Journal of Steroid Biochemistry & Molecular Biology 138 (2013) 348–358

Contents lists available at ScienceDirect

Journal of Steroid Biochemistry and Molecular Biology

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

Integration of ligand and structure-based virtual screening for identification of leading anabolic steroids

Yoanna María Alvarez-Ginarte[a,*], Luis Alberto Montero-Cabrera[a], José Manuel García-de la Vega[b], Alberto Bencomo-Martínez[c], Amaury Pupo[c], Alina Agramonte-Delgado[d], Yovani Marrero-Ponce[e], José Alberto Ruiz-García[f], Hans Mikosch[g]

[a] Laboratory of Theoretical and Computational Chemistry, Faculty of Chemistry, University of Havana, 10400 La Habana, Cuba
[b] Department of Applied Physical Chemistry, Autonomous University of Madrid (UAM), 28049 Madrid, Spain
[c] Center of Molecular Immunology, 11600 La Habana, Cuba
[d] University of Informatics Sciences, La Habana, Cuba
[e] Unit of Computer-Aided Molecular "Biosilico" Discovery and Bioinformatic Research (CAMD-BIR Unit), Faculty of Chemistry-Pharmacy, Central University of Las Villas, Santa Clara, 54830 Villa Clara, Cuba
[f] Biomolecular Chemistry Center, 16042 La Habana, Cuba
[g] Institut für chemische Technologien und Analytik, Vienna University of Technology, TU Wien – Karlspl. 13/164, A-1040 Wien, Austria

ARTICLE INFO

Article history:
Received 29 January 2013
Received in revised form 4 June 2013
Accepted 8 July 2013

Keywords:
Anabolic steroids
Virtual screening
Quantum and physicochemical molecular descriptor
Cluster analysis
QSAR and docking studies

ABSTRACT

Parallel ligand- and structure-based virtual screenings of 269 steroids with anabolic activity evaluated *in vivo* were performed. The quantitative structure–activity relationship (QSAR) model expressed by selected descriptors as the octanol–water partition coefficient, the molar volume and the quantum mechanical calculated charge values on atoms C1, C2, C5, C9, C10, C14 and C17 of the steroid skeleton, expresses structural features of anabolic steroids (AS) contributing to the transport and steroid–receptor interaction. On the other hand, computational simulations of a candidate ligand binding to a receptor study (a "docking" procedure) predict the association of these AS with the human androgen receptor (AR). Fourteen compounds were identified as lead; the most potent was the 7α-methylestr-4-en-3, 17-dione. It was concluded that a good anabolic activity requires hydrogen bonding interactions between both Arg752 and Gln711 residues in the cycles A with O3 atom of the steroid and either Asn705 and Thr877 residues in the cycles D of steroid with O17 atom.

1. Introduction

Drug design is the form of finding medications by means of designs based on its biological actions. The drug is a small organic molecule that activates or inhibits the function of a protein [1,2]. Virtual screening (VS) is a key component in the process of drug design and development. It is normally regarded as the selection of likely drug candidates from large libraries of chemical structures by using computational methodologies. However, the generic definition of VS is wider and may encompass many different methods. There are two broad categories of screening techniques: the ligand-based drug design (LBVS) and the structure-based (SBVS) [3–5].

Ligand-based drug design (indirect drug design) is based in the analysis and comparison of molecular properties and the biological activity for well-known molecules without keeping in mind the structure of the biological receptor responsible for the pharmacological activity [6–8].

In the case of structure-based drug design (direct drug design) when we have the three-dimensional structure of the receptor obtained by experimental methods (crystallography of rays X or NMR experiments) or through the construction of molecular models, we can approach the drug design on the base of the structure. If we do not have the structures, we can still use these technical if we obtain models for homology of enough trust [8–10].

In the peculiar case of anabolic-androgenic steroid (AAS), the so called androgen receptor (AR) [11] has been isolated and together with the existence of a data base that reports the anabolic activity of more than 200 steroids [12,13] it is possible to perform both QSAR and docking studies in these molecules. Anabolic steroids are drugs which mimic the effects of the male steroids testosterone

* Corresponding author. Tel.: +53 7 8738222; fax: +53 7 8733502.
E-mail addresses: yoanna@fq.uh.cu, alvarezmaria.alvarez1@gmail.com (Y.M. Alvarez-Ginarte).

0960-0760/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jsbmb.2013.07.004

(Testo) and dihydrotestosterone (DHT). They increase protein synthesis within cells, which results in the building-up of cellular tissue (anabolism), especially in muscles. Anabolic steroids also have androgenic and virilising properties, including the development and maintenance of masculine characteristics such as the growth of the vocal cords and body hair [14]. They were first isolated, identified and synthesized in the 1930s, and now are used therapeutically in medicine to stimulate bone growth and appetite, induce male puberty, and treat chronic wasting conditions, such as cancer and acquired immune deficiency syndrome or acquired immunodeficiency syndrome (AIDS) [15–17].

In the case of AS, the mechanisms of action differ depending on their molecular specificity. Different types of anabolic steroids bind to the AR with different affinities, depending on their chemical structure [18]. The effect of anabolic steroids on muscle mass is caused in at least two ways: first, they increase the production of proteins; second, they reduce recovery time by blocking the effects of stress hormone cortisol on muscle tissue, so that catabolism of muscle is greatly reduced. It has been hypothesized that this reduction in muscle breakdown may occur through anabolic steroids inhibiting the action of other steroid hormones called glucocorticoids that promote the breakdown of muscles [19].

