,
acidsNum(i):int,
aminoAcids:(int,int)-> AMINOACID,
prostheticGroup:(int)-> MOLECULE,
peptideBond:(int, int, int)->bool,
domains:(int)-> domainType,
mass:real,
.... ),
```

The main attributes that characterize this agent type are:

- polypeptideChainsNum – an integer type attribute that determines the number of polypeptide chains,
- functional attribute AcidsNum(i) the number of amino acids of the corresponding polypeptide chain,

• unctional attribute aminoAcids:(i,j) -> AMINOACID for storing the list of amino acids (i – chain number, j – amino acid number) that are part of the corresponding polypeptide chain,

• functional attribute PeptideBond(k,i,j)-> bool, which determines the presence of a bond between the i-th and j-th amino acids of the k-th chain,

• functional attribute prostheticGroup:(int)-> MOLECULE, which defines (if available) an additional group of non-protein nature that is a component of a complex protein,

• functional attribute domains(i)->domainType, which stores the list of domains of this protein,

• mass – protein mass, and so on.

As you can see, the PROTEIN type agent contains AMINOACID and MOLECULE type agents as attributes. From a biological point of view, an amino acid is also a molecule – an organic substance, the physical structure of which includes an amino group (- NH2) and a carboxyl group (- COOH), and therefore can be formalized as a type of MOLECULE agent. On the other hand, amino acids have their own set of attributes that can be specific to a given type of molecule/substance, which is why we consider amino acids as agents at a higher level of abstraction that uses the MOLECULE agent as an attribute.

Attributes of agents of MOLECULE type, in turn, are represented by numerical values of the following parameters: the set of atoms that comprise it (atoms), the electronic configuration of the molecule (MolOrbital), bond length (d_bond), bond energy (BondEnergy), dipole moment (DipoleMoment), molar mass (M_r), bond order (bondMO – by the method of molecular orbitals, bondV – by the method of valence bonds) and bond type (bondType), etc.

The properties of the atoms are determined by the structure of their nuclei and the number and organization of orbital electrons. Accordingly, the main attributes that characterize substances are the structures of their molecular orbitals and the nuclear models of atoms that are part of the substance. This allows us to consider all the processes of their interactions at the level of quantum interactions.

The set of values of all agent attributes determines its state. It should be noted that the main advantage of algebraic modelling is the ability to consider not only specific states of agents (all attributes have specific values) but also symbolic states (attributes can be specified by formulas – using inequalities, Boolean functions). For example, we can say that the temperature of the environment in which the agents interact is 22 degrees (temperature == 22) and the concentration of nanoparticles is 400 μg/ml (nanoparticle.concentration == 0,4), which is a specific value. In the case where we are considering a symbolic simulation, we can specify that 18<=temperature <= 22 and 0,2<=nanoparticle.concentration <= 0,8.

The functional attribute proteins(i)->PROTEIN defines all proteins considered in the model.

Interactions of agents are defined using formal actions that determine a change in the state of an agent or a change in the values of its attributes. Such actions include, for example, the formation of bonds between atoms, intermolecular interactions, interactions between substances (in particular, the formation of bonds between amino acids, the formation of protein structures, interactions of proteins with other proteins and substances), and so on. Actions are defined in the language of behaviour algebra, which is algebraic specifications. Each action has a precondition for its execution and changing the values of the attributes of the agent/agents. For example, let's consider the CreatePeptideBond action, which determines the formation of a peptide bond between two amino acids:

CreatePeptideBond = Forall(i: AMINOACID, j: AMINOACID, n1:int, k1:int, n2:int, k2:int,)

(i !=j &&(1<= n1 <= i.molecule.atomsNum) &&(1<= k1 <= i.molecule.atomsNum) &&(1<= n2 <= j.molecule.atomsNum) &&(1<= k2 <= j.molecule.atomsNum) && n1 != k1 && n2 != k2 && i.molecule.MolOrbital(n1,k1,-1,1)==2 && i.molecule.atoms(n1).Name == O && i.molecule.atoms(k1).Name == H &&

j.molecule.MolOrbital(n2,k2,-1,1)==2 &&

j.molecule.atoms(n2).Name==N && j.molecule.atoms(k2).Name == H)

->

" PROTEIN #P1: action 'CreatePeptideBond';"

(Protein1.aminoAcids(1,1) = i;

Protein2.aminoAcids(1,2) =j;

i.molecule.MolOrbital(n1-1,n1,-2,1)=2;
i.molecule.atoms(n1)=j.atoms(n2);
j.molecule.atoms(k2)= i.atoms(n1);
j.molecule.MolOrbital(n2,k2,-2,1)=2
);

In the precondition of the action, it will be checked that one of the amino acid residues has a free amino group and the other has a free α-carboxyl group. If this condition is satisfied, a bond will be created between amino acids, and we will "write" them into the corresponding polypeptide chain. The first will be written the amino acid residue with a free amino group, and the second with a free α-carboxyl group. Similarly, other amino acids are added to the chain. We also record the electronic structure of the dipeptide. In this case, since the molecular orbitals of the amino acids are the molecular orbitals of the protein, we change the structure of the amino acids and create the corresponding orbital. All other orbitals remain unchanged.

