Steroids 96 (2015) 81–88

Contents lists available at ScienceDirect

Steroids

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

Structural classification of steroid-binding sites on proteins by coarse-grained atomic environment and its correlation with their biological function

Yasuha Hori-Tanaka [a], Kei Yura [b,c,d,*], Takako Takai-Igarashi [a,e], Hiroshi Tanaka [a,e]

[a] Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, Japan
[b] Graduate School of Humanities and Sciences, Ochanomizu University, 2-1-1, Otsuka, Bunkyo-ku, Tokyo 112-8610, Japan
[c] Center for Informational Biology, Ochanomizu University, 2-1-1, Otsuka, Bunkyo-ku, Tokyo 112-8610, Japan
[d] National Institute of Genetics, 1111, Yata, Mishima, Shizuoka 411-8540, Japan
[e] Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi 980-8573, Japan

ARTICLE INFO

Article history:
Received 14 September 2014
Received in revised form 22 December 2014
Accepted 19 January 2015
Available online 30 January 2015

Keywords:
Function-unknown protein
Principle component analysis
Side effect
Structural bioinformatics
Three-dimensional structure

ABSTRACT

Steroid hormone is extensively used for transmitting variety of biological signals in organisms. Natural steroid hormone is synthesized from cholesterol in adrenal cortex and in sexual gland in vertebrates. Appropriately dosed synthetic steroid hormones can be used for medication. Despite their positive effects as medicine, they sometimes cause significant side effects due to their wide range of actions, and the studies for discovering the mechanisms of side effects were carried out aiming to reduce the side effects. The fundamental cause of the side effects seems to be interactions between the steroid and a non-target protein. To understand the possible range of interaction of steroid molecule, we gathered all the three-dimensional structures of protein–steroid complex determined by X-ray crystallography, compared the atomic environments of the steroid-binding sites in proteins and classified the pattern of steroid binding. Protein Data Bank contained 871 structures of steroid–protein complexes in 382 entries. For this study, we selected 832 steroid binding proteins. Using a newly developed method to describe the atomic environments of these steroid molecules and their function, we were able to separate the environments into six patterns. This classification had a potential to predict the function of function-unknown proteins with a co-crystallized steroid molecule. We speculated that the proteins grouped into the same pattern of nuclear receptors were the candidates of non-targeted proteins causing a side effect by a therapeutic prescription of steroid hormone.

1. Introduction

Steroid is defined as a compound based on the cyclopentaphenanthrene carbon skeleton with methyl groups at C-10 and C-13, and often an alkyl group at C-17 [1]. Steroids are widely used as a signaling molecule in virtually all eukaryotes and these steroids are specifically named steroid hormones. Steroid hormones are further categorized to subgroups including glucocorticoids, mineralocorticoids, oestrogens, androgens, and progestogens [2]. All these steroid hormones are the derivative of cholesterol, a structural component of a membrane. The biological roles of these steroid hormones range from a wide aspect of metabolism, immune function, reproduction, and sex difference [3,4]. Prescribed synthetic steroids can affect these functions, and hence, a steroid hormone has been used for therapeutic drug for a long time [5].

Among the steroid molecules, glucocorticoid is the most widely prescribed steroid for its versatility against a wide range of diseases, and hence the amount of usage has been continuously increasing [4,6]. For the drug usage, a glucocorticoid molecule is synthesized as an inactive form and is transformed to an active form by hydroxylation when it binds to glucocorticoid receptor (GR), one of the nuclear receptor proteins. In the normal cells, the activated GR forms a dimer and transferred to nucleus where GR functions as a transcription factor, and activates the transcription of other genes. In the therapeutic process, however, the GR binds to NFκB or AP1 transcription factors and inhibit the transcription of the genes regulated by those transcription factors [5,6]. Glucocorticoid functions as a molecule to repress genetic switches and eased the effect of disease.

* Corresponding author at: Center for Informational Biology, Ochanomizu University, 2-1-1, Otsuka, Bunkyo-ku, Tokyo 112-8610, Japan. Tel./fax: +81 3 5978 5514.
E-mail address: yura.kei@ocha.ac.jp (K. Yura).

http://dx.doi.org/10.1016/j.steroids.2015.01.015
0039-128X/

A number of side effects have been surfaced in accordance with the increase in the usage of glucocorticoid in therapy [5], and the studies for discovering the mechanisms of the side effect were carried out for a safe and effective usage of glucocorticoid [6,7]. One of the well-studied mechanisms of the side effect in glucocorticoid therapy involves a cross talk with mineralocorticoid receptor (MR). Glucocorticoid has affinity toward MR and the association activates MR as a transcription activator. MR is expressed in selected tissues, such as epithelial cells in the kidney, colon, and salivary glands and non-epithelial cells in the brain and heart where side effects are observed [8]. Another mechanism of the side effect, which is called non-genomic mechanism, involves protein not related to nuclear receptors. In hippocampus, liver and blood cells, for instance, glucocorticoids presumably bind to membrane proteins and cause unintended responses [9,10]. In both cases, the side effect is evidently caused by the interactions between the steroid drug and non-target proteins.

