Computational Biology and Chemistry 58 (2015) 205–221

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem

Research article

An efficient approach for the prediction of ion channels and their subfamilies

Arvind Kumar Tiwari*, Rajeev Srivastava

Department of Computer Science & Engineering, Indian Institute of Technology (BHU), Varanasi, U.P., India

ARTICLE INFO

Article history:
Received 16 March 2015
Received in revised form 25 June 2015
Accepted 8 July 2015
Available online 26 July 2015

Keywords:
Ion channels
Voltage gated ion channels
Ligand gated ion channels
Random forest
Minimum redundancy maximum relevance
Sodium
Potassium
Calcium
Chloride

ABSTRACT

Ion channels are integral membrane proteins that are responsible for controlling the flow of ions across the cell. There are various biological functions that are performed by different types of ion channels. Therefore for new drug discovery it is necessary to develop a novel computational intelligence techniques based approach for the reliable prediction of ion channels families and their subfamilies. In this paper random forest based approach is proposed to predict ion channels families and their subfamilies by using sequence derived features. Here, seven feature vectors are used to represent the protein sample, including amino acid composition, dipeptide composition, correlation features, composition, transition and distribution and pseudo amino acid composition. The minimum redundancy and maximum relevance feature selection is used to find the optimal number of features for improving the prediction performance. The proposed method achieved an overall accuracy of 100%, 98.01%, 91.5%, 93.0%, 92.2%, 78.6%, 95.5%, 84.9%, MCC values of 1.00, 0.92, 0.88, 0.88, 0.90, 0.79, 0.91, 0.81 and ROC area values of 1.00, 0.99, 0.99, 0.99, 0.99, 0.95, 0.99 and 0.96 using 10-fold cross validation to predict the ion channels and non-ion channels, voltage gated ion channels and ligand gated ion channels, four subfamilies (calcium, potassium, sodium and chloride) of voltage gated ion channels, and four subfamilies of ligand gated ion channels and predict subfamilies of voltage gated calcium, potassium, sodium and chloride ion channels respectively.

1. Introduction

Ion channels are membrane proteins that are responsible for electrical signaling by gating the flow of ions across the cell membrane. These are the prominent component of nervous systems. Ion channels are classified by gating that is used for opening and closing the channels. The voltage gated ion channels are open and close based on the voltage gradient across the cell membrane, while ligand gated ion channels open and close based on the ligand binding of the ion channels. The voltage gated ion channels play an important role in generation and propagation of the nerve impulse and in cell homeostasis (Bezanilla, 2005). The dysfunction of ion channels plays an important role in the development of various diseases such as hypertension, defective insulin secretion, cardiac arrhythmias, neurological diseases such as epilepsy and even developmental defects such as osteopetrosis (Jentsch et al., 2004). So it is necessary to know about the structure and function of the ion channels to develop a new drug for these diseases. Ion channels play an important target for antiepileptic drug design, antihypertensive and antipsychotics disorder such as schizophrenia (Abernethy and Schwartz, 1999; Yogeeswari et al., 2004). Currently, there are various computational intelligence techniques based approaches that have been proposed to predict membrane proteins (Shen et al., 2006; Cai et al., 2006; Yang et al., 2007; Chou and Shen, 2007a; Zhao and Ding, 2008; Wang and Yang, 2010; Huang et al., 2014). Similarly various computational intelligence techniques based approaches have been proposed to predict nuclear receptor (Bhasin and Raghava, 2004a; Cai and Li, 2005; Gao et al., 2009; Wang et al., 2011; Xiao et al., 2012; Wang and Xiao, 2014) and G-protein coupled receptor (Bhasin and Raghava, 2004b, 2005; Gao and Wang, 2006; Gu et al., 2010; Peng et al., 2010).

However, few papers reported computational intelligence techniques based methods to predict ion channels and their types. The authors of the papers (Willett et al., 2007; Pourbasheer et al., 2009) have proposed computational intelligence techniques based method to predict the activity of ion channel proteins. In the paper (Liu et al., 2006) a support vector machine based method has been proposed to predict five types of voltage gated potassium channels and obtained accuracy of 98%. The authors of the paper (Saha et al.,

* Corresponding author.
E-mail addresses: arvind.rs.cse12@itbhu.ac.in (A.K. Tiwari), rajeev.cse@iitbhu.ac.in (R. Srivastava).

http://dx.doi.org/10.1016/j.compbiolchem.2015.07.002
1476-9271/

2006) have proposed a support vector machine based method to predict four types of voltage gated ion channels by using amino acid composition and dipeptide composition and obtained overall accuracy of 97.78%. In the paper (Chen and Lin, 2012) a support vector machine based method has been proposed to predict voltage gated potassium channel subfamilies by using amino acid composition and dipeptide composition and obtained overall accuracy of 93.09%. In the paper (Lin and Ding, 2011) the authors proposed a support vector machine based method to predict ion channels and their types by using dipeptide mode of pseudo amino acid composition and obtained overall accuracy of 86.6% to discriminate ion channels from non-ion channels, overall accuracy of 92.6% to classify voltage gated ion channels and ligand gated ion channels and an overall accuracy of 87.8% to predict four types of voltage gated ion channels.

In this paper a random forest based method has been proposed to predict ion channels and their types by using sequence derived properties of a protein. In this paper 857 sequence derived features with seven features vectors such as amino acid composition, dipeptide composition, correlation, and composition, transition and distribution and pseudo amino acid composition are used to predict the ion channels and their types. The minimum redundancy maximum relevance (mRMR) based feature selection is used to improve the predictive accuracy. In this paper the proposed method used four level strategies to predict ion channels and their types. First, it is determined that protein sequence is ion channel or non-ion channel. Second, if protein is classified as ion channels then the method classify the protein into two groups viz voltage gated ion channels or ligand gated ion channels. Third, it is classified into the subfamilies of voltage gated ion channels and ligand gated ion channels. Fourth, it determines the subfamilies of calcium, potassium, sodium and chloride voltage gated ion channels. The two parameter of random forest the size of random subset of features (*mtry*) and the number of trees in the forest (*ntree*) are used to decrease the error rate. Therefore each of the four levels is developed using a random forest classifier with optimized value of *ntree* and *mtry*.

2. Material and methods

In this paper the sequence of the ion channels is extracted from the Uniport (http://www.uniprot.org), ligand gated ion channels

Table 1
Number of sequences belonging to each ion channels and their subfamilies.

Ion/non-ion channels	Families of ion channels	Subfamilies of ion channels	Sub-subfamily of voltage gated ion channels	No. of sequences	No. of sequences	No. of sequences	No. of sequences
Ion channel	Voltage gated ion channels	Calcium	P-Type	190	634	...	2141
			R-Type	51			
			L-Type	280			
			N-Type	25			
			T-Type	88			
		Potassium	Kv1	61	646		
			Kv2	51			
			Kv3	52			
			Kv4	63			
			Kv5	22			
			Kv6	51			
			Kv7	59			
			Kv8.2	42			
			Kv9	52			
			Kv10	55			
			Kv11	59			
			Kv12	52			
			Kv13	27			
		Sodium	Alpha subunits	250	401		
			Beta subunits	151			
		Chloride	ClC1	48	146		
			ClC2	18			
			ClC3	43			
			ClC4	7			
			ClC5	15			
			ClCk	9			
			ClC6	6			
	Ligand gated ion channels	GABAA receptors	27		314	314	
		Glycine receptors	34				
		glutamate receptors	184				
		Nicotinic acetylcholine receptors	69				
Non ion channel				722	----	--	722