Cellular Physiology and Biochemistry Published online: 3 March 2021

Cell Physiol Biochem 2021;55(S3):14-45 DOI: 10.33594/00000336

Accepted: 11 February 2021

© 2021 The Author(s) Published by Cell Physiol Biochem Press GmbH&Co, KG, Duesseldorf www.cellphysiolbiochem.com

This article is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 Interna-tional License (CC BY-NC-ND). Usage and distribution for commercial purposes as well as any distribution of modified material requires written permission.

Review

Computational Ion Channel Research: from the Application of Artificial Intelligence to **Molecular Dynamics Simulations**

Sarah Maskri^a Janosch Menke^a Oliver Koch^{a,b}

^aInstitute of Pharmaceutical and Medicinal Chemistry, Westfälische Wilhelms-Universität Münster, Münster, Germany, ^bCenter for Multiscale Theory and Computation, Westfälische Wilhelms-Universität Münster, Münster, Germany

Key Words

Ion channel • Topology prediction • Structure-based design • Homology modelling • Docking Molecular dynamics simulations
 Machine learning

Abstract

Although ion channels are crucial in many physiological processes and constitute an important class of drug targets, much is still unclear about their function and possible malfunctions that lead to diseases. In recent years, computational methods have evolved into important and invaluable approaches for studying ion channels and their functions. This is mainly due to their demanding mechanism of action where a static picture of an ion channel structure is often insufficient to fully understand the underlying mechanism. Therefore, the use of computational methods is as important as chemical-biological based experimental methods for a better understanding of ion channels. This review provides an overview on a variety of computational methods and software specific to the field of ion-channels. Artificial intelligence (or more precisely machine learning) approaches are applied for the sequence-based prediction of ion channel family, or topology of the transmembrane region. In case sufficient data on ion channel modulators is available, these methods can also be applied for quantitative structureactivity relationship (QSAR) analysis. Molecular dynamics (MD) simulations combined with computational molecular design methods such as docking can be used for analysing the function of ion channels including ion conductance, different conformational states, binding sites and ligand interactions, and the influence of mutations on their function. In the absence of a three-dimensional protein structure, homology modelling can be applied to create a model of your ion channel structure of interest. Besides highlighting a wide range of successful applications, we will also provide a basic introduction to the most important computational methods and discuss best practices to get a rough idea of possible applications and risks.

© 2021 The Author(s). Published by Cell Physiol Biochem Press GmbH&Co. KG

J. Menke and S. Maskri contributed equally to this work.

PD Dr. Oliver Koch

Institute of Pharmaceutical and Medicinal Chemistry Westfälische Wilhelms-Universität Münster, Corrensstraße 48, 48149 Münster (Germany) E-Mail oliver.koch@uni-muenster.de; oliver.koch@agkoch.de; Homepage www.agkoch.de

Cellular Physiology	Cell Physiol Biochem 2021;55(S3):14-45				
and Biochemistry	DOI: 10.33594/000000336 Published online: 3 March 2021	© 2021 The Author(s). Published by Cell Physiol Biochem Press GmbH&Co. KG			
	Menke et al.: Computational Ion Channel Research				

Introduction

Ion channels are important membrane proteins that mediate fast electrical and chemical signalling by regulating passive ion transport across the cell membrane [1]. Ion transport takes place through a central pore as a common structural motif that is formed by four or five transmembrane helices of different subunits in most channels. A structural selectivity filter is used to distinguish the different ion species that can pass the pore. The change between the open and closed state is based on a conformational change that is mainly mediated by changes in the membrane potential (voltage-gated channels) or ligand-binding (ligandgated channels). In the latter case, channel activators stabilize the open conformational state and channel blockers stabilize the closed conformational state. In addition, channel blockers are known that bind to the central pore and block ion permeation. Fig. 1 shows the important features exemplary on a mammalian intermediate-conductance potassium channel (K_{c_3} 3.1) channel. There are over 340 genes reported that encode for ion channels with important functions in a plethora of physiological functions. The importance of ion channels is underlined by many severe diseases (the channelopathies) that are described due to impaired and dysfunctional ion channels. Therefore, ion channels are also attractive drug targets. For a more detailed introduction into ion channels, we would like to refer to Ashcroft [1] and Hille [2] as starting points.

Although there are a number of chemical-biological based experimental methods to study the function of ion channels [4], the use of computational methods is equally important for a better understanding of ion channels and their involvement in (patho)physiology. With this review, we want to provide a broad overview about the successful application of computational methods in ion channel research, combined with basic introductions and best practices to get a rough idea of possible applications and risks. We will start with artificial intelligence/machine learning based approaches, that are used to identify and classify ion channels based on their protein sequence, and quantitative structure-activity relationship (QSAR) approaches for the analysis of small molecule ion channel modulators. This is followed by a structure-based section that describes the successful application of homology modelling, molecular dynamics (MD) simulation and molecular design methods for the analysis of ion channels. This section includes the analysis of ion conductance, different conformational and functional states, binding sites and ligand interactions, and the influence of mutations on ion channel function. We will also give a basic introduction and point out potential pitfalls of these methods.



Fig. 1. Exemplary illustration of important ion channel features using a K_{ca} 3.1 channel cryo-EM structure (pdb 6cnn [3]). a) A tetrameric ion channel (each subunit is coloured differently). b) Overview about the important feature of one subunit: Helix S5/S6 form the ion channel pore surrounded by membrane-embedded helices S1-S4. Helices HA/HB build the binding site for calmodulin which is important for activation. The channel gate regulates ion conductance, and the selectivity filter is responsible for ion specificity. c) Surface representation with clipped surface showing the inner channel pore.

Cellular Physiology	Cell Physiol Biochem 2021;55(S3):14-45				
and Biochemistry	DOI: 10.33594/000000336 Published online: 3 March 2021	© 2021 The Author(s). Published by Cell Physiol Biochem Press GmbH&Co. KG			
	Menke et al.: Computational Ion Channel Research				

Artificial Intelligence in ion channel research

In the last decade, much progress has been made with regards to the application of artificial intelligence (AI) in image recognition and natural language processing [5, 6]. Although AI has already arrived in everyday life through applications such as voice recognition in mobile phones (e.g. Apple's Siri), the term AI probably only became known to the general public after the AlphaGo software won against the world's best professional go player [7].

This progress and the recent improvements have also reached natural sciences. For medicinal chemistry, AI is expected to be a game changer in the drug design and development process through a combination of and creative collaboration between the "mind and machine" [8]. Taking a closer look, artificial intelligence methods already play an important role in several different areas of drug design such as ligand- and structure-based screening or ADMETox (Adsorption, Distribution, Metabolism, Excretion and Toxicity) prediction, but also in the de novo design of new compounds or retro-synthesis prediction AI is applied [9]. More recently, an antibiotic was discovered using AI techniques [10]. Slowly, the first AI applications in the field of ion channels are also arriving and will be discussed in the context of this section.

The main driver of the success of AI is the subfield of Machine Learning (ML). As the name suggests, Machine Learning describes algorithms that are able to "learn" patterns from a given dataset, often in an iterative manner. This "knowledge" can be used to predict or classify data points not included in the original dataset. The distinction between ML and traditional statistical methods is not always a clear cut, as some methods are common in both domains. A general distinction is made by Bzdok et al. [11]. which states that "[s]tatistics draw population inferences from a sample, and machine learning finds generalizable predictive patterns". Two benchmark studies comparing different machine learning algorithms on various chemical data sets show that not only the right choice of algorithm is important for a successful application but also the right kind of input influences the results [12, 13]. This fact is important, as overtime not only the algorithm but also the kind of input for ML methods on ion channels changed. In the following paragraphs, we give a brief introduction into the important machine learning methods and provide an overview of usages of ML for ion-channels. The presented methods will be analysed with a focus on the exact algorithms as well as input used.

The basics of machine learning for ion channel research

As previously stated, the choice of input is as important as the choice of the ML algorithm itself. In the following we briefly cover the input and most frequently used algorithms in more detail.

The reason why so much attention has to be paid to the input data is that ligands, as well as proteins, are not easily convertible to formats that are accessible for typical statistical models [14]. To understand why, one can look at the differences between chemical structures and images. Images are 2D collections of pixels ordered in a rectangular grid. Each pixel has a numeric value associated with it that measures the brightness or colour intensity. Thus, images are already ordered structures with assigned numeric values. Chemical structures, however, are much more complex as they are three-dimensional and flexible. Proteins as well as small molecules can exhibit conformational changes and there is no straightforward way of assigning atoms/amino acids numeric values. Additionally, most algorithms require the data to be represented by a single vector, so the challenge is to convert complex structures into a single line of values. For these reasons, the process of converting the biochemical data into a computer-readable format, sometimes called featurization, is a crucial step in setting up a machine learning model. Especially as the choice of the algorithm limits the choice of possible featurization and vice versa.

Cellular Physiology	Cell Physiol Biochem 2021;55(S3):14-45				
and Biochemistry	DOI: 10.33594/000000336 Published online: 3 March 2021	© 2021 The Author(s). Published by Cell Physiol Biochem Press GmbH&Co. KG			
	Menke et al.: Computational Ion Channel Research				

Support Vector Machines

Support vector machines (SVM) are a relatively simple, yet powerful method suited for classification. SVMs aim to draw the most optimal decision boundary between two classes. This can be imagined as a (hyper) plane between the datapoints that separates both classes even in a high-dimensional space (see Fig. 2a). 'Most optimal' refers to the fact that SVMs maximize the distance between the datapoints and the decision boundary. Initial SVMs were only able to create linear decision boundaries, until Boser and colleagues [15] used the so-called "Kernel-Trick" to allow for non-linear boundaries.

Random Forest

Random Forest (RF) [16, 17] is another machine learning algorithm, that can provide good results with relatively little training and tuning. The approach is based on decision trees. Decision trees combine decisions in a tree-like structure that allow the model to separate different datapoints based on specific properties. RFs work by combining the prediction of many different decision trees and where each decision tree is build based on a slightly different dataset. For this, a random set of datapoints is removed from the original dataset as well as a random set of input variables before building a new decision tree. This ensures that each tree has different data to work with and hence is built differently. To make the final predictions the mean over all predictions made by each decision tree is calculated (see Fig. 2b).

Neural Networks and Deep Learning

While neural networks have attracted much attention in recent years they have been in use for decades for chemical applications [18]. Neural Networks are based on the idea of neurons in the brain. Neurons can receive input of many neurons and combine them into a single output, which in return can be the input for another neuron. Artificial neural networks are built out of layers, each made up of neurons (see Fig. 2c). Each neuron is, in most cases, connected to all neurons of the next layer and forwards the input information to these next neurons. The initial layer is called the input layer, in which the original input is fed to. The input is then passed to the next layer called the hidden layer, lastly the hidden layer passes their output to the output layer. The output layer is the layer in which the prediction is made.

In artificial neural networks, so-called weights manage how inputs are passed from neuron to neuron. The right set of weights allow the neural network to make accurate predictions. However, these weights first have to be learned and cannot simply be derived. The process of "learning" weights (training) is often much more complex and time-



Fig. 2. Visual representation of different machine learning algorithms. (a) Support Vector Machines aim to draw an optimal decision boundary between two categories. (b) Random Forest models build multiple decision trees based on a subset of variables and data. This average of the decision trees is used as the final prediction. (c) Neural Networks transform the input as it passes through the hidden layers. These transformations should allow the network to make accurate classification in the output layer.

Cellular Physiology	Cell Physiol Biochem 2021;55(S3):14-45				
and Biochemistry	DOI: 10.33594/000000336 Published online: 3 March 2021	© 2021 The Author(s). Published by Cell Physiol Biochem Press GmbH&Co. KG			
/	Menke et al.: Computational Ion Channel Research				

consuming than the training of random forest classifiers or SVMs. Deep Learning refers to a neural network which has multiple hidden layers. These additional layers make the model even harder to train but tend to produce better results [19].

Convolutional Neural Networks (CNN) are a special kind of neural networks and were initially developed for the field of image processing. Loosely speaking, CNNs run multiple filters over the image/matrix and can identify important patterns. The advantage of a CNN is that it can deal with 2D-(e.g. images) and 3D inputs, while regular neural networks are only able to use 1D inputs. Long-short term memory (LSTM) networks are another specialized neural network. LSTMs originate from the field of natural language processing and belong to the class of Recurrent Neural Networks. Their advantage is that they are able to process input and taking the order of the input into account as a sequence of data. As an example, while processing a specific word in a sentence, the LSTM is able to take into account what was said earlier in the sentence, so a word can be processed differently depending on what was said before. More specific to biochemistry, LSTMs could process amino acids differently based on the surrounding amino acids.

The application of machine learning in ion-channel research can broadly be summarized into two categories. One is concerned with the prediction of functionality and topology of ion-channels, the second is concerned with the quantitative structure-activity relationship (QSAR) prediction, which aims to predict the activity of a given ligand on one or multiple targets.

Functionality Prediction

One application of machine learning is the identification and classification of ion channels based on their amino acid sequence, which is an important feature for the analysis of new and unknown sequences. The earliest application of machine learning in this context was described by Liu et al. in 2006 [20]. The goal was to classify five different types of voltage-gated potassium-channels purely based on their amino acid sequence. Here, known sequences were featurized using a dipeptide composition that encodes the relative frequency of dipeptides in a protein. They used SVMs, even though SVMS can only be used for binary classification. This means they can only be trained to distinguish between two classes of ion-channels. For that reason, Liu et al. trained five different SVMs, one for each channel type. Their approach was quite successfully classifying almost all channels correctly.

In the same year, VGIchan was released by Saha et al. [21] together with a still available webserver. They shifted the attention to differentiate between different voltage-gated ion channels (potassium, sodium, calcium, chloride). They used a SVM and Dipeptide Composition together with *HMMER* [22] generated profiles for the four types of voltage-gated ion-channels. HMMER is a tool able to search for sequence homologs based on profiles of multiple sequences based on Hidden Markov Model (HMM). Whenever HMMER failed to make a prediction, the SVM predictions were used for a final assessment. Overall an accuracy of 97.78 % was achieved.

