Computational Biology and Chemistry 59 (2015) 126–130

Contents lists available at ScienceDirect

Computational Biology and Chemistry

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

Research article

QSAR modeling of the antimicrobial activity of peptides as a mathematical function of a sequence of amino acids

Mariya A. Toropova[a], Aleksandar M. Veselinović[b,*], Jovana B. Veselinović[b], Dušica B. Stojanović[c], Andrey A. Toropov[d]

[a] Dipartimento di Scienze Farmaceutiche, Universita'degli Studi di Milano, Via L. Mangiagalli 25, 20133 Milano, Italy
[b] University of Niš, Faculty of Medicine, Department of Chemistry, Bulevar Dr Zorana Đinđića 81, 18000 Niš, Serbia
[c] University of Niš, Faculty of Medicine, Public Health Institute Niš, 18000 Niš, Serbia
[d] IRCCS- Istituto di Ricerche Farmacologiche Mario Negri, 20156, Via La Masa 19, Milano, Italy

ARTICLE INFO

Article history:
Received 6 April 2015
Received in revised form 4 August 2015
Accepted 15 September 2015
Available online 21 September 2015

Keywords:
Monte Carlo method
Mastoparan analogs
CORAL software
QSAR
Optimal descriptor
Antimicrobial activity

ABSTRACT

Antimicrobial peptides have emerged as new therapeutic agents for fighting multi-drug-resistant bacteria. However, the process of optimizing peptide antimicrobial activity and specificity using large peptide libraries is both tedious and expensive. Therefore, computational techniques had to be applied for process optimization. In this work, the representation of the molecular structure of peptides (mastoparan analogs) by a sequence of amino acids has been used to establish quantitative structure—activity relationships (QSARs) for their antibacterial activity. The data for the studied peptides were split three times into the training, calibration and test sets. The Monte Carlo method was used as a computational technique for QSAR models calculation. The statistical quality of QSAR for the antibacterial activity of peptides for the external validation set was: $n = 7$, $r^2 = 0.8067$, $s = 0.248$ (split 1); $n = 6$, $r^2 = 0.8319$, $s = 0.169$ (split 2); and $n = 6$, $r^2 = 0.6996$, $s = 0.297$ (split 3). The stated statistical parameters favor the presented QSAR models in comparison to 2D and 3D descriptor based ones. The Monte Carlo method gave a reasonably good prediction for the antibacterial activity of peptides. The statistical quality of the prediction is different for three random splits. However, the predictive potential is reasonably well for all cases. The presented QSAR modeling approach can be an attractive alternative of 3D QSAR at least for the described peptides.

1. Introduction

Over the past decade, bacterial resistance has increased dramatically leading to a huge global health problem (Overbye and Barrett, 2005; Neu, 1992). Bacterial resistance is a natural biological phenomenon that can be associated to their fight for survival. Mechanisms of bacterial resistance to antibiotics are very different and complex and depend both on the bacteria and the structure of antibiotics. However, there are several known mechanisms of antimicrobial resistance. In the first mechanism, bacteria may acquire genes encoding antibiotic-inactivating enzymes, like β- lactamases. In the second mechanism, bacteria may acquire efflux pumps which prevent the antibacterial agent to reach its target site. In the third mechanism, bacteria may acquire several genes for a certain metabolic pathway that ultimately produces altered bacterial cell walls. This leads to loosing the binding site of the antimicrobial agent, or on the another hand, bacteria may acquire mutations that can limit the access of antimicrobial agents to the intracellular target site via the downregulation of porin genes (Radu et al., 2011; Band and Weiss, 2015; Li et al., 2015). Antimicrobial peptides (AMPs) have recently been recognized as suitable leads in several areas of drug discovery due to their high affinity and specificity toward their targets and quite favorable toxicity profiles (Dathe et al., 1996; Levy and Marshall, 2004). A wide range of pathogen agents like bacteria, viruses or fungi can be inhibited with AMPs through the interaction and modulation of microbial membrane permeability (Bowdish and Hancock, 2005; Brandenburg et al., 2010). The main physico-chemical features important for AMP selectivity for the pathogen membrane are net electric charge, hydrophobicity and peptide length (Yandek et al., 2009). However, the exact mechanism of antimicrobial peptides against the membrane of pathogen agents is still unknown. One of the most effective antimicrobial peptides of great medical interest is mastoparan, isolated from wasp venom (Amin et al., 2003; Hirata et al., 2000;