Accordingly, by representing all structural units of the cell and external factors in the form of agents and attributes of the environment, we obtain a formal representation of the model. By formalizing all possible interactions between agents and setting different values of attributes of agents and the environment, we can consider different scenarios of agent behaviour.

So, for example, we consider possible options for activating the process of apoptosis of a cancer cell by a nanoparticle (MOLECULE-type agent), such as the activation of death receptors on the cell surface (external pathway of the apoptosis process activation) or the penetration of nanoparticles into the cell and the start of the mitochondrial pathway of activation of the cell death process. In particular, we consider modelling for different initial states of the environment, i.e., we consider different ratios between factors of activation and inhibition of the apoptosis process and different values of environmental parameters capable of influencing the course of the corresponding reactions – temperature, presence and characteristics of a magnetic field, concentration and sizes of nanoparticles, concentration reactive forms of oxygen and Ca2 ions in the cell, and so on.

At the higher level, the formalization of the model behaviour for this experiment will be present in the following form:

APOPTOSIS_PROCESS = (CONNECTION_NANOPARTICLE_TO_CELL + PENETRATION_NANOPARTICLE_INTO_CELL),

CONNECTION_NANOPARTICLE_TO_CELL = (BINDINGtoTNF-R1 + BINDINGtoFAS-R)),

BINDINGtoTNF = (bindingNanoparticleToTNFR1receptor. bindingTNFR1withTRADD; (bindingTRADDwithRIP; (bindingRIPwithTRAF2. NFkBinducingKinaseActivation. NFkBinhibitorPhosphorylation. InterleukCytokinSynthesis. ApoptosisInhibition + (binding RIPwithRAIDD; CASPASE_CYCLE_ACTIVATION)) + + (bindingTRADDwithFADD ; (binding FADD withcFLIP. ApoptosisInhibition + CASPASE_CYCLE_ACTIVATION)))),

CASPASE_CYCLE_ACTIVATION = (bindingwithProcaspas; (InitiatorCaspasesActivation; (ExecutionerCaspasesActivation + MITOCHONDRIA _APOPTOSIS); MORPHOLO_-ICAL_CELLS_CHANGES) + InitiatorCaspasesInhibition. ApoptosisInhibition),

....

In the first part, we consider two possible options for the interaction of a nanoparticle with a cell – CONNECTION_NANOPARTICLE_TO_CELL behaviour (a set of actions and behaviours describing the external path of apoptosis process activation) and PENETRATION_NANOPARTICLE_INTO_CELL behaviour (a set of actions and behaviours describing the internal path of apoptosis process activation). Next, we consider all sets of actions that describe the possible interactions of agents and changes in the state of the environment depending on the "chosen" path. For example, we consider the external pathway of activation of the apoptosis process, namely activation of the TNF-R1 receptor – BINDINGtoTNF behaviour.

The BINDINGtoTNF behavior consists of such actions as: bindingNanoparticleToTNF-R1receptor – binding of a nanoparticle to the cell death receptor TNF-R1; bindingTNF-R1withTRADD – interaction of the TRADD adapter protein with the TNF-R1 death domain;

bindingTRADDwithRIP – the connection of the adapter protein TRADD with RIP proteins (here we also consider the possible sequence of actions bindingRIPwithTRAF2, NFkBinducing-KinaseActivation, NFkBinhibitorPhosphorylation, InterleukCytokinSynthesis, simulating the activation of NF-kB-inducing kinase and as a result – blocking the process of apoptosis or the action of bindingRIPwithRAIDD (RIP interacts with adapter proteins RAIDD, which can bind caspase 2 and activate apoptotic cell death (behavior CASPASE_CYCLE_ACTIVATION))) or bindingTRADDwithFADD connection of the TRADD adapter protein with FADD proteins (activation of the signaling complex of cell death ((behavior CASPASE_CYCLE_ACTIVATION))) or blocking the process of apoptosis by connecting with the protein c-FLIP (action of binding FADDwithcFLIP)).