In 2011 Lin et al. [23] extended the existing method to not only classify the type of voltage-gated ion channels but to predict whether a given sequence is an ion-channel or not. If an ion channel is predicted, the next step is the prediction whether it is a ligand or a voltage-dependent ion channel, and what type of voltage-dependent ion channel it is. While the choice of SVM was made arbitrarily before, Lin.et al. compared different algorithms and found the SVM to work best. Major changes were introduced by Gao et al. in 2016 with the introduction of PSIONplus [24]. Based on the comparisons made by Lin et al. [23] in 2011 they stuck to an SVM but largely extended the features used as input. Apart from the dipeptide composition, various physiochemical properties, the predicted relative solvent accessibility, and information on secondary structures were used. Additionally, a PSI-Blast [25] approach was implemented to generate the position-specific scoring matrix (PSSM), which is then converted into a feature vector. Their method outperformed the general BLAST

Cellular Physiology	Cell Physiol Biochem 2021;55(S3):14-45				
and Biochemistry	DOI: 10.33594/000000336 © 2021 The Author(s). Published by Published online: 3 March 2021 Cell Physiol Biochem Press GmbH&Co. KG				
	Menke et al Computational Ion Channel Research				

[26] and also previous models such as the VGIchan [20] and the model described by Lin et al. 2011. Simultaneously, Tiwari et al. proposed a novel arguably more efficient method [27]. While the results are not directly comparable, their model was able to predict ion-channels and their subtype with high accuracy and without using any calculation using BLAST. For their models it appeared that the Random Forest classifiers perform better than SVMs. In 2017, IonPredv2.0 by Zhao et al. [28] extended the original work of Lin et al. with a novel pseudo dipeptide composition which led to improvements in performance in comparison to the original paper.

Unfortunately, there were two issues with the methods proposed at that time: (1) Many publications did not provide code or a web server which allows researchers to use and compare methods, and (2) there was no benchmarking dataset which allows objective comparison of different methods. This changed in 2019 with a review published by Gao et al. [29]. They compared the performance of three published methods: VGIchan [21], PSIONplus [24], and IonchanPred 2.0 [28]. They found that IonChanPred 2.0 is the overall best performing model. More importantly, they showed that all models performed worse on the benchmarking dataset than on the datasets which they were trained on in their respective original work. In the same year, two additional models were released: One by Han et al. [30] and a second called DeepIon by Taju et al. [31]. Rather than classifying subtypes of ion-channels, DeepIon focused on distinguishing ion channels, Ion Transporters, and other membrane proteins. The approach proved to be successful to differentiate between the described categories. This work is especially noteworthy as the applied method is completely new in this context. It uses one of the newly described deep learning/artificial intelligence methods: A Convolution Neural Network (CNN). All models up to now could only take vectors (1D) as input. For example, the PSSM in PSIONplus was converted into a vector by summing all rows of the same amino acid and concatenating those rows. DeepIon does not require this concatenation and can apply the CNN directly on the available 2D Feature Matrix. A second advantage is that (convolutional) neural networks can integrate all prediction tasks into one model and allow for multitask and multi-target predictions. This contrasts with the used SVMs that are only able to be used for a binary classification. Multiple SVMs have therefore be trained to overcome this issue.

The most recent work by Gao in 2020 [32] describes an extension of their original model PSIONplus [24] from 2016. They addressed many problems that burdened earlier models. For ones, the prediction is also extended to the subtypes of ligand-gated channels. Second, sequential multi-label (or multitask) classification is made possible by combining predictions from higher-level models with the ones of lower-level ones. This enables the classification of ion channels that belong to more than one subtype of ion-channel. These changes lead to an overall more powerful model, providing better results in less time.

Transmembrane topology prediction

A second use case for machine learning methods in the context of ion channels is topology prediction. In general, the topology of a protein refers to the overall folded 3D structure and the connected secondary structure elements. In case of ion channels, topology predictions are used to predict the ion channel domain that lies within the membrane. This is possible since most transmembrane ion channel domains are build-up of helices that show different amino acid compositions compared to helices lying outside of a membrane. Early models of transmembrane protein topology were simple but with the increasing number of available crystal structures, the models became increasingly complex. Membrane-embedded helices "can be short, long, kinked or interrupted in the middle of the membrane, they can cross the membrane at oblique angles, lie flat on the surface of the membrane, or even span only a part of the membrane and then turn back, forming so-called re-entrant loops" [33]. The first attempts to predict secondary structures within membrane proteins date back to 1982 [34] and many improvements have been made over the last 34 years. Due to the huge amount

Cellular Physiology	Cell Physiol Biochem 2021;55(S3):14-45				
	DOI: 10.33594/00000336	© 2021 The Author(s). Published by			
and Biochemistry	Published online: 3 March 2021	Cell Physiol Biochem Press GmbH&Co. KG			
•	Menke et al.: Computational Ion Channel Research				

of work described, we will strictly focus on ML algorithms used for such predictions. For a complete overview of topology predictions of transmembrane proteins, we refer to a review from 2017 by Almeida et al. [35].

In contrast to function prediction, neural networks were utilized relatively early for secondary structure predictions and were later used for topology predictions of transmembrane proteins. In 1993 Fariselli et al. [36] published their work on predicting α -helix and β -sheet segments using a neural network in combination with a single protein sequence as input. A segment of 17 amino acids was used as input for the network where each amino acid in this sequence was encoded as a binary vector of length 20 (one position per amino acid). Therefore, the final input vector had a length of 340 bits (17 x 20). Burkhard Rost and Chris Sander continued to improve and expand this approach [37–40]. In 1995 they showed that using profiles obtained from multiple alignments improves the performance of their neural network. This neural network is also one of the first that is used to localize transmembrane helices in a given protein sequence. In the early 2000s, HMMs were a popular choice for topology predictions [41–45], but Martelli et al. used them in combination with a neural network [46]. In 2003, they introduced their approach called ENSEMBLE focusing on all- α transmembrane proteins. As their name suggests, they used an ensemble of a single neural network and two Hidden Markov Models. Their average predictions are used to identify the transmembrane segments.

In 2007, David T. Jones proposed a new model called MEMSAT3, which solves the issue of having to use rule-based prediction for C- and N-terminals [47]. As the name suggests, it builds on their previous work starting from 1994 where MEMSAT was first introduced [48]. It classified residues into 5 different classes based on probabilities obtained from curated membrane protein data. MEMSAT2 [49] extended the models by using sequence profiles. MEMSAT3 starts out with a neural network that uses the information of residues based on the Position-Specific Scoring Matrix (PSSM). The output of the Neural Network is then used as input for the MEMSAT algorithm. The neural network used by Jones can directly predict signal-peptides, rendering ad-hoc rules used by previous models obsolete.

MemBrain [50] was introduced in 2008 and uses a more sophisticated version of the k-nearest-neighbour (kNN) algorithm with features obtained through the PSSM. kNN algorithms classify data by assuming that similar objects should belong to the same class. When deciding for a new object to which class it belongs, the models analyse to which class its closest neighbours belong and the majority determines the class of the new object. The more sophisticated algorithm used by MemBrain, among other things, extends the kNN to a multilabel classification task. With this newly proposed method, MemBrain was providing the best classification prediction up to this point. Later the year, OCTOPUS [51] was introduced. Multiple Neural Networks generate enriched input from traditional sources such as the PSSM. These features were then fed into an HMM. What makes this approach unique is that it can classify helices which re-enter or do not completely cross the membrane. Later in the year, SPOCTOPU [52] was published an extension which allows for the classification of signal peptides.

In 2009, TOPCONS [53] combined five published models for topology predictions to create a more powerful combined predictor based on the consensus of the individual models. TOPCONS aggregates the output of every single model and uses them for their newly trained HMM. While TOPCONS was only slightly better than the individual models it provided reliability scores per prediction, which measured how certain the prediction is. In 2015 an updated TOPCONS version was released [54]. The new TOPCONS uses updated models with increased speed and other user experience improvements. With the new model, the overall accuracy could be improved from 83% to 87%. Next to the updated TOPCONS model, a new Consensus mode (CCTOP) [55] was introduced. It uses predictions from 10 different models and utilized them as constraints in a HMM. Unique about this approach is that it allows the user to define constraints and weights the inputs from the different models depending on their accuracy. CCTOP performed better than any of the single models as well as previous consensus models such as TOPCONS.

Cellular Physiology	Cell Physiol Biochem 2021;55(S3):14-45				
	DOI: 10.33594/00000336	© 2021 The Author(s). Published by			
and Biochemistry	Published online: 3 March 2021	Cell Physiol Biochem Press GmbH&Co. KG			
	Menke et al. Computational Ion Channel Research				

A tool using more sophisticated neural networks is DMCTOP [56]. It is based on a multi-scale CNN model trained to predict the topology. The multi-scale approach allows the network to detect dependencies across various ranges resulting in a prediction accuracy that is superior to previously reported methods. Lastly, an updated MemBrain version [57] was released. It uses an ensemble method of two CNNs, one to process the complete sequence and one processing a sliding window of 17 residues. Additionally, it uses two SVMs aiming to predict the N- and C-terminus. The SVMs are also used to predict the orientation of the helices. Fig. 3 shows a MemBrain prediction exemplary on two transmembrane proteins.

Most methods introduced only focus on the predictions of α -helices and β -barrels are often neglected. The decision against this prediction is often made because α -helices are more abundant, and β -barrels are mostly occurring in prokaryotic cells [61]. As the methods do not necessarily diverge greatly, we will only provide a quick overview for β -barrels. HMM are the models of choice for the topological predictions [62–65]. and early benchmarking show that they are the best performing models [66].



Fig. 3. Transmembrane helix prediction using MemBrain [57] for a human C5a anaphylatoxin chemotactic receptor 1 (C5aR, a-d) and a hyperpolarization-activated cyclic nucleotide-gated ion channel (HCN-1, e-h). Structural overview of a) C5aR (pdb 5o9h[58]) and e) a HCN-1 subunit (pdb 5u6o[59]), b/f) Transmembrane Helix Propensity, c/g) predicted topology and d/h) sequence information about secondary structure and membrane regions taken from the corresponding pdb entries (www.rcsb.org.[60]) extended with information about MemBrain predictions.

Cell Physiol Biochem 2021;55(S3):14-45 DOI: 10.33594/00000336 Published online: 3 March 2021 Cell Physiol Biochem Press GmbH&Co. KG Menke et al.: Computational Ion Channel Research

QSAR

Quantitative structure-activity relationship describes mathematical models that determine a relation between a compound and their activity on specific proteins [67]. Molecules are often described in terms of molecular descriptors to establish these mathematical models and the models then relate these descriptors to the activity of the molecule [68]. Molecular descriptors are numeric representations which describe certain (bio)chemical or physical properties of molecules. A special class of descriptors are molecular fingerprints, which aim to represent molecules and their substructures in a vector. By selecting the appropriate descriptors in combination with the right model it is possible to predict the activity of molecules on specific proteins. For this, a database of compounds with known activities and/or inactivities on a specific protein target is required. By statistical methods, like partial least square regression or machine learning methods, a combination of molecular descriptors is identified that is best suited to predict affinity. For selecting the best QSAR model and calculating statistical parameters for validation, a test dataset of previously unseen molecules is used. A 2nd dataset, the external validation dataset, is applied on the final selected QSAR model for assessing the performance of the QSAR model. A lot can go wrong when creating a QSAR model, and the inexperienced reader is advised to read the following manuscripts on best practice: One by Cherkasov et al. [67] and the other by Alexander Trophsa [69]. The final aim of a OSAR model is to predict if and how a given compound will be active on a specific protein or not.

Early uses of neural networks for OSAR date back to the 1990s [70–72]. However, these models were tiny compared to today's standards. One of the first neural networks used as a QSAR model for an ion-channel was trained with 57 compounds and their network used 8 hidden nodes [73]. Thanks to hardware improvements, new algorithms and increasing availability of data, models are much larger nowadays. The ion-channel that has attracted the most attention with regards to OSAR is the hERG channel, since it is an off-target and responsible for severe side-effects of potential drugs. The early machine learning models for hERG activity predictions are well described by Anthony Klon in 2010[74]. From this review, it becomes apparent that a wide variety of ML algorithms are utilized with the already mentioned SVMs, Neural Networks, and the Random Forest algorithm being the most popular. It can also be seen that the available data has increased over time. Models in the 90s were built on small datasets, the models introduced later were trained on at least 300 compounds. However, not all models are actually trained to predict the binding affinity of compounds to hERG. Rather they perform a binary classification (active or not), this allows some models to not be trained on hERG activities directly but on dofetilide displacement [75] or more general torsades de points (TdP) measures [76, 77].

Many models use as input a variety of descriptors and molecular fingerprints that describe the physicochemical properties of the input molecules. These input feature vectors are often much bigger than what is used in the prediction of topology or functionality. Since QSAR models do not necessarily have to take into account the protein structure, they do not need to deal with sequential data (e.g. amino acid sequence), making it easier to generate features. Free software exists which can quickly calculate up to 1800 different descriptors for every single molecule in the dataset [78]. Commercial software packages offer more than 4000 descriptors [79].

In an article from 2014 by Braga and colleagues [80], it was revealed that most hERG QSAR Models so far do not comply with published good-practice rules as well as the OECD guidelines [81] for QSAR Models. Next to insufficient predictive power, many models do not pass the Y-scrambling test. The dependent variable that the QSAR Model is supposed to predict is thereby randomly scrambled. Thus, the relation of structure to activity is now randomized. A good QSAR model should perform much better on the regular dataset than the Y-scrambled dataset. Lastly, the applicability domain is rarely assessed. One would like to know when the model is reliable and when it is not. Therefore, Braga et al. [82] introduced a model that follows these guidelines. It is a consensus model that combines different machine

Cellular Physiology and Biochemistry Cell Physiol Biochem 2021;55(S3):14-45 DOI: 10.33594/000000336 © 2021 The Author(s). Published by Cell Physiol Biochem Press GmbH&Co. KG Menke et al.: Computational Ion Channel Research

learning algorithms with different inputs. The chosen models are Random Forest Model, SVM, and a Gradient Boosting Machine (GBM). Gradient Boosting Machines, like RF models, rely on Decision Trees. They provided in their work the hERG liability prediction, assess whether the compound is within the applicability domain, and also provided an overview of which substructures of the molecule are responsible for the prediction.

A new model was developed in 2019 by Konda et al. [83]. They also acknowledged the same flaws of previous models and built a consensus model. They started out by computing a vast number of descriptors and fingerprints, leading to a combined input of 16.000 bits. With extensive variable selection, the optimal set of descriptors and models were chosen. They compared their model on three external validation sets and found that they performed better than previous hERG QSAR models. Unfortunately, they did not provide the code or a web server to make use of their model. Lastly, a QSAR by Siramshetty et al. [84] analysed the challenges that one can come across when building a QSAR model for hERG. Especially, the data from public databases causes an issue as activity measures are quite heterogeneous. Further, they showed that the choice of fingerprint for their models did not make a significant difference. However, the activity cut-off value chosen has a strong impact on the performance of the QSAR model. The cut-off value determines which compounds are considered active and which inactive based on the activity of that compound. This converts the problem of predicting the exact activity measures for compounds to a binary classification task (active vs. inactive).