* Corresponding author. Fax: +381 18 4238770.
E-mail address: aveselinovic@medfak.ni.ac.rs (A.M. Veselinović).

http://dx.doi.org/10.1016/j.compbiolchem.2015.09.009
1476-9271/

Hori et al., 2001; Sukumar et al., 1997). Mastoparan analogs isolated from different wasp species have improved pharmacological effects making them especially interesting for the pharmaceutical industry due to their different selectivities against *Bacillus subtilis* and *Escherichia coli* and low hemolytic activity (Cherkasov and Jankovic, 2004; dos Santos Cabrera et al., 2008, 2011; Leite et al., 2011). Since mastoparans have wide therapeutic applicability and extremely low side effects, the design and synthesis of new mastoparan derivatives are an important trend in preclinical and clinical studies (Bartie et al., 2008; Fan et al., 2011). Unfortunately, the application of methods by which the antimicrobial activity of new derivatives of mastoparan is calculated without requiring the presence of microbial membranes is necessary due to limited knowledge about the mastoparan mechanism of action at pathogen membranes.

Computational methods are a valuable tool in designing and evaluating new compounds because they are fast and cost-effective. Furthermore, they do not require complicated chemical synthesis and testing procedures. Especially quantitative structure—activity relationships (QSAR) and quantitative structure—property relationship (QSPR) techniques are important for drug design since in the absence of a targeted protein X-ray structure they are the only available methods. Therefore QSARs are a chemoinformatic tool to predict the biological activity of potential therapeutic agents by means of the analysis available experimental data (García et al., 2011; Mullen et al., 2011). One of the most important steps in building an accurate QSAR/QSPR model is the selection of descriptors used for encoding molecules under study. As a starting point for building a computational model, descriptors are calculated for as many molecular attributes as possible. However, in further model development a crucial step is the selection of the most appropriate descriptors for predicting the desired activity/property. Usually the selection of effective descriptors is problem-dependent and there is no universal rule to achieve this goal. Also, when QSAR models are built with 3D representation of molecules and descriptors derived from them, new problems emerge because complex softwares with high computation time and complex computer resources are typically used to keep into account the tridimensional structure (Avram et al., 2012; Du et al., 2007; Cerovsky et al., 2008; Zhou et al., 2010). Therefore, the QSAR modeling where 3D representation is excluded is an attractive approach. For simpler molecules, the widely used paradigm for building up a QSPR/QSAR model is the following: Endpoint = Function (system of atoms). In the first approximation, biochemical processes are not accompanied by the destruction of the amino acids. Hence, these phenomena can be expressed by the paradigm: Endpoint = Function (sequence of amino acids). The second paradigm is a very attractive alternative to the first paradigm for the cases of peptides and proteins, because the representation of amino acids by symbols A, R, N, etc. is very compact. In addition, this approach, if successful, can indicate a particular function of an amino acid (that is an autonomic system of atoms) in a biochemical process.

The CORAL software (Veselinović et al., 2013; Achary, 2014; Toropova and Toropov, 2014a; CORAL, 2015) gives the possibility to carry out the calculation with a string of symbols which represent peptide structures, using the same algorithms which have been involved in the SMILES-based QSPR/QSAR modeling (Toropov et al., 2012). The aim of the present study is to build QSAR models for calculating the antibacterial activity of mastoparan analogs based

Table 1
Experimantal and calculated pMIC data for antibacterial peptides.