Behaviors CASPASE_CYCLE_ACTIVATION, MITOCHONDRIA_APOPTOSIS); MORPH-?LOGICAL_CELLS_CHANGES correspond to the processes of activation of the caspase cycle, opening of mitochondrial channels and activation/maintenance of the cell apoptosis process, morphological changes of the cell as a result of achieving the irreversibility of the apoptosis process, respectively, and also consist of sets of actions that model the interactions between the relevant agents participating in these processes (caspases, Bcl-2 family proteins, cytochrome, APAF-1, etc.).

Since the task of this experiment is to determine such properties as reaching the final stage of cell apoptosis under different values of environmental parameters, and the main biochemical marker of apoptosis is the cleavage of chromosomal DNA, which leads to an increase in the concentration of Ca2 ions in the cell nucleus and the presence of dependent proteases in the cell nucleus (DFF, AIF, EndoGin, etc.), under the action of which DNA fragmentation occurs, the formula that determines the achievement of the final stage of the apoptosis process in the cell is defined as:

ConcentrationCa2>NormalConcentration && (AIFinNucleus==1 || DFFinNucleus==1 || EndoGinNucleus==1) && DNAfragmentation==1

The reachability of this formula indicates the reachability of the specified property under the given initial parameters. For example, in the experiment, it was determined that under the condition of the advantage of connections of the adapter protein TRADD with DD-containing FADD proteins and, accordingly, under the conditions that the activity of procaspase 8 will be higher than the activity of the c-FLIP inhibitor, and the activity of bcl2 family proteins capable of inducing apoptosis will be higher than the activity of its inhibitors (procaspase8Concentration > cFLIPConcentration && bcl2Inhi-bitionActivity < bcl2InductionActivity) the mitochondrial pathway of cell apoptosis will additionally be launched and, accordingly, the proapoptotic protein – AIF will be released and translocated to the nucleus, due to which the degradation phase of cell apoptosis will be reached. Therefore, the achievement of the specified property will be proven.

We also consider the need to model reversible biochemical processes and reactions (reactions that can co-occur in the forward and reverse directions), which is determined by the behavioural equation and the prerequisites for each action. So, for example, if there is a deficiency or an excess of certain agents capable of influencing the development of the apoptosis process or caspase cycle at a certain stage (ATP level, temperature, acidity, etc.), we can return to the previous phase (transition to the execution of an action or fragment of behaviour, which corresponds to the given condition).

As a result of running the model, we get a scenario/set of possible scenarios (depending on whether we perform concrete or symbolic modelling) of behaviour and determine the reach of the stage of the onset of morphological changes in the cell depending on the ratio of factors inducing or inhibiting the process of cell apoptosis.

It is worth noting that such a model is not complete and ready for real use in the medical field. These are only the first steps of an experiment that shows the possibilities of using the proposed method.

Conclusions

The search for new approaches and methods to the modelling and research of experiments in biology,

in particular, the processes that occur in cells under various scenarios of interaction with external and internal factors of influence (radiotherapy, nanotechnology, inflammatory processes, etc.), remains an open issue. Research on pathological apoptosis and the possibility of using this type of cell death in medical practice (oncology treatment, cardiovascular diseases, human immunodeficiency virus, etc.) deserve special attention.

Hybrid models and methods for systems biology and medicine (including working with hybrid formalisms such as temporal and hybrid automata), combining models by integrating combinatorial and continuous constraints, and using machine learning to design models and define their parameters are important steps in solving open issues in the field of modelling and research of relevant processes and systems. However, despite the availability of more and more data on existing proteins and nucleic acids, modelling methods and tools, the development and use of a wide range of combined methods and tools for modelling and computing large molecular systems is one of the main challenges facing scientists at the intersection of natural sciences and exact sciences.