Besides hERG, QSAR models also exist for other channels. The voltage-gated sodium ion channel 1.5 ($Na_v 1.5$) is another ion channel also related to TdP. Khalfia et al. [85] built a QSAR model for this channel using a variety of machine learning models and found a gradient boosting machine to work best. They followed OECD guidelines and obtained statistical accuracy above 0.8. A different paper considered with TdP argues that QSAR models specific to a single channel are not able to take into account the multi-channel effect and thus not capable of making an accurate assessment of cardiotoxicity. They focus on a model that classifies torsadogenic drugs using an SVM [86].

Another target is the voltage-gated sodium channel 1.7 ($Na_v 1.7$), involved in the pain generation [87]. Kong et al. [88] trained many different ML models to identify $Na_v 1.7$ inhibitors. They showed that an RF with a CDK [89] fingerprint delivered the best performance in discriminating actives from inactives. More interestingly they also compared their model to a Graph Neural Network (sometimes called Graph Convolutions Network). Because GNNs are thought to be well suited to handle molecular structures without having to compute a vectoral representation, a recent surge of their application in chemistry has been observed [90]. However, Kong et al. did not see any benefits from their GNN with regards to prediction quality. Similar results were also found for a hERG QSAR model which did not benefit from the usage of GNNs [91].

Additionally, Kong and colleagues investigated whether a fingerprint obtained from an autoencoder could be a better-suited input than traditional fingerprints. Autoencoders are neural networks that aim to first encode the input into a single dense vector, and then in a second step reconstruct the original input from the encoded vector. These autoencoders can be trained without labelled data, as the input also represents the required output. The dense vector obtained from the autoencoder can also be used as a fingerprint, which is done by Kong et al. [88]. This, however, did not result in a better performance than traditional fingerprints. They also used the dense vector to optimize the synthetic accessibility and drug-likeness of compounds identified by their models. By changing the dense vector slightly, one can obtain altered molecules. One of the compounds identified by the model was experimentally validated and found to be an actual inhibitor of the $Na_v 1.7$.

Lastly, we introduce two articles that do not focus on ion-channels specifically, but rather include them as a subset in their testing. The first is DeepAffinity [92] by Karimi and colleagues. It aims to predict the binding affinity between a ligand and a protein using both ligand and protein information. For this, they use two RNNs one for the protein sequence and one for the ligand, represented in the SMILES format. Similar to Kong et al., an autoencoder

Cellular Physiology	Cell Physiol Biochem 2021;55(S3):14-45				
and Biochemistry	DOI: 10.33594/000000336 © 2021 The Author(s). Published by Published online: 3 March 2021 Cell Physiol Biochem Press GmbH&Co. KC				
	Menke et al.: Computational Ion Channel Research				

is used to pre-train the model. Pretraining is done if not enough data is available to get well performing weights for the neural network. Pretraining allows the neural network to adapt their weight to the data that is going to be used. This pre-training should already induce some "syntax knowledge" of SMILES and/or protein sequence into the model. Later training for predicting binding-affinity should then be easier. The encoder of the pre-trained autoencoder is used in the actual training of the model. They follow up the RNN with a single CNN layer and later combine the input from the SMILES and protein sequence. While the model performed overall better than baseline models, looking specifically at ion-channels a simpler RF could outperform their model. Most likely this is due to the fact that only 15,000 out of almost 500,000 samples in the dataset were ion channel modulators. So, one would expect that the models perform better for more frequently occurring subclasses. Another proposed method was introduced by Wang et al. [93] in 2020. They combine the ligand with features obtained from the PSSM of the protein as input for an LSTM. This approach outperforms more traditional machine learning models even for ion-channels. Here a bigger proportion of the data were ion channels, although overall their dataset contains less structures than DeepAffinity.

As it can be seen, the complexity of QSAR models increased over time. However, the choice of a model is much more diverse than for topology predictions. This can be attributed to the nature of the predictions: Topology predictions must take into account the order of the amino acid sequence. Regular random forests or neural networks are not capable of doing so, HMMs are for that reason used in most structure-based models. The choice of model is also depending on available features that can be utilized by a model. QSAR models can rely on many sources such as molecular descriptors and molecular fingerprints, which are easily computed. Something what topology or functionality predictions cannot rely on. Lastly, much more activity data is available than protein structures, hence more complex models can be trained based on abundance of data.

Misc

In this section, we introduce two recent papers which do not fit into any of the previously mentioned categories. The first automates the event-detection for patch-clamp measurements. Patch-clamp techniques aim to quantify the dysregulation of ion-channels. However, the recordings are noisy and must first be cleaned, usually done through human supervision. The aim is to identify if and how many ion channels are open. To automate this cleaning process "Deep-Channel" [94] was developed. It uses a combination of CNN and LSTM to allow for a more effective assessment of long and short-term dependencies. One of the biggest challenges was that little to no labelled data with known ground truth exist. Practitioners performing patch-clamp can only guess from the data how many channels are open, but there is no way to identify how many channels did actually open. To overcome this process, they used a HMM to generate data which is reflective of the closing and opening of ion channels. The trained models were then compared to humans and it was found the model is much faster with comparable accuracy, especially for data in which five ion-channels are measured. Lastly, a work from Rao and colleagues uses an SVM to build a simple model that can be used to identify hydrophobic gates. Hydrophobic gates can block the passing of ion channels even when the pore is not blocked by any steric occlusion. Their SVM uses only the local hydrophobicity and radius of the pore to classify whether this is blocked through by the displacement of water [95].

Cellular Physiology	Y Cell Physiol Biochem 2021;55(S3):14-45 DOI: 10.33594/000000336 © 2021 The Author(s). Published by Published online: 3 March 2021 Cell Physiol Biochem Press GmbH&Co. KG Menke et al.: Computational Ion Channel Research Menke et al.: Computational Ion Channel Research				
and Biochemistry					

Summary and Discussion

We have shown various applications of machine learning projects focusing on ionchannels. While the building of QSAR models is done for all kinds of target classes, the prediction of topology and sub-types is a challenge unique to ion-channels. Especially for these structure-based challenges, models are often simpler than used for other structurebased tasks where different variants of neural networks are more frequently used [96].

One can see a clear improvement for the classification of ion-channels over time. Initial models were only interested in differentiating between specific subtypes of a specific ionchannel. Today's models are able to recognize ion channels, identify whether they are ligand- or voltage-gated and additionally can also identify the specific subtype. However, this success is not due to more advanced models but rather an influx of available data and some better feature generation. The most recent model PSIONplusm builds on SVMs just like early models from 2006.

However, while the SVM is easy to train with relatively little data, it also has drawbacks like the binary classification problem. This forces an approach to train different SVMs to classify ion-channels completely. In the latest edition of PSIONplus^m, the authors overcame this issue, but other algorithms would provide a more straightforward solution for the issue. Traditional Neural Networks are often used for multi-task challenges and have proven to be quite successful [97–99]. However, neural networks require more data than SVMs or random forest to be trained successfully. While the available data increased over the last years, it might be not enough for traditional neural networks. Even for SVMs Gao et al. (2020) states that improvements can only be made with more labelled data [32].

Topology predictions are much more complex than ion-channel classification. Rather than having to make a single, global prediction for the whole sequence, topology prediction needs to identify segments within each sequence. Further, topology prediction can focus on tasks with increasing complexity. Early models only attempted to identify transmembrane segments, while recent models aim to identify signal peptides, re-entrances, and orientation of helices. Featurization also differs between the two tasks. As ion-channel classification only requires global predictions, features that are computed across the complete sequence are often used. Many models are using amino acid composition or di/tripeptide compositions and beyond that physicochemical properties are computed based on the complete sequence. Some models include some form of amino acid composition as their input for topology predictions but most of the time features extracted from PSSM profiles are used. The differences can also be seen in the choice of model. A frequently used algorithm for topology predictions is the HMM. Hidden Markov Models are well-suited for such tasks as they can take into account the sequence in which the amino acids are ordered. More recent models rely on LSTM or multi-scale CNN which also have similar capabilities. These models are much more complex than the ones used in the ion-channel classification. This can be achieved as more training data is available. As predictions are made for each residue in a sequence, models have "more opportunities" to learn. Ion-channel predictions do not focus on a single residue. Here the dominant choice is the SVM. It is much simpler and cannot take into account the sequence ordering. This is also not necessary as most features used are global ones.

Links to all software discussed so far can be found in Table 1. Our recommendation for functionality prediction would be to use PSIONplusm by Gao et al. It is the most recent model which offers the most in-depth prediction while providing arguably the most accurate predictions. For topology predictions, not as clear. In the paper introducing the third iteration of MemBrain, it appeared that this version provides the most accurate predictions. Additionally, it can perform predictions for orientation which other software cannot do. However, other webservers provide almost similar performance and, in some scenarios, even better. These include MEMSAT-SVM and CCTOP. As all of these can be accessed easily via a web browser, we believe it is advisable to use all of them and compare the results.

For QSAR modelling of ion channel modulator activity, a more diverse set of algorithms is used. It was shown that more complex models such as GNNs or RNN+CNN combinations

Cellular Physiology	Cell Physiol Biochem 2021;55(S3):14-45				
and Biochemistry	DOI: 10.33594/000000336 Published online: 3 March 2021	© 2021 The Author(s). Published by Cell Physiol Biochem Press GmbH&Co. KG			
	Menke et al.: Computational Ion Channel Research				

Table	1.	Online	Tools	using	Artificial	Intelligence	specific to	Ion-channels.	Only	webservers	which	were
access	sible	e at poi	nt of p	oublicat	tion are n	nentioned						

Name	Link to Website
Ion Channel Classification	
VGIchan [21]	http://crdd.osdd.net/raghava/vgichan/
PSIONplus [24]	https://sourceforge.net/projects/psion/
IonchanPred v2.0 [28]	http://lin-group.cn/server/IonchanPredv2.0/webServer.html
DeepIon [31]	http://biologydeep.com/deepion/app/identification/
PSIONplusm [32]	https://yanglab.nankai.edu.cn/PSIONplusm
Topology Predictions	
HMMTOP [42]	http://www.enzim.hu/hmmtop/index.php
Phobius [43]	https://phobius.sbc.su.se/
OCTOPUS[51] /SPOCTOPUS [52]	https://octopus.cbr.su.se/
TOPCONS [53]	https://old.topcons.net/
TOPCONS (updated) [54]	https://topcons.cbr.su.se/pred/
CCTOP [55]	http://cctop.enzim.ttk.mta.hu/?_=
DMCTOP [56]	https://github.com/NENUBioCompute/DMCTOP
MemBrain [57]	http://www.csbio.sjtu.edu.cn/bioinf/MemBrain/
QSAR	
Pred Herg 2.0 [82]	http://predherg.labmol.com.br/

did not always lead to improved model performance. The biggest issue that ion-channel faces with regards to artificial intelligence is the data or the lack thereof. Complex models which worked well for many target classes did not produce the desired results for ion-channels. One way to fix the reliance on an increased volume of data is to utilize transfer learning, multi-task and pre-training [98, 100–103]. These methods allow the researcher to utilize the power of larger datasets for more specific challenges and could help build more sophisticated models for ion-channels.

Computational approaches for structure-based analysis of ion channels

The availability of structural information on ion channels is important in order to get an in-depth insight into their function and how mutations lead to dysfunctional ion channels that are responsible for severe channelopathies. As a matter of fact, no ion channel structure was known before 1998 when Doyle et al. [104] described for the first time a crystal structure of a potassium channel from *Streptomyces lividans* with a resolution of 3.2 angstroms (see Fig. 4a/b). This structure contains a selectivity pore of 12Å long where negatively charged moieties are pointing towards the inside of the pore to balance the positive charges of the potassium ions. Although the crystallisation of membrane bound proteins for structure determination is extremely difficult and the experimental needs demanding [105], numerous three-dimensional structures of ion channels are available nowadays. However, the number is small compared to other transmembrane families such as G-protein coupled receptors (GPCRs). The first structure determination of a TRP ion channel (see Fig. 4c/d) without any crystallisation via cryogenic electron microscopy (cryo-EM) in 2013 [106] has had a massive impact on the field of ion channel structure determination. Although X-ray crystallography is still the method of choice for a detailed and high-resolution overview on ion channels, cryogenic electron microscopy (cryo-EM) is an emerging field that has raised hopes of gaining access to more ion channel structures. An overview of cryo-EM can be found in the paper by Nygaard and colleagues [107].





Fig. 4. The first ion channel x-ray and cryo-EM structures. a) cartoon representation and b) surface representation (one subunit hidden, magenta spheres: K⁺) of the first x-ray structure (pdb 1bl8 [104]). c) cartoon representation and d) surface representation of the first cryo-EM structure (pdb 3j5p [106]).

Due to the limited number of available high-resolution X-ray structures and the low resolution of cryo-EM structures, computational methods are of utmost importance for the detailed analysis of ion channel (dys)function, ligand-binding and the development of drugs to cure channelopathies. Many researchers have applied computational methods to this day and achieved major accomplishments in the field. In the following paragraph, we will give a brief introduction into important structure-based computational methods used in ion channel research. Subsequently successful applications of these methods are discussed, leading to a better understanding of ion channels. Such computational methods include homology modelling, molecular docking and molecular dynamics (MD) simulations.

The basics of structure-based computational methods

Homology modelling

Homology modelling or comparative protein modelling is being applied for decades now [108]. Although the methods used improved over the years [109] the underlying workflow remains the same (Fig. 5) [110]. The most important step is the initial template selection of a known protein structure most suited to build the basis for modelling. This is done by a pairwise sequence comparison of the target sequence to all known x-ray structures

available in the protein data bank (PDB, www. wwpdb.org/) [111] It is expected that below a sequence identity of 25% it is difficult to create a successful homology model. The next step is the alignment of the target sequence and the sequence of the selected template. It is advisable to not rely on a simple pairwise sequence alignment but use multiple sequence alignments or more sophisticated and specialized alignment methods. The model building starts by taking the 3D backbone of the template structure, followed by the modelling of gaps and missing loop regions. Finally, the sidechains of the target structure are reconstructed. Afterwards, the models are optimised via energy minimisation and validated based on different quality assessment like sidechain clashes or unlikely backbone or sidechain torsion angles. A good webserver for quality assessment is the Continuous Automated Model Evaluation



Fig. 5. Homology modelling workflow.

Cellular Physiology	Cell Physiol Biochem 2021;55(S3):14-45	
and Biochemistry	DOI: 10.33594/000000336 Published online: 3 March 2021	© 2021 The Author(s). Published by Cell Physiol Biochem Press GmbH&Co. KG
	Menke et al.: Computational Ion Channel Research	

project (CAMEO, www.cameo3d.org) [112] Here regular quality assessments of the registered webservers are reported. Potential inaccuracies are also discussed by Haddad et al. [110]: Inappropriate template selection, errors in the final target-template alignment, problematic sidechain packing or problematic loop modelling due to local differences between target and template structure.