Split1	Split 2	Split 3	Mastoparan analogs	Sequence of amino acides	pMIC expr	pMIC calc (split1)	pMIC calc (split2)	pMIC calc (split3)
TRN	TRN	TRN	MP	INWLKLGKKMMSAL	5.040	4.729	4.725	4.731
TRN	TRN	TRN	MP-2	INWLKLGKKLLSAL	4.600	4.722	4.673	4.689
TRN	TRN	TRN	MP-5	INWLKLGKKMMSAI	4.450	4.626	4.669	4.685
TRN	TRN	TRN	MP-6	SNWLKLGKKMMSAL	4.390	4.520	4.489	4.465
CLB	TRN	TRN	PDDA	INWKKIFQKVKNLV	4.920	4.899	4.981	4.940
TRN	TRN	TRN	PDDA-1	INWKKIFEKVKNLV	4.350	4.823	4.780	4.752
TRN	CLB	TRN	PDDA-2	INWKKIFEKVKDLV	5.300	4.918	4.823	4.861
TRN	TRN	VLD	PDDA-3	IDWKKIFEKVKNLV	5.070	4.918	4.823	4.861
TRN	TRN	TRN	PDDA-4	IDWKKIFEKVKDLV	4.750	5.013	4.866	4.970
CLB	TRN	CLB	PDDA-5	INWSKIFEKVKNLV	4.760	4.664	4.620	4.596
TRN	VLD	CLB	PDDA-6	INWSSIFEKVKNLV	4.390	4.504	4.460	4.441
VLD	CLB	VLD	PDDA-7	INWSSIFESVKNLV	4.050	4.344	4.301	4.285
CLB	TRN	TRN	PDDA-8	INWSSIFESVSNLV	4.000	4.184	4.141	4.129
TRN	CLB	CLB	PDDA-9	INWKKIFEKVSNLV	4.640	4.664	4.620	4.596
TRN	TRN	TRN	PDDA-10	INWKKIFESVKNLV	5.000	4.664	4.620	4.596
TRN	VLD	CLB	PDDA-11	INWKSIFEKVKNLV	4.820	4.664	4.620	4.596
TRN	VLD	TRN	PDDA-12	NIWKKIFEKVKNLV	4.690	4.823	4.780	4.752
VLD	CLB	CLB	PDDB	INWLKLGKKILGAL	4.800	4.754	4.688	4.743
CLB	TRN	CLB	PDDB-1	INWLKLGKKILGAI	4.600	4.650	4.632	4.697
TRN	TRN	TRN	PDDB-2	INWLRLGRRILGAL	4.760	4.795	4.799	4.794
TRN	CLB	VLD	PDDB-3	INFLKLGKKILGAL	4.640	4.651	4.571	5.032
TRN	TRN	TRN	PDDB-4	INWKKLGKKILGAL	4.800	4.601	4.555	4.586
VLD	TRN	CLB	PDDB-5	INSLKLGKKILGAL	4.000	4.168	3.999	4.463
VLD	VLD	TRN	PMM	INWKKIASIGKEVLKAL	4.580	4.675	4.604	4.649
CLB	TRN	TRN	PMM-1	INWKKIASIGKEVLKAI	4.560	4.572	4.548	4.603
TRN	VLD	VLD	PMM-6	INWKKIASIGKEVLKA	4.150	4.435	4.395	4.404
VLD	VLD	CLB	PMM-7	INWKKIASIGKEVLK	4.000	4.179	4.172	4.191
TRN	TRN	TRN	PMM-8	INWKKIASIGKEVL	4.000	4.092	4.095	4.103
TRN	TRN	TRN	PMM-9	INWKKIASIGKEV	4.000	3.851	3.886	3.859
VLD	TRN	VLD	PMM-10	IWNKIAKSIGKVLEKAL	4.460	4.675	4.604	4.649
TRN	TRN	TRN	PMM-12	NIWKKIASIAKEVLKAL	5.000	4.868	4.839	4.830
VLD	TRN	TRN	PMM-13	KNWKKIASIGKEVLKAL	4.220	4.626	4.527	4.538
TRN	TRN	TRN	PMM-14	SNWKKIASIGKEVLKAL	4.520	4.466	4.368	4.383

TRN: trainig set; CLB: calibration set; VLD: validation set.