One of the approaches that, in our opinion, will allow us to solve most of the open issues is the neurosymbolic approach, that is, the approach that, in this case, will combine the methods of neural networks and algebraic modelling. The main advantage of applying the proposed approach is that it provides an opportunity to derive consequences from existing laws and, therefore – can provide new facts and theories that will allow solving complex problems. In other words, the use of an algebraic approach in combination with the methods of neural networks allows the determination and formal proof of certain properties of objects (in this case – charged particles, atoms, organic and inorganic substances, cells, viruses, etc.) or processes, and also, the searching of objects or the necessary values of their parameters that correspond to the specified properties.

Although research on the application of the proposed approach in the field of biology is not yet complete, and we are currently working on expanding the base of formalized knowledge, the first obtained results of modelling interatomic and intermolecular interactions, intracellular processes indicate that the proposed neurosymbolic approach is effective and promising for modelling biochemical processes and biological systems.

The following steps are the study of the necessary initial values of the agent attributes and the modelled environment, the expansion and complication of the model, the conduct of forward concrete modelling (to compare the obtained results with the data of biological experiments) and backward modelling on the obtained models, as such, which will allow determining the necessary parameters of the attributes of the agents and environment required to reach the specified property. In particular, we see the possibility of continuing the research for modelling the following experiments:

1) Study of the effects of the use of nanoparticles for the controlled production of ROS and photothermia aimed at the induction of oxidative stress and selective death of tumour cells (reaching the degradation phase of tumour cell apoptosis – forward modelling; determination of the necessary initial state of the environment under which it is possible to achieve the desired behaviour scenario (temperature indicators, concentration and structures of substances, acidity, characteristics of cells and nanoparticles under investigation, etc.) – backward algebraic modelling).

2) Study of the influence of different doses of radiation on DNA damage, increase in the level of ROS (reactive oxygen species) as a prerequisite for the intracellular initiation of the process of pathological cell apoptosis (achievement of the degradation phase of tumour cell apoptosis – forward modelling; determination of the necessary initial state of the environment under which it is possible to achieve the desired behaviour scenario (indicators of temperature, concentration and structure of substances, acidity, characteristics of the cell under investigation, etc.) – backward algebraic modelling).

3) Study of the effect of cofactors, activators and inhibitors on the change in the activity of enzymes (metaloproteases, glutaminase) (achieving the degradation phase or blocking the process of cell apoptosis – forward modelling; determination of

the necessary initial state of the environment under which it is possible to achieve the desired scenario of behaviour (temperature indicators, concentration and structure of substances, acidity, characteristics of the cell under study, etc.) – backward algebraic modelling).

REFERENCES

1. Pande, V., Tran, A., 16 Open Problems in Engineering Biology, [online]. Available at: <https://a16z.com/16-open-problems-in-engineering-biology/>[Accessed 01 Nov. 2023].
2. Top Five Open Problems in Bioinformatics (2021), [online]. Available at: <https://homolog.us/blogs/bioinfo/2021/07/12/open-problems-bioinformatics/> [Accessed 01 Nov. 2023].
3. Open problems in single-cell analysis, [online]. Available at: <https://openproblems.bio/> [Accessed 01 Nov. 2023].
4. L hnemann, D., K ster, J., Szczurek, E. et al., 2020. "Eleven grand challenges in single-cell data science". Genome Biol 21(31).
5. Jones, B., 2017. "Clinical radiobiology of proton therapy: modeling of RBE". Acta Oncologica, 56(11), pp. 1374–1378.
6. Chen, Y., Ahmad, S., 2011. "Empirical model estimation of relative biological effectiveness for proton beam therapy". Radiation Protection Dosimetry, 149(2), pp. 116–123.
7. Dahle,, T. J., Rykkelid, A. M., Stokkev g, C. H., Mairani, A., G rgen, A., Edin, N. J., R rvik, E., Fj ra, L. F., Malinen, E., Ytre-Hauge, K. S., 2017. "Monte Carlo simulations of a low energy proton beamline for radiobiological experiments". Acta oncologica (Stockholm, Sweden), 56(6), pp. 779–786.
8. Mathpal, D., Masand, M., Thomas, A., Ahmad, I., Saeed, M., Zaman, G.S., Kamal, M., Jawaid, T., Sharma, P.K., Gupta, M.M., Kumar, S., Srivastava, S.P., Balaramnavar, V.M., 2021. "Pharmacophore modeling, docking and the integrated use of a ligand- and structure-based virtual screening approach for novel DNA gyrase inhibitors: synthetic and biological evaluation studies". RSC Advances. Vol. 11(55), pp. 34462–34478.