A good starting point for the interested reader are "the ten quick tips for homology modelling" described by Haddad et al. [110], and the results of the regular "Critical Assessment of Techniques for Protein Structure Prediction (CASP) challenges" [113]. Table 2 shows a summary of easy to use and successful webserver for homology modelling. In addition, Modeller [114] (https://salilab.org/modeller/) is a well-known free-for-academics software tool that allows local modelling creation.

Molecular Dynamics Simulations

The basic idea behind molecular dynamics (MD) simulations is the calculation of the position of each atom of a system as a function of time, based on Newton's laws of motion [121]. The basis is a calculation of the forces acting on each atom and the subsequent updating of the position and velocity. These calculations are repeated on a very short time scale leading to a trajectory of protein dynamics over time. This trajectory describes the internal motions and the resulting conformational changes. The first MD simulation of a protein was performed in 1977 with the simulation of a globular protein (bovine pancreatic trypsin inhibitor) [122].

Meanwhile, MD simulations have become the most popular method to study and draw a chronological series of events characterizing the structural and functional aspects of protein behaviour with impact on molecular biology and drug discovery [123]. Indeed, this method brought the understanding of molecular systems, such as ion channels, to another level, picturizing then a succession of conformational changes from the opening and closing, the ion permeation, the external stimuli and the binding of modulators. In contrast to homology modelling with available web-servers, the application of molecular dynamics simulations still needs experiences and in-depth knowledge about the underlying mechanism. This is especially true for ion channels, where the structures have to be embedded in a membrane for a realistic simulation. The most widely applied software packages are described in Table 3.

Computational molecular design and virtual screening

After the target selection step, either by selecting a protein structure from the Protein Data Bank or a homologous model built starting from a template protein, computational molecular design or structure-

lecular design or structurebased virtual screening can be applied for the identification of binding sites of known ligands or for the structure-based identification of new ion channel inhibitors. The most popular method in structure-based design is molecular docking, where the binding mode of a small molecule is predicted within a protein binding site. In the early 1890s, Emil Fischer was the first to introduce the idea of the conventional "key in a lock" model [124], he stated that

Table 2. Well-known homology modelling servers

Name	Link to Website
I-Tasser [115]	https://zhanglab.ccmb.med.umich.edu/I-TASSER/
Phyre2 [116]	http://www.sbg.bio.ic.ac.uk/phyre2/index.cgi
PSIPRED [117]	http://bioinf.cs.ucl.ac.uk/psipred/
RaptorX [118]	http://raptorx.uchicago.edu/
Robetta [119]	http://www.robetta.org/
SWISS-MODEL [120]	https://swissmodel.expasy.org/

Table 3. Well-known software packages for applying MD simulations

Name	Homepage	Cost
Amber	https://ambermd.org/	low cost for academics
CHARMM	https://www.charmm.org/	low cost for academics
GROMACS	http://www.gromacs.org	free
NAMD	http://www.ks.uiuc.edu/Research/namd/	free for academics
YASARA	http://www.yasara.org/	low cost for academics

Cellular Physiology	Cell Physiol Biochem 2021;55(S3):14-45	
and Biochemistry	DOI: 10.33594/000000336 Published online: 3 March 2021	© 2021 The Author(s). Published by Cell Physiol Biochem Press GmbH&Co. KG
,	Menke et al.: Computational Ion Channel Research	

Name	Degree of flexibility	Cost
DOCK [127]	partial protein flexibility	free of charge for academics
GOLD [128]	partial protein flexibility	available at most university due to an often existing academic licence of the Cambridge Structural Database
AutoDock [129] and Autodock Vina [130]	partial protein flexibility	open source
Glide [131]	flexibility of side chains of protein	Cost associated, part of Schrödinger modelling suite

Table 4. The most popular applied dockin	g software packages
--	---------------------

a unique conformation of a ligand is necessary to fit and generate a complementarity unique binding site conformation of the target. Based on this idea, molecular docking was the first time described by Kuntz et al. in 1982 [125]. They developed and tested a set of algorithms that can explore the binding of a ligand to a receptor of a known rigid structure. The conformational space of a ligand is sampled either before placing these rigid conformations into a binding site (rigid docking), or the conformations are adapted while placing the conformations (flexible docking). Although the "key and lock" model introduced earlier was useful to predict and identify favourable binding modes, major problems were reported. This model implies a rigidity of the catalytic site, thus lacking degrees of freedom. It also does not take into account the water moieties, which are in fact important molecules in many cases for the binding process. Nowadays, it is more precise to use flexible docking approaches, either semi-flexible docking approach, where usually the ligand is flexible and protein rigid, or a flexible docking where both the ligand and partly the protein are flexible. It is then no longer a "key and lock "model but a "hand and glove" model [126], here bond rotations, sidechains rotations are sampled in order to minimize the free energy and avoid steric clashes.

The most popular applied docking software packages are described in Table 4.

Ion conductance

Not long after the publication of the K⁺ channel crystal structure [104], scientists took advantage of the increasing power of computers by starting the first biomolecular simulations of that ion channel. Berneche et al. inaugurated the simulation of *KcsA* K⁺ channel in 2000 [132]. The article unveiled a stabilization of the potassium ions with a water-filled cavity in the center of that transmembrane protein. Right before, Guidoni et al. attempted a molecular dynamic simulation of the same protein permeated by sodium and potassium ions [133]. This article introduced the cation selective penetration of this protein, thus repulsing anions. They also revealed a salt bridge between Asp80 with Arg89 which is a key stabilizing point for that ion channel. Few months later, it was Allen et al. who described a molecular dynamic simulation of *KcsA* potassium channel, noticing a stabilization of the system with a double potassium ions [134]. With the increasing computational power more detailed analysis are possible, also with respect to ion conductance. In the following part, we will discuss some interesting analyses regarding ion conductance using MD simulations.

The availability of an open-conformation of a pore-only construct of a bacterial Na_vM sodium channel (pdb 4f4l [135]) allowed Ulmschneider et al. to perform detailed analysis of ion passage [136]. The pore was embedded into a lipid bilayer and subjected to MD simulations. Due to missing voltage-sensing domains, restraints were applied to keep the pore open. In addition, an electric field was needed to drive the ions through the channel. The simulation revealed five distinct ion binding sites and a sodium conductance of ~33pS, whereas the latter is in agreement with experimental validation. The pore is selective for

Cellular Physiology	Cell Physiol Biochem 2021;5	5(S3):14-45	
and Biochemistry	DOI: 10.33594/000000336 Published online: 3 March 2021	© 2021 The Author(s). Published by Cell Physiol Biochem Press GmbH&Co. KG	30
	Menke et al.: Computational Ion Cha	annel Research	

sodium, but potassium can also pass through the pore with a larger barrier at the first ion binding site (S_0) preventing further translocation. Interestingly, lipid tails enter and exit channel fenestrations, small openings in the pore region. These openings are also expected to be entry portals for hydrophobic molecules. Overall, the simulation showed that the sodium ions are hydrated and can be accommodated next to each other. This contrasts with potassium channels, where potassium has to be dehydrated to pass through the channel. Additionally, they predicted five different ion binding sites (S_0 - S_4), whereas the S_1 , S_2 and S'4 ion binding sites were confirmed by a higher resolution x-ray structure of the pore [137].

Later, Li et al. used a Na, Rh channel structure from marine alpha proteobacterium HIMB114 (pdb 4dxw [138]) for the analysis of Na⁺ translocation. Unfortunately, the available structure shows the closed conformation. Therefore, an open state model was built using Modeller and an open $Na_{\nu}M$ structure. The structures were embedded in a lipid bilayer, and during MD simulations a transmembrane ion concentration and an electric field were applied. The binding free energies for Na⁺ was calculated by the free energy perturbation method. They adopted the simulation protocol described by Ulmschneider et al. [136]. The Na_vRh -open and the Na_vMs exhibited a similar level of ion permeability and the conductance of Na_{Rh} channels was calculated as 68 +/- 23 pS which is in comparable magnitude to the experimental reported values. More interestingly, they also mutated the four Ser180 residues at the construction site to the DEKA (Asp, Glu, Lys and Ala) motif of mammalian Na_{ν} channels. Their analysis showed that the DEKA motif exhibited a favourable ion binding site. This motif seems to retard ion permeation, since the DEKA mutant exhibited a smaller electric current than the WT Na, Rh channel. The conductance was estimated as 43 +/- 22 pS, also slightly lower as the WT Na, Rh, but in agreement with experimental results of eukaryotic *Na*, *1.4* channels.

These examples show the power of MD simulations for the analysis of ion permeability and channel conductance. Experimental determined values could be reproduced even by using homology models or in-silico mutants, which allows for a much better structural understanding than wet-lab experiments alone. In the following, we would like to shortly highlight additional examples. For a more detailed overview we refer to a review by Zhekova et al. [139].

Another important class of ion channels are potassium channels, since the control of K⁺ flux is important for the regulation of the transmembrane potential. Recently, Kopec et al. analysed the regulation of K⁺ flow of a calcium-gated prokaryotic potassium channel MthK by extensive MD simulation [140]. It is assumed that potassium channels have an activation gate and a selectivity filter gate that are allosterically coupled and regulate channel opening, closing and inactivation. Based on x-ray structures with different conformations of the activation gate as a starting point, the MD simulation showed that the K⁺ current is regulated by the opening of the activation gate. This gate was not physically blocking the K⁺ flow at any time. Interestingly, a wide opening of the activity gate leads to a water entry to the selectivity filter which is then destabilized. Already in 2000, Shrivastava et al. analysed the ion permeation through the bacterial potassium channel KscA [141]. In a very short simulation time of 5 ns from today's point of view, they could identify a concerted movement of K⁺ ions and water within the selectivity filter. They also suggested breathing motions that leads to the opening of the intracellular gate and allowing a K⁺ ion to leave the channel. A similar study presented by the group showing a similar behaviour for the mammalian Kir6.2 channel [142]. These examples show impressively, what was already possible 20 years ago with limited computational resources and what is possible today.

The last article to be presented is by Furini and Domene. It focuses more on the development of MD simulation techniques rather than on an analysis of specific ion channels [143]. This should highlight that method development is improving the quality and time scales of ion channel analyses. In the study of ion channels, it is important to investigate ion permeation, selectivity, and gating. All-atom molecular dynamics simulations played an important role in linking ion channel structure and function but required sampling several conformational states and accounting for large numbers of particles. For the analysis of ion

Cellular Physiology	Cell Physiol Biochem 2021;55(S3):14-45	
and Biochemistry	DOI: 10.33594/000000336 Published online: 3 March 2021	© 2021 The Author(s). Published by Cell Physiol Biochem Press GmbH&Co. KG
	Menke et al : Computational Ion Channel Research	

conductance on a reasonable time frame, an external electric field must be applied. The numbers of ions that traverse the channel per unit time are then counted and an estimate of the conductivity is obtained. Several advanced methods have been proposed as an alternative method to unbiased all-atom MD simulations, among which is metadynamics [143]. Here, an external bias potential to accelerate sampling along selected collective variables (CVs) is introduced. This bias potential discourages visiting regions of the configurational space already explored. In addition, the bias potential provides an estimate of the free energy as a function of the collective variables chosen once the simulation has converged. Metadynamics aims at enhancing rare events and reconstructing the underlying free energy landscape as a function of a set of order parameters, the already introduced CVs. This approach has several characteristics that have proven useful for the study of ion channels. Metadynamics can be used to accelerate state transitions along a predefined set of CVs, and at the same time render the free energy profile along them. Thus, a natural choice for the CVs in this case is the displacement of the ions along the pore axis. In their review, Furini and Domene describe exemplary studies of ion conduction with this algorithm that have predicted permeation pathways and the related binding free energy profiles. In addition, studies addressing efficient sampling of ion channel conformations during permeation are also described.

Analysis of functional states

As mentioned earlier, ion channels can change the state between open/closed or they can be inactivated. These different states are based on conformational changes that are mainly mediated by changes in the membrane potential (voltage-gated channels) or ligand-binding (ligand-gated channels). X-ray crystallography or cryo-EM can only show a static picture of the different ion channels states, sometimes in low resolution. Computational methods (especially MD simulations) are powerful techniques to analyse the different functional states in more detail and to gain insights into the required conformational changes and how they are induced. In the following, we will discuss studies highlighting the use of computational methods for the analysis of functional states of different ion channels.

In a recent study, Dämgen and Biggin analysed the open state of glycine receptors using MD simulations [144]. The glycine receptor (GlyR) belongs to the pentameric ligand-gated ion channels (pLGIC) that is selective for chloride ions and the endogenous agonist is the amino acid glycine. Upon binding of a neurotransmitter, here glycine, the receptor of this superfamily undergoes a conformational change to the open state. GlyRs have been the focus of structural studies leading to various structures of the GlyRs with agonist, antagonist, and modulators bound by cryo-EM [145] and X-ray crystallography [146–148]. Different functional states are available annotated to closed, open, and desensitized states, whereas MD simulations have provided useful insight into the functional annotation of states to ion channel structures, which is difficult to infer from structural information alone. It is assumed that two main constriction points are existing in the channel pore: A ring of five hydrophobic residues (usually leucine) in the middle of the TMD at the 90 position of the pore lining M2 helices (the L90gate) and, in the case of the GlyR, a ring of five proline residues at the 20 position near the intracellular mouth of the channel pore (the P20gate). The L90gate seems to close that gate when antagonists bound. The P-20gate gates the channel in the desensitized-like structures.

Among the available structures, one glycine bound cryo-EM structure (pdb 3JAE) [145] is of particular interest due to an unusually wide-open pore conformation. The distance from the pore centre to the backbone C_{α} at the -2' position is about 2 Å wider in this structure compared other open-like structures of the pLGIC superfamily. Previous work by Gonzalez-Gutierrez et al. found this open-wide pore to have a four times higher single-channel conductance than the experimental value [149]. In such simulations, artificial restraints were applied to the protein backbone so that it stays very close to the original cryo-EM model. Interestingly, if no such restraints are applied, the structure undergoes a so-

Cell Physiol Biochem 2021;55(S3):14-45 DOI: 10.33594/000000336 Published online: 3 March 2021 Cell Physiol Biochem Press GmbH&Co. KG Menke et al.: Computational Ion Channel Research

called hydrophobic collapse to a distinct state with a significantly smaller pore radius [150]. This behaviour has also been observed in simulations of other open (but not "wide-open") structures of the pLGIC family initiated via the non-polar interaction of the 90gate residues. A hydrophobic collapse or hydrophobic gating occurs in narrow hydrophobic pores where the energetically favoured expulsion of water prohibits ion permeation. This means that a pore whose dimensions are theoretically large enough to allow for ion permeation is not necessarily ion-permanent due to hydrophobic effects.

