European Journal of Pharmaceutical Sciences 47 (2012) 273–279

Contents lists available at SciVerse ScienceDirect

European Journal of Pharmaceutical Sciences

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

Chemoinformatics in anti-cancer chemotherapy: Multi-target QSAR model for the in silico discovery of anti-breast cancer agents

Alejandro Speck-Planche [a,*], Valeria V. Kleandrova [b], Feng Luan [a,c], M. Natália D.S. Cordeiro [a,*]

[a] REQUIMTE/Department of Chemistry and Biochemistry, University of Porto, 4169-007 Porto, Portugal
[b] Faculty of Technology and Production Management, Moscow State University of Food Production, Volokolamskoe shosse 11, Moscow, Russia
[c] Department of Applied Chemistry, Yantai University, Yantai 264005, People's Republic of China

ARTICLE INFO

Article history:
Received 8 February 2012
Received in revised form 22 March 2012
Accepted 9 April 2012
Available online 17 April 2012

Keywords:
Anti-BC activity
BC cell lines
Linear discriminant analysis
mt-QSAR
Quantitative contributions
In silico design

ABSTRACT

The discovery of new and more efficient anti-cancer chemotherapies is a field of research in expansion and growth. Breast cancer (BC) is one of the most studied cancers because it is the principal cause of cancer deaths in women. In the active area for the search of more potent anti-BC drugs, the use of approaches based on Chemoinformatics has played a very important role. However, until now there is no methodology able to predict anti-BC activity of compounds against more than one BC cell line, which should constitute a greater interest. In this study we introduce the first chemoinformatic multi-target (mt) approach for the in silico design and virtual screening of anti-BC agents against 13 cell lines. Here, an mt-QSAR discriminant model was developed using a large and heterogeneous database of compounds. The model correctly classified 88.47% and 92.75% of active and inactive compounds respectively, in training set. The validation of the model was carried out by using a prediction set which showed 89.79% of correct classification for active and 92.49% for inactive compounds. Some fragments were extracted from the molecules and their contributions to anti-BC activity were calculated. Several fragments were identified as potential substructural features responsible for anti-BC activity and new molecules designed from those fragments with positive contributions were suggested as possible potent and versatile anti-BC agents.

1. Introduction

Breast cancer (BC), also known as malignant breast neoplasm, is a type of cancer originating from breast tissue, and constitutes the most common cancer as well as the leading cause of cancer deaths in women (14%) around the world (Roses, 2005). In 2006, more than 214,000 new BC cases were diagnosed, and it was estimated that close to 50,000 women died of the same disease in that year (Gradishar and Wood, 2008). Important advances have been realized for the search of new targets for anti-BC chemotherapy (Han et al., 2008; Lipkowitz, 2003; Sim et al., 2008), and more than 2000 compounds have been tested for anti-BC activity (EBI-Team, 2010). However, the search for new and more efficient anti-BC chemotherapies still remains a challenge for the scientific community.

Chemoinformatics based approaches have been essential in drug design (Gasteiger, 2003), preventing and/or rationalizing the serendipitous discovery of compounds with dissimilar biological activities (Kubinyi, 1993). In the field related with the medicinal chemistry and current pharmaceutical design of more effective anti-BC agents, Chemoinformatics has played a very important role (Liao et al., 2009; Markovic et al., 2011; Munteanu et al., 2009; Nagar et al., 2008; Nicolle et al., 2009; Pick et al., 2011; Roy and Roy, 2010; Vilar et al., 2009). However, the majority of these computational approaches employ for the discovery of anti-BC agents, small and homogeneous databases of compounds, studying usually only one biomolecular target like a protein or cell line associated with BC. This element prevents the exploration of structural patterns which could be related with development of anti-BC activity. If the compound shows anti-BC activity against one specific BC cell line, this does not convert the compound in an anti-BC agent. On the other hand, the fact that a compound with anti-BC activity can act through a defined mechanism of action may be one of the main causes of resistance by the different BC cell lines (Rivera and Gomez, 2010). In this sense, the discovery of highly active compounds against several important and well studied BC cell lines should constitute a major goal, in order to avoid probably, the prevalence of resistance to anti-BC drugs. For this reason, the first step should be the development of a computer-aided methodology to achieve the prediction of anti-BC activity of compounds against several of the most studied BC cell lines. In an attempt to overcome this problem, we introduce the first chemoinformatic multi-target (mt) approach for the virtual screening, prediction and in silico design of potent and versatile anti-BC compounds

* Corresponding authors. Fax: +351 220402659.
E-mail addresses: alejspivanovich@yahoo.es (A. Speck-Planche), ncordeir@fc.up.pt (M.N.D.S. Cordeiro).

0928-0987/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ejps.2012.04.012

against several BC cell lines. This approach allows the computer-aided drug discovery of anti-BC agents against 13 BC cell lines by employing an mt-QSAR model.

2. Materials and methods

2.1. Molecular descriptors

In the last 20 years, the quantity of molecular descriptors has been increased in considerable way (Todeschini and Consonni, 2000). More than 4000 descriptors have been reported in the literature (Todeschini and Consonni, 2009). In this work, we selected the descriptors known as functional group counts (FGC) and atom-centered fragments (ACF). These descriptors have been successfully applied in several QSAR studies (Speck-Planche et al., 2011a,b, 2009). They encode important information about specific fragments or functional groups in the molecules (Talete-srl, 2005), such as their abilities to participate in hydrophobic and dispersive interactions (Viswanadhan et al., 1989), or to exhibit a defined chemical reactivity (Speck-Planche et al., 2011a,b, 2009). These fragment-based descriptors, mentioned above, have some similarities with the variables used in a Free-Wilson analysis (Kubinyi, 1993).