9. Lin, X., Li, X., Lin, X., 2020. "A Review on Applications of Computational Methods in Drug Screening and Design". Molecules, vol. 25(6):1375.
10. Beentjes, C.H.L., Baker, R.E., 2019. "Quasi-Monte Carlo Methods Applied to Tau-Leaping in Stochastic Biological Systems". Bull Math Biol, vol. 81, pp. 2931–2959.
11. Bitencourt-Ferreira, G., Pintro, V., de Azevedo, W., 2019. "Docking with AutoDock4". Methods in Molecular Biology, pp. 125–148.
12. Hughes-Oliver, J.M., Brooks, A.D., Welch, W.J., Khaledi, M.G., Hawkins, D., Young, S.S., Patil, K, Howell, G.W., Ng, R.T., Chu, M.T., 2012. "ChemModLab: a web-based cheminformatics modeling laboratory". In Silico Biol, 11(1-2), pp. 61–81.
13. Morency, L., Gaudreault, F. and Najmanovich, R., 2018. "Applications of the NRGsuite and the Molecular Docking Software FlexAID in Computational Drug Discovery and Design". Methods in Molecular Biology, pp. 367–388.
14. Pirhadi, S., Sunseri, J., Koes, D., 2016. "Open source molecular modeling". Journal of Molecular Graphics and Modelling, 69, pp. 127–143.
15. Liu, K., Sun, X., Jia, L., Ma, J., Xing, H., Wu, J., Gao, H., Sun, Y., Boulnois, F., Fan, J, 2019. "Chemi-Net: A Molecular Graph Convolutional Network for Accurate Drug Property Prediction". Int. J. Mol. Sci., 20, 3389.
16. Das, B., Mucahit Kutsal, Resul Das, 2022. "Effective prediction of drug–target interaction on HIV using deep graph neural networks". Chemometrics and Intelligent Laboratory Systems, 230, 104676.
17. Shtar, G, Rokach, L, Shapira, B, 2019. "Detecting drug-drug interactions using artificial neural networks and classic graph similarity measures". PLoS ONE, 14(8): e0219796.
18. Ramsundar, B, Liu, B, Wu, Z, Verras, A, Tudor, M, Sheridan, RP, Pande, V., 2017. "Is Multitask Deep Learning Practical for Pharma?". J Chem Inf Model, 57(8), pp. 2068–2076.
19. Pineda, J., Midtvedt, B., Bachimanchi, H. et al., 2023. "Geometric deep learning reveals the spatiotemporal features of microscopic motion". Nat Mach Intell 5, pp. 71–82.
20. Li, X, Xu, Y, Lai, L, Pei, J, 2018. Prediction of Human Cytochrome P450 Inhibition Using a Multitask Deep Autoencoder Neural Network. Mol Pharm, 15(10), pp. 4336–4345.
21. Sharma, M., Deswal, S., 2022. "Drugs–Protein affinity score prediction using deep convolutional neural network". Expert Systems, 39(10), e13154.
22. Kuenzi, B.M., et all., 2020. "Predicting drug response and synergy using a deep learning model of human cancer cells". J Elsevier Cancer Cell, 38(5):, pp.1535–6108.
23. Gentile, F., Yaacoub, J. C., Gleave, J., Fernandez, M., Ton, A. T., Ban, F., ... & Cherkasov, A., 2022. "Artificial intelligence–enabled virtual screening of ultra-large chemical libraries with deep docking". Nature Protocols, 17(3), pp. 672–697.

24. Letychevskyi, O., Tarasich, Y., Peschanenko, V., Volkov, V., Sokolova, H., 2022. "Algebraic Modeling of Molecular Interactions". Communications in Computer and Information Sciencethis link is disabled, 1635 CCIS, pp. 379–387.
25. AI drug discovery: assessing the first AI-designed drug candidates to go into human clinical trial, [online]. Available at: <shttps://www.cas.org/resources/cas-insights/drug-discovery/ai-designed-drug-candidates>[Accessed 01 Nov. 2023].