The basic hypothesis of Dämgen and Biggin was that the wide-open atomic model fitted into the cryo-EM density map may not correspond to an energetically stable representation of the open state under physiological conditions. They developed a careful equilibration protocol based on MD simulation for the exploration of other open state configurations, which agree with the cryo-EM density while preventing the pore from collapsing. Using this protocol, they identified an alternative conformation that remains open with a hydrated pore that allows selectively for frequent chloride ion permeation. This is based on a leucine side chain conformation that is not discernible in the original density map. These results unify previous, seemingly contradictory viewpoints and provide a way forward with regards to model transitions between states in the pLGIC family. To summarize, Dämgen and Biggin found a structural explanation for the hydrophobic collapse seen in some MD simulations and a stable open state seen in other MD simulations. This in particular shows the danger of low-resolution cryo-EM structures, since both starting points (leading to a hydrophobic collapse or staying stable) can be fitted to the cryo-EM density map.

The next presented study by Vijayan et al. deals with the GABA_A receptor, another pLGIC that is selective for Cl⁻ [151]. In absence of a structure, they built homology models based on two other pLGIC x-ray structures in open and closed state. The models were validated with molecular dynamics simulation and reproducing realistic binding modes of GABA via docking. Subsequently, elastic network modelling (ENM) was combined with normal mode analysis for the analysis of needed conformational changes between the closed and open state. Elastic network modelling is a coarse-grained method that treats proteins as networks of coupled harmonic oscillators and masses representing the starting structure. A "Twist to Turn" global motion was identified, accompanied by tilting and rotation of the M2 helices along the membrane normal. This rationalises the structural transition and at the same time gives an indication of a possibly conserved gating mechanism within the pLGIC.

The next two studies to be discussed deal again with voltage-gated potassium channels. Monticelli et al. used steered MD simulations for the analysis of possible coupling mechanisms between the motion of the voltage sensors and the opening of the pore in K_vAP channels [152]. For this, a complete model of the K_vAP ion channels was created by combining the available full protein structure with a high-resolution structure of the voltage sensor domain (VSD). During MD simulations, the VSD was pulled from the intracellular to the extracellular side using an applied force. This was assumed to mediate the conformational change from the closed to the open state. As a result, a coupling mechanism could be proposed that is based on a charged gating rather than conformational changes. Glass et al. analysed the influence of β -subunits on the central pore (α -subunits) of voltage gated sodium (Na_v) channels [153]. These β -subunits are known to modulate the voltage sensitivity and regulate the ion channel trafficking. They could show that different subunits (β 1 and β 3) show distinct differences in behaviour within a lipid bilayer.

The last discussed study by Schreiber et al. analysed the influence of inhibitors on the N-methyl-D-aspartate (NMDA) receptors [154]. NMDA receptors are ligand-gated glutamate receptor consisting of seven different subunits belonging to three groups (GluN1-3) that form various hetero-tetrameric receptors. Glutamate can bind to the ligand binding domain (LBD) of GluN1 and GluN3 and glycine to the LBD of GluN3. Binding of both agonists results in channel opening. An extra-terminal domain (ATD) also modulates channel opening. Ifenprodil is a known selective modulator of the GluN2B subunit binding to the interface within the ATD between GluN1 and GluN2B. MD simulations combined with site-directed mutagenesis and chemical modification of the ifenprodil scaffold revealed an aromatic interaction that prevents the needed reorientation of the α 5-helix for channel opening.

Cellular Physiology	Cell Physiol Biochem 2021;55(S3):14-45	
and Biochemistry	DOI: 10.33594/000000336 Published online: 3 March 2021	© 2021 The Author(s). Published by Cell Physiol Biochem Press GmbH&Co. KG
	Menke et al.: Computational Ion Channel Research	

Binding site analysis and ligand-target interaction

The identification of the binding site and binding mode of endogenous and exogenous ligands is another important area for a better understanding of ion channel function and drug development, as even voltage-gated ion channels can be modulated by small molecules. Molecular docking is the method of choice for predicting the binding mode of a known ligand in case the binding site is known. If not, docking alone can be misleading and further (experimental) validation is needed.

In a recent work, Ladefoged et al. analysed the binding mode of vortioxetine to the human 5-hydroxytryptamine_{3A} (5HT_{3A}) receptor [155]. 5-HT₃ receptors are ligand gated ion channels that are modulated by the neurotransmitter serotonin. The antidepressant drug vortioxetine modulates several 5-HT receptors and is known to have an antagonist effect on the 5-HT₃ receptor. It was expected that vortioxetine binds to the orthosteric binding site, but this was not shown before. Ladgefort et al. started their analysis with building a homology model of human 5-HT_{3A} receptor in an inactive conformation using the software MODELER [108]. The model was based on three template structures, since the existing homologous mouse 5-HT_{3A} structure showed an unclear conformational state. They additionally integrated structural information of an inactive human GABAa receptor and an antagonist-bound 5-HTBP receptor for the final modelling. Vortioxetine was then docked into the orthosteric binding site leading to six different potential bioactive binding modes with different interaction patterns. Subsequent molecular dynamics simulations combined with calculation of the relative free energy of binding using an MM-PBSA approach showed that two of the six potential binding poses represents the strongest and most stable binding mode.

For further validation, site-directed mutagenesis in the h5-HT_{3A} receptor was performed and the effect of single point mutations was measured on the vortioxetine IC_{50} . The mutations were selected for discrimination between the six different potential binding modes. The result of this mutation study clearly showed that four of the possible binding modes do not correspond to the actual bioactive binding mode. Only one binding mode can explain the results of the mutation study. The binding mode that shows one of the two strongest and most stable binding modes. Interestingly, another binding pose was not stable during the MD simulation and ended up in a conformation very similar to this most likely bioactive conformation. Overall, Ladefoged et al. concluded that vortioxetine has a unique inhibitory mechanism, as it shares common amino acids in the binding site as already existing 5-HT_{3A} ligands, but it also includes amino acids that were not reported previously. This study contributed to new insights into the inhibition of the 5-HT_{3A} receptor.

Besides this, the study highlights the power of computational methods as well as the dangers. Docking alone can be misleading as shown here by creating six potential binding modes. Additional information should be used for the identification of the near native binding pose. It could be helpful to compare the interaction pattern to already known complex structures, either by overlaying or by e.g. processing and comparing protein-ligand interaction fingerprints [156]. MD simulations could be another validation method, but especially in the case of membrane-bound ion channels the application of this method needs expert knowledge. As described, even MD simulations can lead to ambiguous binding modes, leaving a carefully selected site-directed mutagenesis study as the final choice for the validation of an assumed binding mode.

Another pitfall is the use of homology models, even when carefully evaluated and combined with mutagenesis studies. Brown et al. created $K_{ca}^2.3$ and $K_{ca}^3.1$ channel homology models for the structural elucidating of known small molecule activator binding and their selectivity for both channels [157]. The analysed activators were known to bind to the C-terminal Calmodulin (CaM)-binding domain, so they focused their modelling on this part. Using the Rosetta modelling suite [158] they created the C-terminal CaM binding domain in complex with CaM for the $K_{ca}^3.1$ and $K_{ca}^2.3$ channel based on a high-resolution crystal structure of a $K_{ca}^2.2$ channel CaM-binding domain in complex with CaM and the small

Cellular Physiology	Cell Physiol Biochem 2021;55(S3):14-45		
	DOI: 10.33594/00000336	© 2021 The Author(s). Published by	
and Biochemistry	Published online: 3 March 2021	Cell Physiol Biochem Press GmbH&Co. KG	
	Manka at al. Computational Ion Channel Passarsh		

molecule activator NS309 (pdb 4J9Z [159]). Due to significant differences in a loop region forming the CaM-binding domain between K_{ca} 3.1 and K_{ca} 2 channels, this region was predicted using Rosetta's loop prediction method [160]. Based on previous site-directed mutagenesis studies, it was known that several activators bind in a similar region as NS309 in the K_{ca} 2.2 channel. Therefore, the selectivity of these activators was first analysed. Rosetta based ligand docking [161] and validation using site-directed mutagenesis indicated that all K_{ca} activators show a hydrogen bond to CaM-M51 in both channels. A closer look into two derivatives (SKA-121 and SK-111) with selectivity for K_{ca} 3.1 over K_{ca} 2.3 revealed K_{ca} 3.1 R362 as an important part of a hydrogen bond network that explains this selectivity. Unfortunately, the availability of the full-length K_{ca} 3.1 structure [3] later showed that the created homology model was partly wrong (see Fig. 6). Shim et al. further analysed K_{ca} 3.1 and could confirm the binding site of K_{ca} 3.1 activators near the helix S4-S5 linker [162]. They analysed the binding mode of the activator SKA111 in more detail and could reveal a binding site between the S45A helix and the CaM N-lobe via docking and mutagenesis studies. They could further explain that the previously identified important residues do not interfere directly with activator binding, but with CaM-binding. This presumably also influences the activator binding site.

To summarize, the de novo prediction of an uncertain loop region has led to a wrong homology model that seems to be valid, since predicted mutagenesis studies showed the expected results but for the wrong reasons. So, one has to be careful if a very low sequence similarity is occurring in a region that seems to be important for ligand binding. In a different study, this group created a homology model of the K_{ca} 3.1 channel pore region based on a $K_{1.2}$ - $K_{2.1}$ chimeric channel structure (pdb 2R9R [163]) and an open *KcsA* structure (pdb 3FB5) [164]. Here, the pore region shows a high similarity to the finally solved K_{ca} 3.1 structure. Together with site-directed mutagenesis they could validate the binding site of several channel inhibitors.

In the following, we would like to give an overview about other interesting examples regarding the identification of potential binding site and ligand interactions. In a recent study, Nguyen et al. analysed the interaction of lidocaine on the human cardiac sodium channel $hNa_v 1.5$ using homology modelling and molecular dynamics simulations [165]. Key interaction residues for the binding of antiarrhythmic and local anaesthetic drugs were already known before this analysis. They build a homology model of $hNA_v 1.5$ in a partially open state based on the cryo-EM structure of $eeNa_v 1.4$. It can be expected that the homology model is of high quality, since both channels share a sequence identity of 84% in the pore-forming transmembrane region. Docking studies revealed a possible similar binding site of antiarrhythmic and local anaesthetic drugs. Most interestingly, they could show the two different access pathways of lidocaine using microsecond MD simulation on a supercomputer as already proposed by Hille in 1977 [166]: A hydrophobic pathway between the lipid membrane and through a hydrophilic intracellular gating pathway.



Fig. 6. a) Binding of SKA-111 (yellow) in the calmodulin (CaM, orange red) binding region of the $K_{ca}3.1$ (pink) homology model provided by Brown et al. [157]. b) Overlay of CaM binding regions of $K_{ca}3.1$ homology model and cryo-EM structure. c) CaM (kakhi) binding to the CaM binding region of the later solved cryo-EM $K_{ca}3.1$ structure (pdb 6cno [3]).

Cellular Physiology	Cell Physiol Biochem 2021;55(S3):14-45	
and Biochemistry	DOI: 10.33594/000000336 Published online: 3 March 2021	© 2021 The Author(s). Published by Cell Physiol Biochem Press GmbH&Co. KG
	Menke et al.: Computational Ion Channel Research	

In 2019, Faulkner et al. reported new binding sites of fentanyl, which is a pain reliever analgesic, with *Gloeobacter violaceus* ligand-gated ion channel (GLIC) [167]. The new identified interactions of this anaesthetic with the channel are different from the conventional binding modes observed for other anaesthetics. As the mechanism of action of this drug is still unclear, they employed molecular dynamics simulation with three runs of 500ns on the apo form of GLIC protein inserted in a phospholipid bilayer. Four fentanyl molecules were added to the simulation box in each system. The opioid molecule initiated at 20ns of the simulation an interaction with the extracellular hairpin loop of the channel, before extending to the binding site of the GLIC channel for the rest of the simulation in two out of three runs, indicating the strength of the interaction. These novel binding sites lead to conformational changes that were not reported before, as a closure of the helix pore creating a hydrophobic gate formed by 233-Ile and 240-Ile residues.

In a work by Yuan et al. four K⁺ channel scorpion toxins were analysed for their binding to the $K_v 1.2$ channel [168]. This is of interest as an exemplary analysis to discuss here since peptides in general are an important class of channel activity modulators. The analysed toxins all belong to the group of α -K⁺ channel toxins (α -KTxs) and share a similar folding pattern consisting of one helix and an antiparallel β -sheet. The basis for this analysis was a $K_v 1.2$ - $K_v 2.1$ paddle chimera X-ray structure with bound charybdotoxin (ChTx, pdb 4JTA) [169]. The structure of the four α -KTxs were overlaid onto the bound ChTx and the ChTx was deleted afterwards. Molecular dynamics simulations of these complexes led to the identification of important hydrophobic patches, hydrogen-bonds, and salt bridges between this channel and the toxins. Four $K_v 1.2$ -specific interacting amino acids (D353, Q358, V381, and T383) are identified as important for the first time. This discovery might help to design highly selective $K_v 1.2$ -channel inhibitors by altering amino acids of these toxins binding to the four channel residues.

In the next presented study, Li et al. used docking-based virtual screening as the final validation of an identified ligand binding site in the gating charge pathway of *KCNQ2* channels [170]. Voltage-gated potassium channels are built up from a pore domain and a voltage-sensor domain (VSD). The channel opening or closing is mediated via a residue translocation in the VSD through a physical gating charge pathway in response to membrane potential changes. The compound ztz240 is a known KCNQ2 activator and was used as a chemical probe throughout this study. A site-directed mutagenesis study suggests the binding of ztz240 to the open-state voltage-sensor domain. A homology model was built using an open-state $K_v 1.2$ channel structure (pdb 2R9R [163]) and ztz240 was docked afterwards into the identified binding pocket. A subsequent MD simulation was used for the refinement of the binding pose in the KCNQ2 model. An alanine scanning of the binding pocket residues further validated the binding pocket. A docking-based virtual screening identified nine compounds of five different chemotypes with effects on the outward current of the KCNQ2 channel.

In a recent study, Brömmel and colleagues developed novel fluorescently labelled small-molecules targeting the K_{ca} 3.1 channel, which displayed promising results in staining experiments [171]. The modelling of the synthesized dye with senicapoc as a starting point highlighted a perfect insertion in the pore of the protein, corroborating a similar binding of the senicapoc moiety described in the rosetta model of this protein [164].

