Chapter 10

Evolutionary Intelligence–Based Feature Descriptor Selection for Efficient Identification of Anti–Cancer Peptides

Deepak Singh  
National Institute of Technology, Raipur, India

Dilip Singh Sisodia  
https://orcid.org/0000-0001-9845-290X  
National Institute of Technology, Raipur, India

Pradeep Singh  
National Institute of Technology, Raipur, India

ABSTRACT

A novel evolutionary-based feature selection model for ACPs identification that will explore the relationships hidden across the various feature descriptors is explored in this chapter. In this model, the authors amalgamate the nine feature descriptors from the three groups of peptide feature descriptors including amino acid composition (three descriptors), grouped amino acid composition and composition/transition/distribution (three descriptors). The proposed model integrates these features to unfold the hidden association between the diverse features in peptide classification. However, the inclusion of irrelevant, redundant, and noisy attributes in the model building process phase can result in poor predictive performance and increased computation. Hence, evolutionary-based feature selection is utilized in the model that involves a combination of search and feature utility estimation by ReliefF score. Through extensive experiments on benchmark dataset, it is demonstrated that the proposed model achieves improved performance.

DOI: 10.4018/978-1-7998-2120-5.ch010
INTRODUCTION

Early detection and the use of surgery, radiation therapy, and chemo-therapeutic drugs including aromatase inhibitors can reduce cancer (Tan, 2016). However, the unfavorable side effects of the medication makes the patients remain traumatized. Therefore, the research and discovery for target-specific and less side-effect cancer therapy is still undergoing. Traditional methods for the treatment of cancer includes surgery, radiation therapy and chemotherapy, which may also depend on the location, stage of the disease, and the patient condition. Despite advances, these methods are expensive and can often exhibit damaging effects on normal cells (F. Liu, B, Sun, Liu, & Wang, 2015). Additionally, there is a growing concern that cancer cells may develop resistance to chemotherapy and molecularly-targeted therapies. Moreover, cancer cells are known to develop multidrug resistance through a broad range of mechanisms, which not only makes these cells resistant to the drug in use for treatment, but also several other compounds. As soon as the molecular mechanism behind cancer (or, as a matter of fact, any disease) is understood, the next logical step is to discover a desirable remedy for it (Yang et al., 2018). Therefore, in view of the above, there is an urgent need to discover and design novel anti-cancer drugs to combat this deadly disease.

ACPs kill’s cancer cells by interacting with the anionic cell membrane components of cancer cells, without impairing the normal cells. These peptides do not harm the body’s physiological functions, facilitating a new path for cancer treatment (Khosravian, Kazemi Faramarzi, Mohammad Beigi, Behbahani, & Mohabatkar, 2013). Anti-cancer peptides are safer than synthetic drugs, and have greater efficacy, selectivity, and specificity. ACPs represent a promising line of treatment. These are short peptides (typically 5–50 amino acids in length) that exhibit high specificity, high tumour penetration and ease of synthesis and modification, in addition to low cost of production. Traditionally, ACPs were identified and characterized using biochemical experiments (Hossain, Yasmin, Hosen, & Nabi, 2018). They are derived from protein sequences and to identify ACPs from a protein sequence seems to be highly expensive, time-consuming, and overly complex to be utilised in a high-throughput manner (Guida et al., 2018). Therefore, it is essential to develop sequence-based computational methods to rapidly identify potential ACP candidates from the sequencing data prior to their synthesis.

The development of computational methods using a many machine learning models and peptide features were successfully applied to identify potential ACPs candidates. Tyagi et al. (Tyagi et al., 2013) proposed a model to identify ACPs using amino acid composition and binary profiles as the input of support vector machine (SVM). Shortly afterwards, Hajisharifi et al. (Hajisharifi, Piryaiee, Mohammad Beigi, Behbahani, & Mohabatkar, 2014b), using Chou’s pseudo amino acid composition and the local alignment kernel based method, also proposed a model to do the same. Both methods produced relatively promising results and have indeed played a vital role in stimulating the development of this area. A crucial factor for the success of a prediction method is composition of the feature set (Li & Wang, 2016). The ideal feature set should capture the major and subtle patterns from the sequence to differentiate actual positives from negatives.

Existing methods has shown that compositional, physicochemical, and structural properties; sequence order; and the pattern of terminal residues are the most frequently adopted features for ACP predictions (Wei, Zhou, Chen, Song, & Su, 2018). Another solution that emerges is integrating these various kinds of features as the input to train a classifier and build a predictive model. However, this could cause increase in dimension or on the other hand, simply integrating diverse features can lead to information redundancy and thereby influence the predictive performance (Khushaba, Al-ani, Al-jumaily, & Box,