26. Insertion Model Creator system, [online]. Available at <https://rd.litsoft.com.ua/> [Accessed 01 Nov. 2023].
27. Letychevskyi, O., Peschanenko, V., Poltoratskyi, M., Tarasich, Yu., 2020. "Platform for modeling of algebraic behavior: Experience and conclusions". CEUR Workshop Proceedings, 2732, pp. 42–57.
28. Letichevsky, A., Gilbert, D., 1999. "A Model for Interaction of Agents and Environments". In: Bert D., Choppy C., Mosses P.D. (eds) Recent Trends in Algebraic Development Techniques., WADT 1999, LNCS, vol. 1827, pp. 311–328.
29. Letychevskyi, O., Peschanenko, V., Volkov, V., 2022. Algebraic Virtual Machine and Its Applications. Communications in Computer and Information ScienceThis link is disabled., 1698 CCIS, pp. 23–41.

Received 22.09.2023

Ю.Г. Тарасич, доктор філософії, докторант, Інститут кібернетики імені В.М. Глушкова НАН України,
ORCID: https://orcid.org/0000-0002-6201-4569, Scopus Author ID 56436890300,
03187, м. Київ, просп. Академіка Глушкова, 40, Україна,
yutarasich@gmail.com

Г.О. Солошенко, аспірантка, Херсонський державний університет, Україна,
ORCID: https://orcid.org/0000-0001-9622-310X; Scopus Author ID 57878437800,
73000, м. Херсон, вул. Університетська, 27, Україна,
hannasoloshenko@gmail.com

НЕЙРОСИМВОЛЬНИЙ ПІДХІД У БІОЛОГІЧНИХ ДОСЛІДЖЕННЯХ

Вступ. Моделювання та вивчення процесів і методів регуляції міжклітинних та внутрішньоклітинних сигнальних каскадів, що беруть участь у процесі запрограмованої загибелі клітин, та пошук речовин, здатних впливати на активацію чи гальмування процесу апоптозу клітин та способів їх транспортування до заданої клітини, є однією з багатьох актуальних і відкритих проблем біологічних досліджень.

Мета. Дослідити задачу пошуку методів створення нових ліків для розробки ефективних методів лікування, для чого розробити новий підхід до моделювання біохімічних процесів та біологічних систем – нейросимвольний підхід, заснований на формалізмі алгебри поведінки та алгебраїчної мови моделювання *APLAN* у поєднанні з нейромережевими методами.

Методи. Безпечним і швидким методом розв'язання даних задач, який не потребує дослідження живих організмів, є комп'ютерне молекулярне моделювання. За останнє десятиліття запропоновано та розроблено багато підходів та інструментів. До таких методів належать сучасні методи штучного інтелекту, засновані на технології нейронних мереж, і методи моделювання взаємодій у біологічних і хімічних процесах на різних рівнях абстракції. Нейронні мережі використовуються для отримання представлення лігандів, білкових сполук тощо, а також для побудови прогнозних моделей властивостей молекулярних сполук, які широко використовуються в дослідженнях відкриття ліків. Методи моделювання як для неперервних, так і для дискретних моделей застосовуються з використанням різних підходів: статистичного, імовірнісного, імітаційного та візуального. Найбільш відомі та використовувані методи молекулярного моделювання включають метод докінгу, метод молекулярної динаміки та метод Монте-Карло.

Результати. На сьогоднішній день розроблено багато програмних засобів, які підтримують перлічені методи, проте розглянуті підходи та інструменти моделювання мають низку недоліків, що може мати критичне значення для проведення експериментів. У роботі представлено новий підхід до моделювання біохімічних процесів та біологічних систем, заснований на формалізмі алгебри поведінки і алгебраїчної мови моделювання *APLAN* у поєднанні з нейромережевими методами, так званий нейросимвольний підхід. Зокрема, розглядається можливість багаторівневого моделювання (від рівня атомної будови речовин і квантово-механічних взаємодій до рівня взаємодії біологічних об'єктів) та моделювання біологічних систем як складних гібридних систем, що поєднують дискретні та безперервні процеси. Також представлено короткий огляд сучасних досліджень використання нейромережевих методів у біологічних дослідженнях.

Висновки. Наведений підхід до моделювання біохімічних процесів та біологічних систем може бути використаний при створенні нових ліків та розробки ефективних методів лікування.

Ключові слова: *молекулярне моделювання, алгебраїчне моделювання, методи нейронних мереж, штучний інтелект, моделювання біологічних експериментів, моделювання апоптозу клітин.*