Virtual Screening approaches

Computational methods are not only valuable for the analysis of ion channels, their binding sites and interactions of known ligands. Computational molecular design methods can also be applied for virtual screening approaches for the rational and fast identification of new modulators. In case of available ion channels structures (or validated homology models) molecular docking is the method of choice for structure-based virtual screening. This is especially interesting, since functional screening of ion channels is often very time

Cellular Physiology	Cell Physiol Biochem 2021;55(S3):14-45	
and Biochemistry	DOI: 10.33594/000000336 Published online: 3 March 2021	© 2021 The Author(s). Published by Cell Physiol Biochem Press GmbH&Co, KG
	Menke et al Computational Ion Channel Research	

consuming and needs measurements of transfected cells. We will discuss in the following some successful virtual screening examples.

In 2011, Nury et al. [172] could solve a X-ray structure of propofol bound to a *Gloeobacter* violaceus ligand-gated ion channel (GLIC, pdb 3P50), which is a bacterial homolog of GABA. receptors. Heusser et al. 2013 used this as a starting point for the docking-based identification of new GLIC and GABA, modulators [173]. It was already known that propofol, isoflurane, and midazolam are anaesthetics that play the role of positive allosteric modulators (PAMs) for GABA, receptors. In the GLIC complex structure, propofol binds in a lipophilic intra-subunit cavity of the transmembrane domain, which is highly conserved in the GABA, receptor and other pentameric ligand-gated ion channels (pLGICs). Therefore, it can be expected that the GABAA receptor shows a very similar propofol binding site. Heusser et al. started with a MD simulation and selected a best performing snapshot based on an initial validation experiment. Here, propofol together with 100 very similar molecules were docked into 302 MD snapshots. One snapshot was selected for the final virtual screening that allowed to select propofol out of the 100 decoys with a near native binding mode. Over 153,000 commercially available compounds were then docked into the GLIC propofol binding site. A set of 22 molecules were selected for functional testing on recombinant GLIC based on their predicted ranking and 13 of them displayed a modulation of this ion channel. Afterwards, they were tested on GABA,. An experimental validation was then followed by mutation studies in the possible binding site of GLIC and GABA_A. Among the selected compounds, one of them showed a similar binding action as propofol on GLIC and GABA, and has been affected by the mutation that decreased its action on both targets. This study has provided inestimable information about mutant proteins of GLIC and GABAA and the importance of the amino acids responsible for the interaction with new PAMs. It has clarified the binding site and the conformational changes that occur to the protein allowing an anaesthetic modulation. In addition, new modulators could be identified using virtual screening.

The next two articles describe the identification of dual active inhibitors. This polypharmacology-based modulation is an important approach in medicinal chemistry, with the hope that compounds show synergistic effects when modulation more than one protein target [174]. In the first article discussed, Kowal et al. searched for dual active compounds interacting as an inhibitor on acetylcholinesterase (AChE) and as a positive allosteric modulator for the α 7 nicotinic acetylcholine receptor (nAChR) using computational molecular design methods [175]. Even when both targets differ with respect to their structure and function, their binding pockets efficiently recognise the same neurotransmitter. They started their work with galantamine, a marketed drug that already shows the desired dual activity. Galantamine was already solved in complex with AChE and an acetylcholine binding protein (AChBP) which shows a high similarity to nAChR. They built a homology model of human α 7 nAChR with galantamine based on this AChBP structure and an α 1 nAChR extracellular subunit. They performed a structure-based virtual screening using a dataset of 87,250 natural products and an in-house database containing 250 lycopodium alkaloids. The compounds were first docked into the α 7 nAChR homology model, and the highest scored compounds were then docked into AChE. A visual inspection led to the selection of 13 compounds for testing. These compounds were tested for activity on AChE and α 7 nAChR and two compounds showed the desired dual activity on both receptors. Additional four compounds were identified as nAChR antagonists.

In a similar study, the same group performed virtual screening to obtain dual hit molecules as acetylcholinesterase (AChE) inhibitors and as an α 7 nicotinic acetylcholine receptor (α 7 nAChR) agonists [176]. This time they based their study on an X-ray structure of AChE co-crystallized with donepezil and a homology model of α 7 nAChR which was constructed using an (α 4)2(β 2)3nAChR structure and an AChBP structure as templates. They docked a library of 3,848,234 drug-like molecules into both protein targets and analysed the intersecting high-scoring compounds. Based on visual inspection focusing on the docking pose and molecular diversity, 15 of these compounds were purchased for *in vitro* validation. Two compounds showed dual activity with AChE inhibition and activation of the α 7 nAChR.

Cellular Physiology	Cell Physiol Biochem 2021;55(S3):14-45	
and Biochemistry	DOI: 10.33594/000000336 Published online: 3 March 2021	© 2021 The Author(s). Published by Cell Physiol Biochem Press GmbH&Co. KG
Menke et al: Computational Ion Channel Research		annel Besearch

In an alternative virtual screening approach, Callejo et al. used a ligand-based comparison for the identification of new acid-sensing ion channel 3 (ASIC3) modulators [177]. They compared the 3D shape and the chemical similarity of a known ASIC3 modulator to FDA-approved drugs library of 1884 compounds. The top 150 drugs were visually inspected for their similarity to the query compound 2-guanidine-4-methylquinazoline (GMQ) and five were selected for testing. One of these drugs (Guanabenz, GBZ) activates ASIC3 at physiological pH. They also tested sephin1 as a GBZ derivative. Three homology models of rASIC3 were built based on available chicken ASIC1 structures in the open, closed and desensitized states. The GMQ, GBZ and sephin1 were docked and the binding site and the ligand interactions discussed. To summarize, Callejo et al. were able to identify a new ASIC3 modulator using a ligand-based virtual screening, that leads to the prediction of five drugs as potential ASIC3 modulators that were tested. This shows the power of ligand-based molecular design methods in case a structure and binding site is unknown.

Summary and Discussion

In this second part of our review, we focused on structure-based computational methods that rely on the availability of structures of the ion channel of interest. In case these structures are not available, several examples are shown where created homology models were sufficient enough to get new insights about ion channel function and ligand binding. In most cases, homology modelling is combined with MD simulations for the validation of these models. Apart from this, MD simulation is a powerful tool to analyse ion channel function. In combination with docking, potential ligand binding sites and ligand interaction can be analysed in detail. In order to fully understand the needed conformational changes for function, it is essential to obtain different functional states of the ion channel from closed or open gating to conductive or non-conductive selectivity filters. The understanding of the structural insights of the conformational changes that occur rely then on the ability of the 3D resolved structures to sample these different states. The different external stimuli trigger a cascade of conformational changes leading to the activation of the ion channel, the selectivity filter is afterward undergoing an ion flux then returning to a non-conductive state [178].

This review displayed a series of articles where scientists took advantage of the increasing power of computational advances to reveal new insights about ion channel function. It was shown that it is possible to calculate the ion conductance using MD simulation in a similar range as experimentally determined values. However, structure-based computational methods allow much more than this, especially a structural understanding that is difficult to obtain using wet-lab experiments alone. This includes an understanding of ions selectivity, occupancy and translocation in the selectivity filters, conformational changes and the reason for activation and inactivation, binding site identification, mutations altering the functional behaviour, and many more. These series of events can be computationally performed but should be validated experimentally using well-designed electrophysiological experiments, site-directed mutagenesis, or ligand testing. The reason is that computational analysis can point towards possible hypotheses of ion channel function, that are otherwise not possible, but can also be misleading.

There are still unsolved and highly complex events that cannot be accessed from molecular dynamics simulations yet such as spontaneous lipid interaction with the protein, allosteric modulation, ligand reaching the binding pocket. This leads to high expectations in the future in new advanced techniques such as metadynamics [143], umbrella sampling [179] and steered MD [152] that should allow us to analyse such complex events on a long time scale.

Cellular Physiology and Biochemistry Cell Physiol Biochem 2021;55(S3):14-45 DOI: 10.33594/000000336 © 2021 The Author(s). Published by Published online: 3 March 2021 Cell Physiol Biochem 2021;55(S3):14-45 DOI: 10.33594/000000336 Published online: 3 March 2021 Cell Physiol Biochem Press GmbH&Co. KG Menke et al.: Computational Ion Channel Research

Conclusion

Ion channels are one of the three most important protein families in the field of drug discovery. To date, there are still great uncertainties and much that remains undiscovered within this important protein family. We demonstrated in this review the strength and abilities of computational methods in ion channel research to support wet lab experiments and to get a better understanding of ion channel function and the interaction of modulators. Indeed, the computational methods are widely spread nowadays and applied within the scientific community, either in academia or in industrial pharmaceutical drug research. As it goes from structure-based drug design methods that use the structure of the protein, to ligand-based methods that focus purely on ligands. Such methods are used to start from scratch looking into massive databases in order to suggest few molecules to be tested and validated experimentally. It could be referred to as finding a needle in a haystack. In the era where data is abundant, methods such as data mining and knowledge discovery, machine learning and QSAR, similarity searching, pharmacophore modeling, homology modeling, docking studies, and biomolecular simulation can drastically reduce the resources required to find this needle. As an example, when experimental research represents proteins as rigid and static, MD simulations shed light on the flexible and mobile protein. To assemble a valid model for such simulations of the ion channel or any membrane protein in general, few parameters have to be taken into account, such as the lipid bilayer that surrounds the protein and the lipid interactions resulting, or the allosteric modulation of few molecules interacting with the protein.

Despite the above advantages, there are still limitations in the field of computational molecular design. The ligand-based methods still rely on the availability of data that are correct and curated, whereas structure-based methods heavily rely on 3D structures that are sometimes of bad quality or not available for interesting targets. Even when a protein structure is well characterized there could still be gaps and black holes such as allosteric modulations, protein conformation and flexibility, or promiscuity that cannot be easily accessed. The usage of these information could thus be risky and induce wrong understanding of the behaviour of a protein. Every scientist that applies these methods has to be also aware of the limitations of the used methods and algorithms. There is the danger that computational methods always produce results, even when they are misleading or wrong. Therefore, a basic understanding is still needed when applying computational methods and only an experienced computational scientist will recognize these pitfalls. We are still far away from only "pressing a button". Especially, the field of machine learning has yet to write its success stories as much efforts have been made to develop models, but different to traditional computational models has not seen much experimental validation.

However, if applied correctly, computational approaches are powerful methods to accelerate the understanding of ion channel function and the development of chemical probes or potential drugs.

Acknowledgements

This work was supported by the Research Training Group (GRK2515) "Chemical biology of ion channels (Chembion)" funded by the Deutsche Forschungsgemeinschaft (DFG), which is gratefully acknowledged.

Disclosure Statement

The authors declare that they have no conflicts of interest.

Cell Physiol Biochem 2021;55(S3):14-45 DOI: 10.33594/000000336 Published online: 3 March 2021 Cell Physiol Biochem Press GmbH&Co. KG Menke et al.: Computational Ion Channel Research

References

- 1 Ashcroft FM: From molecule to malady. Nature 2006;440:440–447.
- 2 Hille B: Ion Channels of Excitable Membranes, ed 3, Cell Press, 2001.
- 3 Lee CH, MacKinnon R: Activation mechanism of a human SK-calmodulin channel complex elucidated by cryo-EM structures. Science 2018;360:508–513.
- 4 Braun N, Sheikh ZP, Pless SA: The current chemical biology tool box for studying ion channels. J Physiol 2020;598:4455–4471.
- 5 He K, Zhang X, Ren S, Sun J: Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision and Pattern Recognition 2016;770-778.
- 6 Vaswani A, Brain G, Shazeer N, Parmar N, Uszkoreit J, Jones L, et al.: Attention Is All You Need. Adv Neural Inf Process Syst 2017;30.
- 7 Silver D, Huang A, Maddison CJ, Guez A, Sifre L, Van Den Driessche G, et al.: Mastering the game of Go with deep neural networks and tree search. Nature 2016;529:484–489.
- 8 Schneider P, Walters WP, Plowright AT, Sieroka N, Listgarten J, Goodnow RA, et al.: Rethinking drug design in the artificial intelligence era. Nat Rev Drug Discov 2020;19:353–364.
- 9 Yang X, Wang Y, Byrne R, Schneider G, Yang S: Concepts of Artificial Intelligence for Computer-Assisted Drug Discovery. Chem Rev 2019;119:10520–10594.
- 10 Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, et al.: A deep learning approach to antibiotic discovery. Cell 2020;180:688–702.
- 11 Bzdok D, Altman N, Krzywinski M: Statistics versus machine learning. Nat Methods 2018;15:233–234.
- 12 Mayr A, Klambauer G, Unterthiner T, Steijaert M, Wegner JK, Ceulemans H, et al.: Large-scale comparison of machine learning methods for drug target prediction on ChEMBL. Chem Sci 2018;9:5441–5451.
- 13 Wu Z, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, Pappu AS, et al.: MoleculeNet: a benchmark for molecular machine learning. Chem Sci 2018;9:513–530.
- 14 David L, Thakkar A, Mercado R, Engkvist O: Molecular representations in AI-driven drug discovery: a review and practical guide. J Cheminform 2020;12:1–22.
- 15 Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on Computational learning theory. New York, ACM, 1992, pp 144–152.
- 16 Breiman L: Random forests. Mach Learn 2001;45:5–32.
- 17 Breiman L: Bagging predictors. Mach Learn 1996;26:123-40.
- 18 Zupan J, Gasteiger J: Neural Networks in Chemistry and Drug Design, ed 2. USA, John Wiley & Sons, 1999.
- 19 Goodfellow I, Bengio Y, Courville A: Deep Learning. MIT Press; 2016.
- 20 Liu LX, Li ML, Tan FY, Lu MC, Wang KL, Guo YZ, et al.: Local sequence information-based support vector machine to classify voltage-gated potassium channels. Acta Biochim Biophys Sin 2006;38:363–371.
- 21 Saha S, Zack J, Singh B, Raghava GPS: VGIchan: prediction and classification of voltage-gated ion channels. Genomics Proteomics Bioinformatics 2006;4:253–258.
- 22 Eddy SR: Profile hidden Markov models. Bioinformatics 1998;14:755–763.
- 23 Lin H, Ding H: Predicting ion channels and their types by the dipeptide mode of pseudo amino acid composition. J Theor Biol 2011;269:64–69.
- 24 Gao J, Cui W, Sheng Y, Ruan J, Kurgan L: PSIONplus: accurate sequence-based predictor of ion channels and their types. PLoS One 2016;11:e0152964.
- 25 Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al.: Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
- 26 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. J Mol Biol 1990;215:403–410.
- 27 Tiwari AK, Srivastava R: An efficient approach for the prediction of ion channels and their subfamilies. Comput Biol Chem 2015;58:205–221.
- 28 Zhao YW, Su ZD, Yang W, Lin H, Chen W, Tang H: IonchanPred 2.0: a tool to predict ion channels and their types. Int J Mol Sci 2017;18:1838.
- 29 Gao J, Miao Z, Zhang Z, Wei H, Kurgan L: Prediction of ion channels and their types from protein sequences: Comprehensive review and comparative assessment. Curr Drug Targets 2019;20:579–592.
- 30 Han K, Wang M, Zhang L, Wang Y, Guo M, Zhao M, et al.: Predicting ion channels genes and their types with machine learning techniques. Front Genet 2019;10:399.