In an effort for dissociating the anabolic activity from sometimes unwanted androgenic effects; many anabolic-androgenic steroids have been created. But the synthesis of a compound which demonstrates a complete dissociation between anabolic and androgenic properties has not been achieved to date [12]. Our investigation group has reported several articles about the specific QSAR models for different families of steroids. These studies allowed better structural interpretations of which are the molecular features that influence in the activity anabolic or androgenic activities of these steroids, as well as, to identify new lead and hit molecules of anabolic steroids. These models were reported for the congeneric series of AAS: 17β-hydroxy-5α-androstane [20], 4,5α-dihydrotestosterone [21], testosterone [22] and 19-nortestosterone [23].

We also publish a general predictive linear discriminant analysis (LDA)-assisted QSAR model, enabling selection of novel drug-like steroids with good anabolic to androgenic ratio (AAR) [12]. In this study, the considered AAS were derivative molecules of the congeneric series selected in the specific QSAR models reported by our group. These steroids have diverse structural patterns and AAR behavior. If the anabolic ratio is greater than one, it indicates a higher trend toward anabolic effect and, therefore can classify the drug as an anabolic steroid. A measure lower than one, in turn, assesses the steroid as androgenic. However, using this rate a steroid can also be classified as a good anabolic if it has high both values of anabolic and androgenic activity. For this reason it is necessary to have a classification QSAR model for the anabolic activity only, which allows making a correct classification to identify steroids with a high anabolic activity and a low androgenic effect. Therefore the first objective of our work is (1) obtain general predictive QSAR models for the anabolic activity of steroids with diverse structural patterns and anabolic activity behavior.

The combination of obtained results with both the QSAR and docking computational methods can be very promising in the search of lead AAS because the structure-based drug design can be carried out to complement or to go beyond QSAR studies.

For this reason our second objective is (2) obtaining the best conformation in terms of predicted interactions with the human AR by docking simulations of 269 AS, to help to interpret the already elucidated structure–activity relationships in this chemical series. We are taking an advance step in the search of a quick and effective methodology to discover new AS.

2. Methods

2.1. Data set of anabolic steroids

To obtain a QSAR model that classifies to the steroids with good or bad anabolic activity we use the chemical information contained in a great number of compounds with experimental anabolic activity used to establish a QSAR equation. We selected a group of 269 steroids [12,13]. The anabolic activity of each steroid was evaluated *in vivo* as well as collected and reported by Vida in 1969 [13]. Until now, there is no other standardized reference, where the values of the anabolic and androgenic activities find been reported for this kind of molecules.

The details of the *in vivo* methods used to determine the biological activity experimental were ex-plained by Vida [13]. In the determination of the anabolic activity the levator ani muscle (LA) was isolated from each rat. This organ was weighted and the activity was calculated by the differences in weight of the levator ani muscles between the control and active groups [12,13].

The molecular structures and their experimental values of biological activities for the steroid of the data set are shown in the Supporting Information (Tables SI1 and SI2, respectively).

2.2. The estimation of molecular properties

A molecular descriptor (molecular properties) can be defined as the final result of a logical and mathematical procedure that transforms the information coming from certain properties of the molecules (so much theoretical as experimental) that is coded inside a symbolic representation in an useful number to predict, by means of statistical correlations, the result of a certain experiment or appraisable property.

The specific action of the drugs is determined by its properties, among those that are the hydrophobic, electronic and steric. As hydrophobic descriptor, we estimate the decimal logarithm of the octanol–water partition coefficient (*Log P*), related with the transport of the drugs through the biological membranes [24]. The electronic and steric properties determine the interaction of the drug with the receptor [25,26]. The estimated electronic descriptors were: hydration energy, E_{H_2O} [27], polarizability, P [28], electric dipole moment, μ, heat of formation, ΔH_f, electronic energy, EE, total energy, E_T, e_{HOMO} energy, e_{LUMO} energy, net atomic charges of C atoms in the steroid backbone, q_1 to q_{19} [29], electrophilicity index, ω, chemical potential, U, chemical hardness, η and chemical softness, S [30]. Each molecule was optimized geometrically and the electronic descriptors were calculated using the MOPAC v. 6 program [31] by the parametric method PM3 semi-empirical Hamiltonian [32].

As steric descriptor we calculated: approximate surface area, *ASA*, grid surface area, *GSA*, molar volume, *VM* [33,34], and molar refractivity, *MR* [24]. In the Supporting Information (Table SI3) are given the MDs calculated in this study.

2.3. Cluster analysis (CA) for selection of the training and test series

The analysis of clusters (CA) is constitutes by a group of technical where the objects (molecules) or cases are classified in groups relatively homogeneous calls clusters. The CA is used in QSAR studies to design the training and test series, as well as to demonstrate the structural diversity of the used database [35].

This method includes several different classification algorithms and it allows organizing the significant data observed in the structures.

Many algorithms of AC have been created they belong to two categories: the hierarchical AC and the non-hierarchical. The