We also selected the topological/substructural descriptors known as spectral moments of the bond adjacency matrix (μ_k) which are the essence of the TOPS-MODE (**TOP**ological **S**ubstructural **MO**lecular **DE**sign) approach. These descriptors have been widely employed in QSAR studies (Estrada and Peña, 2000a,b; Perez Gonzalez et al., 2003), and for the assessment of dissimilar toxicological profiles (Estrada et al., 2003; Helguera et al., 2007; Helguera et al., 2008). Basically, μ_k descriptors encode the molecular structure by mean of the bond adjacency matrix $\boldsymbol{B}$, which is a square table of order m (the number of chemical bonds in the molecule). The elements of this matrix (e_{ij}) are equal to 1 if bonds i and j are adjacent (which means that i and j are incident in the same vertex or atom) and 0 otherwise. In mathematical terms (Estrada, 1996, 1997, 1998), the μ_k is the sum of the main diagonal elements (e_{ii}) of the matrix $\boldsymbol{B}^k$:

$$\mu_k = \mathrm{Tr}(B^k) = \sum_{i=1}^{s} (e_{ii})^k \quad (1)$$

where Tr means the trace of the matrix, i.e., the sum of the diagonal entries of the matrix. The elements (e_{ii}) are bond weights which represent different physicochemical properties such as dipole moments, standard distances, or mathematical expressions involving atomic weights such as polarizability, polar surface and hydrophobicity. An important aspect about μ_k descriptors is the possibility to calculate the quantitative contribution of any fragment to the desired activity (Estrada and Peña, 2000a).

2.2. Data set: calculation of the descriptors and development of the mt-QSAR model

The data set was formed by 865 compounds with anti-BC activity against 13 BC cell lines (EBI-Team, 2010). Not all the compounds were tested against all the BC cell lines. We had also 102 drugs extracted from the Merck Index. These drugs exhibit other profiles which do not include anti-BC activity and for this reason they have been used as inactive (O'Neill et al., 2006). The FGC and ACF were calculated using DRAGON v5.3 (Talete-srl, 2005). The μ_k descriptors (from order 1 to 15), were calculated using MODESLAB v1.5 (Estrada and Gutiérrez, 2002–2004). In this case, the spectral moments were weighted by physicochemical properties such as dipole moment, molar refractivity, and the Abraham term related with the dipolarity/polarizability relationship. Linear discriminant analysis (LDA) was used to construct the classifier model (van de Waterbeemd, 1995). The essential element in this study was the organization of a spreadsheet containing the data used as input for the LDA, taking into consideration that we were not dealing with a classical LDA problem. For this reason we employed one of the diverse mt-QSAR methodologies which have been widely applied by González-Díaz and coworkers in drug design (Prado-Prado et al., 2011, 2008; Vina et al., 2009). In this specific case we employed a similar methodology to that developed for the prediction of enzyme classes from 3D structure in *Leishmania infantum* (Concu et al., 2009). The same methodology has been generalized to the prediction of GSK-3 inhibitors (Garcia et al., 2011), and a detailed description of the procedure applied here, was recently published in the work of Speck-Planche and coworkers for the assessment of resistance risk in agrochemical fungicides (Speck-Planche et al., 2011b). Here, we formulated a binary discriminant function for the classification of compounds: those which belonged to a particular active group (inhibitory activity against a specific BC cell line) and compounds that did not belong to this group (inactive). For this, we employed an approach to define the groups predicted in each case. Thus, we used the following steps:

- First, the 865 compounds belonging to the group of compounds with anti-BC activity were divided according to their activity against the 13 BC cell lines. Each BC cell line had a cutoff value of anti-BC activity (Table 1), in which the compounds were considered as active. The variable of activity selected was the IC_{50}, which measures the cytotoxic activity as the ability of the compounds to inhibit at 50% the proliferation of the different BC cell lines (EBI-Team, 2010).
- We created a data file by assigning to each compound 286 structural variables (inputs): 97 FGC, 54 ACF and 135 descriptors like μ_k. We had also one output variable and one classification variable related with the type of BT cell line (BCCL). This last variable is an auxiliary variable which was not used to construct the model. Thus, the row for each compound input contained 288 elements in total.
- The output variable is a dummy variable (Boolean) called anti-BC activity (A_{BC}); $A_{BC} = 1$ if the compound has anti-BC activity against any of the 13 BC cell lines and -1, otherwise ($A_{BC} = -1$). The last case corresponded to the 102 drugs which were considered as inactive. We can repeat each of these 102 drugs more than one time in the raw data. In fact, we repeated each compound 13 times corresponding to the 13 BC cell lines. In this work, we used only the BCCL code to relate a compound with its corresponding type of BC cell line, and thus, these compounds entered only once. Conversely, inactive compounds (decoys) have more than one line entry with different BCCL classes.

Table 1
Cutoff values for anti-BC activity in different cell lines.

BC cell line	Cutoff[a]
MCF7	⩽0.001
MDA-MB-231	⩽0.050
MDA-MB-435	⩽0.198
MDA-MB-468	⩽9.100
MX1	⩽11.00
SK-BR-3	⩽0.050
T47D	⩽2.000
Bcap37	⩽45.50
BT-20	⩽7.187
BT-474	<24.66
BT-549	<45.51
MCF7R	⩽0.762
MCF7S	⩽6.825

[a] IC_{50} values are expressed in μM.