Cellular Physiology and Biochemistry Cell Physiol Biochem 2021;55(S3):14-45 DOI: 10.33594/000000336 © 2021 The Author(s). Published by Cell Physiol Biochem Press GmbH&Co. KG

- 31 Taju SW, Ou YY: DeepIon: Deep learning approach for classifying ion transporters and ion channels from membrane proteins. J Comput Chem 2019;40:1521–1529.
- 32 Gao J, Wei H, Cano A, Kurgan L: PSIONplusm Server for Accurate Multi-Label Prediction of Ion Channels and Their Types. Biomolecules 2020;10:876.
- 33 von Heijne G: Membrane-protein topology. Nat Rev Mol Cell Biol 2006;7:909–918.
- 34 Kyte J, Doolittle RF: A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982;157:105–132.
- 35 Almeida JG, Preto AJ, Koukos PI, Bonvin AMJJ, Moreira IS: Membrane proteins structures: A review on computational modeling tools. Biochim Biophys Acta Biomembr 2017;1859:2021–2039.
- 36 Fariselli P, Compiani M, Casadio R: Predicting secondary structures of membrane proteins with neural networks. Eur Biophys J 1993;22:41–51.
- 37 Rost B, Sander C: Improved prediction of protein secondary structure by use of sequence profiles and neural networks. Proc Natl Acad Sci 1993;90:7558–7562.
- 38 Rost B, Sander C: Prediction of protein secondary structure at better than 70% accuracy. J Mol Biol 1993;232:584–599.
- 39 Rost B, Sander C: Secondary structure prediction of all-helical proteins in two states. Protein Eng Des Sel 1993;6:831–836.
- 40 Rost B, Sander C: Combining evolutionary information and neural networks to predict protein secondary structure. Proteins Struct Funct Bioinforma 1994;19:55–72.
- 41 Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL: Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 2001;305:567–580.
- 42 Tusnady GE, Simon I: The HMMTOP transmembrane topology prediction server. Bioinformatics 2001;17:849–50.
- 43 Käll L, Krogh A, Sonnhammer ELL: A combined transmembrane topology and signal peptide prediction method. J Mol Biol 2004;338:1027–1036.
- 44 Käll L, Krogh A, Sonnhammer ELL: An HMM posterior decoder for sequence feature prediction that includes homology information. Bioinformatics 2005;21:251-257.
- 45 Viklund H, Elofsson A: Best alpha-helical transmembrane protein topology predictions are achieved using hidden Markov models and evolutionary information. Protein Sci 2004;13:1908–1917.
- 46 Martelli PL, Fariselli P, Casadio R: An ENSEMBLE machine learning approach for the prediction of all-alpha membrane proteins. Bioinformatics 2003;1:205-211.
- 47 Jones DT: Improving the accuracy of transmembrane protein topology prediction using evolutionary information. Bioinformatics 2007;23:538–544.
- 48 Jones DT, Taylor WR, Thornton JM: A model recognition approach to the prediction of all-helical membrane protein structure and topology. Biochemistry 1994;33:3038–3049.
- 49 Jones DT: Do transmembrane protein superfolds exist? FEBS Lett 1998;423:281-285.
- 50 Shen H, Chou JJ: MemBrain: improving the accuracy of predicting transmembrane helices. PLoS One 2008;3:e2399.
- 51 Viklund H, Elofsson A: OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. Bioinformatics 2008;24:1662–1668.
- 52 Viklund H, Bernsel A, Skwark M, Elofsson A: SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology. Bioinformatics 2008;24:2928–2929.
- 53 Bernsel A, Viklund H, Hennerdal A, Elofsson A: TOPCONS: consensus prediction of membrane protein topology. Nucleic Acids Res 2009;37:465-468.
- 54 Tsirigos KD, Peters C, Shu N, Käll L, Elofsson A: The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. Nucleic Acids Res 2015;43:401-407.
- 55 Dobson L, Reményi I, Tusnády GE: CCTOP: a Consensus Constrained TOPology prediction web server. Nucleic Acids Res 2015;43:408-412.
- 56 Wang H, Yang Y, Yu J, Wang X, Zhao D, Xu D, et al.: DMCTOP: topology prediction of alpha-helical transmembrane protein based on deep multi-scale convolutional neural network. 2019 IEEE International Conference on Bioinformatics and Biomedicine, San Diego, CA, USA, 2019, pp 36–43.
- 57 Feng SH, Zhang WX, Yang J, Yang Y, Shen HB: Topology Prediction Improvement of α-helical Transmembrane Proteins Through Helix-tail Modeling and Multiscale Deep Learning Fusion. J Mol Biol 2020;432:1279–1296.

Cellular Physiology and Biochemistry Cell Physiol Biochem 2021;55(S3):14-45 DOI: 10.33594/000000336 © 2021 The Author(s). Published by Cell Physiol Biochem Press GmbH&Co. KG

- 58 Robertson N, Rappas M, Doré AS, Brown J, Bottegoni G, Koglin M, et al.: Structure of the complement C5a receptor bound to the extra-helical antagonist NDT9513727. Nature 2018;553:111–114.
- 59 Lee CH, MacKinnon R: Structures of the Human HCN1 Hyperpolarization-Activated Channel. Cell 2017;168:111-120.
- 60 Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al.: The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
- 61 Schulz G: beta-barrel membrane proteins. Curr Opin Struct Biol 2000;10:443-447.
- 62 Hayat S, Elofsson A: BOCTOPUS: improved topology prediction of transmembrane β barrel proteins. Bioinformatics 2012;28:516–522.
- 63 Bagos PG, Liakopoulos TD, Spyropoulos IC, Hamodrakas SJ: PRED-TMBB: a web server for predicting the topology of beta-barrel outer membrane proteins. Nucleic Acids Res 2004;32:W400-404.
- 64 Tsirigos KD, Elofsson A, Bagos PG: PRED-TMBB2: improved topology prediction and detection of betabarrel outer membrane proteins. Bioinformatics 2016;32:665-671.
- 65 Singh NK, Goodman A, Walter P, Helms V, Hayat S: TMBHMM: a frequency profile based HMM for predicting the topology of transmembrane beta barrel proteins and the exposure status of transmembrane residues. Biochim Biophys Acta 2011;1814:664–670.
- 66 Bagos PG, Liakopoulos TD, Hamodrakas SJ: Evaluation of methods for predicting the topology of beta-barrel outer membrane proteins and a consensus prediction method. BMC Bioinformatics 2005;6:7
- 67 Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, et al.: QSAR modeling: Where have you been? Where are you going to? J Med Chem 2014;57:4977–5010.
- 68 Grisoni F, Consonni V, Todeschini R: Impact of Molecular Descriptors on Computational Models. Methods Mol Biol 2018;1825:171–209.
- 69 Tropsha A: Best practices for QSAR model development, validation, and exploitation. Mol Inform 2010;29:476–488.
- 70 Andrea TA, Kalayeh H: Applications of neural networks in quantitative structure-activity relationships of dihydrofolate reductase inhibitors. J Med Chem 1991;34:2824–2836.
- 71 Wikel JH, Dow ER: The use of neural networks for variable selection in QSAR. Bioorg Med Chem Lett 1993;3:645–651.
- Aoyama T, Suzuki Y, Ichikawa H: Neural networks applied to pharmaceutical problems. III. Neural networks applied to quantitative structure-activity relationship (QSAR) analysis. J Med Chem 1990;33:2583–2590.
- 73 Maddalena DJ, Johnston GAR: Prediction of receptor properties and binding affinity of ligands to benzodiazepine/GABAA receptors using artificial neural networks. J Med Chem 1995;38:715–724.
- 74 Klon AE: Machine learning algorithms for the prediction of hERG and CYP450 binding in drug development. Expert Opin Drug Metab Toxicol 2010;6:821–833.
- 75 O'Brien SE, de Groot MJ: Greater than the sum of its parts: combining models for useful ADMET prediction. J Med Chem 2005;48:1287–1291.
- 76 Yang SY, Huang Q, Li LL, Ma CY, Zhang H, Bai R, et al.: An integrated scheme for feature selection and parameter setting in the support vector machine modeling and its application to the prediction of pharmacokinetic properties of drugs. Artif Intell Med 2009;46:155–163.
- 77 Xue Y, Li ZR, Yap CW, Sun LZ, Chen X, Chen YZ: Effect of molecular descriptor feature selection in support vector machine classification of pharmacokinetic and toxicological properties of chemical agents. J Chem Inf Comput Sci 2004;44:1630–1638.
- 78 Moriwaki H, Tian YS, Kawashita N, Takagi T: Mordred: a molecular descriptor calculator. J Cheminform 2018;10:4.
- 79 Kode chemoinformatics: ALVADESC 1.0. URL: https://chm.kode-solutions.net/products_alvadesc.php.
- 80 Braga R, Alves V, Silva M, Muratov E, Fourches D, Tropsha A, et al.: Tuning hERG Out: Antitarget QSAR Models for Drug Development. Curr Top Med Chem 2014;14:1399–1415.
- 81 OECD: Principles for the validation, for regulatory purposes, of (quantitative) structure-activity relationship models, 2004. URL: https://www.oecd.org/chemicalsafety/risk-assessment/37849783.pdf.
- 82 C Braga R, M Alves V, FB Silva M, Muratov E, Fourches D, Tropsha A, et al.: Tuning HERG out: antitarget QSAR models for drug development. Curr Top Med Chem 2014;14:1399–1415.
- 83 Konda LSK, Praba SK, Kristam R: hERG liability classification models using machine learning techniques. Comput Toxicol 2019;12:100089.

Cellular Physiology and Biochemistry Cell Physiol Biochem 2021;55(S3):14-45 DOI: 10.33594/000000336 © 2021 The Author(s). Published by Cell Physiol Biochem Press GmbH&Co. KG

42

- 84 Siramshetty VB, Chen Q, Devarakonda P, Preissner R: The Catch-22 of predicting hERG blockade using publicly accessible bioactivity data. J Chem Inf Model 2018;58:1224–1233.
- 85 Khalifa N, Kumar Konda LS, Kristam R: Machine learning-based QSAR models to predict sodium ion channel (Nav 1.5) blockers. Future Med Chem 2020;12:1829–1843.
- 86 Lancaster MC, Sobie EA: Improved prediction of drug-induced Torsades de Pointes through simulations of dynamics and machine learning algorithms. Clin Pharmacol Ther 2016;100:371–379.
- 87 Yeomans DC, Levinson SR, Peters MC, Koszowski AG, Tzabazis AZ, Gilly WF, et al.: Decrease in inflammatory hyperalgesia by herpes vector-mediated knockdown of Nav1. 7 sodium channels in primary afferents. Hum Gene Ther 2005;16:271–277.
- 88 Kong W, Tu X, Huang W, Yang Y, Xie Z, Huang Z: Prediction and Optimization of NaV1. 7 Sodium Channel Inhibitors Based on Machine Learning and Simulated Annealing. J Chem Inf Model 2020;60;2739-2753.
- 89 Willighagen EL, Mayfield JW, Alvarsson J, Berg A, Carlsson L, Jeliazkova N, et al.: The Chemistry Development Kit (CDK) v2.0: atom typing, depiction, molecular formulas, and substructure searching. J Cheminform 2017;9:33.
- 90 Sun M, Zhao S, Gilvary C, Elemento O, Zhou J, Wang F: Graph convolutional networks for computational drug development and discovery. Brief Bioinform 2020;21:919–935.
- 91 Cai C, Guo P, Zhou Y, Zhou J, Wang Q, Zhang F, et al.: Deep learning-based prediction of drug-induced cardiotoxicity. J Chem Inf Model 2019;59:1073–1084.
- 92 Karimi M, Wu D, Wang Z, Shen Y: DeepAffinity: interpretable deep learning of compound--protein affinity through unified recurrent and convolutional neural networks. Bioinformatics 2019;35:3329–3338.
- 93 Wang Y Bin, You ZH, Yang S, Yi HC, Chen ZH, Zheng K: A deep learning-based method for drug-target interaction prediction based on long short-term memory neural network. BMC Med Inform Decis Mak 2020;20:49.
- 94 Celik N, O'Brien F, Brennan S, Rainbow RD, Dart C, Zheng Y, et al.: Deep-Channel uses deep neural networks to detect single-molecule events from patch-clamp data. Commun Biol 2020;3:1–10.
- 95 Rao S, Klesse G, Stansfeld PJ, Tucker SJ, Sansom MSP: A heuristic derived from analysis of the ion channel structural proteome permits the rapid identification of hydrophobic gates. Proc Natl Acad Sci 2019;116:13989–13995.
- 96 Torrisi M, Pollastri G, Le Q: Deep learning methods in protein structure prediction. Comput Struct Biotechnol J 2020:18:1301-1310.
- 97 Rodriguez-Perez R, Bajorath J: Multitask machine learning for classifying highly and weakly potent kinase inhibitors. Acs Omega 2019;4:4367–4375.
- 98 Simoes RS, Maltarollo VG, Oliveira PR, Honorio KM: Transfer and multi-task learning in QSAR modeling: advances and challenges. Front Pharmacol 2018;9:74.
- 99 Sosnin S, Vashurina M, Withnall M, Karpov P, Fedorov M, Tetko IV: A survey of multi-task learning methods in chemoinformatics. Mol Inform. 2019;38:615–621.
- 100 Cai C, Wang S, Xu Y, Zhang W, Tang K, Ouyang Q, et al.: Transfer learning for drug discovery. J Med Chem 2020;63:8683–8694.
- 101 Li X, Fourches D: Inductive transfer learning for molecular activity prediction: Next-Gen QSAR Models with MolPMoFiT. J Cheminform 2020;12:1–15.
- 102 Wang S, Guo Y, Wang Y, Sun H, Huang J: SMILES-BERT: large scale unsupervised pre-training for molecular property prediction. Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, Niagara Falls, NY, USA, 2019, pp 429–436.
- 103 Alley EC, Khimulya G, Biswas S, AlQuraishi M, Church GM: Unified rational protein engineering with sequence-based deep representation learning. Nat Methods 2019;16:1315–1322.
- 104 Doyle DA, Cabral JM, Pfuetzner RA, Kuo A, Gulbis JM, Cohen SL, et al.: The structure of the potassium channel: Molecular basis of K+ conduction and selectivity. Science 1998;280:69–77.
- 105 Moraes I, Evans G, Sanchez-Weatherby J, Newstead S, Stewart PDS: Membrane protein structure determination The next generation. Biochim Biophys Acta 2014;1838:78-87.
- 106 Liao M, Cao E, Julius D, Cheng Y: Structure of the TRPV1 ion channel determined by electron cryomicroscopy. Nature 2013;504:107–112.
- 107 Nygaard R, Kim J, Mancia F: Cryo-electron microscopy analysis of small membrane proteins. Curr Opin Struct Biol 2020;64:26–33.

Cellular Physiology and Biochemistry Cell Physiol Biochem 2021;55(S3):14-45 Dol: 10.33594/000000336 © 2021 The Author(s). Published by Cell Physiol Biochem Press GmbH&Co. KG

- Menke et al.: Computational Ion Channel Research
- 108 Sali A, Blundell TL: Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.
- 109 John B, Sali A: Comparative protein structure modeling by iterative alignment, model building and model assessment. Nucleic Acids Res 2003;31:3982–3992.
- 110 Haddad Y, Adam V, Heger Z: Ten quick tips for homology modeling of high-resolution protein 3D structures. PLOS Comput Biol 2020;16:e1007449.
- 111 Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, Costanzo L Di, et al.: Protein Data Bank: The single global archive for 3D macromolecular structure data. Nucleic Acids Res 2019;47:520–528.
- 112 Haas J, Barbato A, Behringer D, Studer G, Roth S, Bertoni M, et al.: Continuous Automated Model EvaluatiOn (CAMEO) complementing the critical assessment of structure prediction in CASP12. Proteins Struct Funct Bioinforma 2018;86:387–398.
- 113 Croll TI, Sammito MD, Kryshtafovych A, Read RJ: Evaluation of template-based modeling in CASP13. Proteins Struct Funct Bioinforma 2019;87:1113–1127.
- 114 Webb B, Sali A: Comparative protein structure modeling using MODELLER. Curr Protoc Bioinforma 2016; DOI: 10.1002/cpbi.3.
- 115 Yang J, Zhang Y: I-TASSER server: New development for protein structure and function predictions. Nucleic Acids Res 2015;43:174–181.
- 116 Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE: The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 2015;10:845–858.
- 117 Buchan DWA, Jones DT: The PSIPRED Protein Analysis Workbench: 20 years on. Nucleic Acids Res 2019;47:402–407.
- 118 Xu J, Mcpartlon M, Li J: Improved protein structure prediction by deep learning irrespective of co-evolution information. bioRxiv 2020; DOI: 10.1101/2020.10.12.336859.
- 119 Kim DE, Chivian D, Baker D: Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res 2004;32:526:531
- 120 Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al.: SWISS-MODEL: Homology modelling of protein structures and complexes. Nucleic Acids Res 2018;46:296–303.
- 121 Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. Nat Struct Biol 2002;9:646–652.
- 122 McCammon JA, Gelin BR, Karplus M: Dynamics of folded proteins. Nature 1977;267:585–590.
- 123 Hollingsworth SA, Dror RO: Molecular Dynamics Simulation for All. Neuron 2018;99:1129–1143.
- 124 Lichtenthaler FW: 100 Years "Schlüssel-Schloss-Prinzip": What Made Emil Fischer Use this Analogy? Angew Chemie Int Ed English 1995;33;2364–2374.
- 125 Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE: A geometric approach to macromolecule-ligand interactions. J Mol Biol 1982;16:269–288.
- 126 Dias R, de Azevedo Jr. W: Molecular Docking Algorithms. Curr Drug Targets 2008;9:1040–1047.
- 127 Ewing TJA, Makino S, Skillman AG, Kuntz ID. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. J Comput Aided Mol Des 2001;15:411–428.
- 128 Jones G, Willett P, Glen RC, Leach AR, Taylor R: Development and validation of a genetic algorithm for flexible docking. J Mol Biol 1997 267:727–748.
- 129 Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, et al.: Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 1998;19:1639–1662.
- 130 Trott O, Olson AJ: AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 2009;31:455–461.
- 131 Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al.: Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. J Med Chem 2004;47:1739– 1749.
- 132 Bernèche S, Roux B: Molecular dynamics of the KcsA K+ channel in a bilayer membrane. Biophys J 2000;78:2900–2917.
- 133 Guidoni L, Torre V, Carloni P: Potassium and sodium binding to the outer mouth of the K+ channel. Biochemistry 1999;38:8599–8604.
- 134 Allen TW, Kuyucak S, Chung SH: Molecular dynamics study of the KcsA potassium channel. Biophys J 1999;77:2502–2516.

Cellular Physiology and Biochemistry Cell Physiol Biochem 2021;55(S3):14-45 DOI: 10.33594/000000336 © 2021 The Author(s). Published by Published online: 3 March 2021

- 135 McCusker EC, Bagnéris C, Naylor CE, Cole AR, D'Avanzo N, Nichols CG, et al.: Structure of a bacterial voltagegated sodium channel pore reveals mechanisms of opening and closing. Nat Commun 2012;3:1–8.
- 136 Ulmschneider MB, Bagnéris C, McCusker EC, DeCaen PG, Delling M, Clapham DE, et al.: Molecular dynamics of ion transport through the open conformation of a bacterial voltage-gated sodium channel. Proc Natl Acad Sci 2013;110:6364–6369.
- 137 Naylor CE, Bagnéris C, DeCaen PG, Sula A, Scaglione A, Clapham DE, et al.: Molecular basis of ion permeability in a voltage?gated sodium channel. EMBO J 2016;35:820–830.
- 138 Zhang X, Ren W, Decaen P, Yan C, Tao X, Tang L, et al.: Crystal structure of an orthologue of the NaChBac voltage-gated sodium channel. Nature 2012;486:130–134.
- 139 Zhekova HR, Ngo V, da Silva MC, Salahub D, Noskov S: Selective ion binding and transport by membrane proteins – A computational perspective. Coord Chem Rev 2017;345:108–136.
- 140 Kopec W, Rothberg BS, de Groot BL: Molecular mechanism of a potassium channel gating through activation gate-selectivity filter coupling. Nat Commun 2019;10:5366.
- 141 Shrivastava IH, Sansom MSP: Simulations of ion permeation through a potassium channel: Molecular dynamics of KcsA in a phospholipid bilayer. Biophys J 2000;78:557–570.
- 142 Capener CE, Shrivastava IH, Ranatunga KM, Forrest LR, Smith GR, Sansom MSP: Homology modeling and molecular dynamics simulation studies of an inward rectifier potassium channel. Biophys J 2000;78:2929–2942.
- 143 Furini S, Domene C: Computational studies of transport in ion channels using metadynamics. Biochim Biophys Acta 2016;1858:1733–1740.
- 144 Dämgen MA, Biggin PC: A Refined Open State of the Glycine Receptor Obtained via Molecular Dynamics Simulations. Structure 2020;28:130-139.
- 145 Du J, Lü W, Wu S, Cheng Y, Gouaux E: Glycine receptor mechanism elucidated by electron cryo-microscopy. Nature 2015;526:224–229.
- 146 Huang X, Chen H, Michelsen K, Schneider S, Shaffer PL: Crystal structure of human glycine receptor-α3 bound to antagonist strychnine. Nature 2015;526:277–280.
- 147 Huang X, Shaffer PL, Ayube S, Bregman H, Chen H, Lehto SG, et al.: Crystal structures of human glycine receptor ?3 bound to a novel class of analgesic potentiators. Nat Struct Mol Biol 2017;24:108–13.
- 148 Huang X, Chen H, Shaffer PL. Crystal Structures of Human GlyRα3 Bound to Ivermectin. Structure 2017;25:945-950.
- 149 Gonzalez-Gutierrez G, Wang Y, Cymes GD, Tajkhorshid E, Grosman C: Chasing the open-state structure of pentameric ligand-gated ion channels. J Gen Physiol 2017;149:1119–1138.
- 150 Cerdan AH, Martin NÉ, Cecchini M: An Ion-Permeable State of the Glycine Receptor Captured by Molecular Dynamics. Structure 2018;26:1555-1562.
- 151 Vijayan RSK, Trivedi N, Roy SN, Bera I, Manoharan P, Payghan P V., et al.: Modeling the closed and open state conformations of the GABAA ion channel Plausible structural insights for channel gating. J Chem Inf Model 2012;52:2958–2969.
- 152 Monticelli L, Robertson KM, MacCallum JL, Tieleman DP: Computer simulation of the KvAP voltage-gated potassium channel: Steered molecular dynamics of the voltage sensor. FEBS Lett 2004;564:325–332.
- 153 Glass WG, Duncan AL, Biggin PC: Computational Investigation of Voltage-Gated Sodium Channel β3 Subunit Dynamics. Front Mol Biosci 2020;7:40.
- 154 Schreiber JA, Schepmann D, Frehland B, Thum S, Datunashvili M, Budde T, et al.: A common mechanism allows selective targeting of GluN2B subunit-containing N-methyl-D-aspartate receptors. Commun Biol 2019;2:1–14.
- 155 Ladefoged LK, Munro L, Pedersen AJ, Lummis SCR, Bang-Andersen B, Balle T, et al.: Modeling and mutational analysis of the binding mode for the multimodal antidepressant drug vortioxetine to the human 5-HT3A receptor. Mol Pharmacol 2018;94:1421–1434.
- 156 Jasper JB, Humbeck L, Brinkjost T, Koch O: A novel interaction fingerprint derived from per atom score contributions: exhaustive evaluation of interaction fingerprint performance in docking based virtual screening. J Cheminform 2018;10:15.
- 157 Brown BM, Shim H, Zhang M, Yarov-Yarovoy V, Wulff H: Structural Determinants for the Selectivity of the Positive KCa3.1 Gating Modulator 5-Methylnaphtho [2,1-d]oxazol-2-amine (SKA-121). Mol Pharmacol 2017;92:469–480.

Cellular Physiology and Biochemistry Cell Physiol Biochem 2021;55(S3):14-45 DOI: 10.33594/000000336 © 2021 The Author(s). Published by Published online: 3 March 2021 Cell Physiol Biochem Press GmbH&Co. KG

- 158 Rohl CA, Strauss CEM, Misura KMS, Baker D: Protein Structure Prediction Using Rosetta. Methods Enzymol 2004;383:66–93.
- 159 Zhang M, Pascal JM, Zhang JF: Unstructured to structured transition of an intrinsically disordered protein peptide in coupling Ca2+-sensing and SK channel activation. Proc Natl Acad Sci 2013;110:4828–4833.
- 160 Wang C, Bradley P, Baker D: Protein-Protein Docking with Backbone Flexibility. J Mol Biol 2007;373:503– 519.
- 161 Meiler J, Baker D: ROSETTALIGAND: Protein-small molecule docking with full side-chain flexibility. Proteins Struct Funct Genet 2006;65:538–548.
- 162 Shim H, Brown BM, Singh L, Singh V, Fettinger JC, Yarov-Yarovoy V, et al.: The Trials and Tribulations of Structure Assisted Design of KCa Channel Activators. Front Pharmacol 2019;10:972.
- 163 Long SB, Tao X, Campbell EB, MacKinnon R: Atomic structure of a voltage-dependent K+ channel in a lipid membrane-like environment. Nature 2007;450:376–382.
- 164 Nguyen HM, Singh V, Pressly B, Jenkins DP, Wulff H, Yarov-Yarovoy V: Structural insights into the atomistic mechanisms of action of small molecule inhibitors targeting the KCa3.1 channel pore. Mol Pharmacol 2017;91:392–402.
- 165 Nguyen PT, DeMarco KR, Vorobyov I, Clancy CE, Yarov-Yarovoy V: Structural basis for antiarrhythmic drug interactions with the human cardiac sodium channel. Proc Natl Acad Sci 2019;116:2945–2954.
- 166 Hille B: Local anesthetics: Hydrophilic and hydrophobic pathways for the drug-receptor reaction. J Gen Physiol 1977;69:497–515.
- 167 Faulkner C, Plant DF, De Leeuw NH: Modulation of the Gloeobacter violaceus Ion Channel by Fentanyl: A Molecular Dynamics Study. Biochemistry 2019;58:4804–4808.
- 168 Yuan S, Gao B, Zhu S: Molecular dynamics simulation reveals specific interaction sites between scorpion toxins and Kv1.2 channel: Implications for design of highly selective drugs. Toxins (Basel) 2017;9:354.
- 169 Banerjee A, Lee A, Campbell E, MacKinnon R: Structure of a pore-blocking toxin in complex with a eukaryotic voltage-dependent K+ channel. Elife 2013; DOI: 10.7554/eLife.00594.
- 170 Li P, Chen Z, Xu H, Sun H, Li H, Liu H, et al. The gating charge pathway of an epilepsy-associated potassium channel accommodates chemical ligands. Cell Res 2013;23:1106–1118.
- 171 Brömmel K, Maskri S, Maisuls I, Konken CP, Rieke M, Pethő Z, et al.: Synthesis of Small-Molecule Fluorescent Probes for the In vitro Imaging of Calcium-Activated Potassium Channel KCa3.1. Angew Chemie Int Ed Engl 2020;59:8277–8284.
- 172 Nury H, Van Renterghem C, Weng Y, Tran A, Baaden M, Dufresne V, et al.: X-ray structures of general anaesthetics bound to a pentameric ligand-gated ion channel. Nature 2011;469:428–431.
- 173 Heusser SA, Howard RJ, Borghese CM, Cullins MA, Broemstrup T, Lee US, et al.: Functional validation of virtual screening for novel agents with general anesthetic action at ligand-gated ion channels. Mol Pharmacol 2013;84:670–678.
- 174 Proschak E, Stark H, Merk D: Polypharmacology by Design: A Medicinal Chemist's Perspective on Multitargeting Compounds. J Med Chem 2019;62:420–444.
- 175 Kowal NM, Indurthi DC, Ahring PK, Chebib M, Olafsdottir ES, Balle T: Novel approach for the search for chemical scaffolds with dual activity with acetylcholinesterase and the α7 nicotinic acetylcholine receptor—a perspective for the treatment of neurodegenerative disorders. Molecules 2019;24:446.
- 176 Oddsson S, Kowal NM, Ahring PK, Olafsdottir ES, Balle T: Structure-based discovery of dual-target hits for acetylcholinesterase and the α7 nicotinic acetylcholine receptors: In silico studies and in vitro confirmation. Molecules 2020;25:2872.
- 177 Callejo G, Pattison LA, Greenhalgh JC, Chakrabarti S, Andreopoulou E, Hockley JRF, et al.: In silico screening of GMQ-like compounds reveals guanabenz and sephin1 as new allosteric modulators of acid-sensing ion channel 3. Biochem Pharmacol 2020;174:113834.
- 178 Ostmeyer J, Chakrapani S, Pan AC, Perozo E, Roux B: Recovery from slow inactivation in K+ channels is controlled by water molecules. Nature 2013;501:121–124.
- 179 Torrie GM, Valleau JP: Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. J Comput Phys 1977;23:187–199.