Interactions between steroid receptors and steroid hormones have been extensively studied from the viewpoints of protein structure and evolution [11]. The study demonstrated that the evolution of the interactions started with oestrogen molecule, then the receptor was duplicated to interact with other related steroids. These studies provided a significant insight to the interactions between steroid hormone receptors and steroid hormones. However, the side effect of therapeutic usage of the steroids seems to stem from the interactions between the steroid hormone and different types of proteins, hence studies on the interactions between steroids and many types of proteins are required toward the understanding of the mechanism of the side effect.

Progress in the field of structural biology has significantly increased the number of data for the coordinates of atoms in proteins. These data for structures of proteins are mostly obtained by X-ray crystallography, and the proteins are often crystallized with small molecules including steroid molecules [12]. We, therefore, searched for the known interaction patterns between steroid molecules and proteins in the protein three-dimensional (3D) structure database and classified the interaction patterns. The classification identified all the known patterns of interactions between proteins and steroid hormones. In addition, the classification resulted in finding proteins that interact with steroid hormone in a similar way to nuclear receptors. In this paper, we report these interaction patterns, which may provide novel insights to the design of a therapeutic steroid hormones with less side effect.

2. Experimental

2.1. Molecules with a steroid backbone from Protein Data Bank

Entries including both protein and a molecule with a steroid backbone were extracted from Protein Data Bank [13]. We defined the steroid backbone based on the description in IUPAC Compendium of Chemical Terminology – the Gold Book [1], which states "a molecule with three six-membered rings and one five-membered ring structure, made of 17 carbon atoms, and one of the six-membered rings shares one covalent bond with another six-membered ring which shares a different covalent bond with the other six-membered ring, and both covalent bonds are separated by one and only one covalent bond. The five-membered ring attaches to one of the flanking six-membered rings sharing one and only one covalent bond. The other flanking six-membered ring is located *trans* to the five-membered ring." We built an in-house program that resolved the definition above using a chemical bond connectivity matrix (The program can be shared on request). After retrieving PDB entries with the molecule fulfilling the above definition, we compared the 3D structures of the molecules against an oestradiol bound to human 17-beta-hydroxysteroid-dehydrogenase (PDB ID: 1FDT), an arbitrarily chosen steroid molecule in PDB, to ensure that the automatically selected molecules had the same 3D structures to the steroid backbone. The correspondence of atoms in the two molecules for the comparison was determined based on the chemical bond connectivity matrix, too. The detail of method involving the matrix and the application of the method to other small molecules will be explained elsewhere. A molecule with a large deviation from the oestradiol backbone structure was discarded. It should be noted that we did not use a representative set of proteins in PDB intentionally. This is because there were cases that the similar proteins interact with different small molecules with different interaction patterns.

2.2. Detecting interactions between steroid and protein

The interaction between the steroid backbone and a protein residue was defined based on the contact between atoms. When two heavy atoms were within the distance of 5.0 Å, the atoms were considered to be in contact, hence the atoms interact. This cutoff distance maybe too long, when hydrogen bond interaction is considered. Yet we tried to take other interactions such as electrostatic and hydrophobic interaction into account and set the cutoff distance longer. Heavy atoms were defined as carbon, oxygen, nitrogen, or sulfur atoms, which were normally observed by X-ray crystallography measurement. An amino acid residue with the interacting atom to any heavy atoms of the steroid backbone was considered as an interacting residue. A steroid molecule without any interaction to a protein was discarded from the data set.

2.3. Comparison of the steroid-binding site in three-dimensional structures

Locations of atoms in different proteins around the same steroid backbones were compared. Heavy atoms from proteins that interact with a heavy atom of the steroid backbone were considered. The locations of these atoms in different coordinate systems were described by spanning a common local coordinate around steroid molecules as shown in Fig. 1. C-9 carbon atom (or the corresponding atom) in the steroid backbone was defined as the origin of the coordinates. A vector from the origin to C-14 carbon atom (or the corresponding atom) in the steroid backbone was then defined as a direction of *X*-axis and a vector from the origin to C-11 carbon atom was defined as a tentative *Y*-axis. The *X*- and the tentative *Y*-axes were not necessarily orthogonal, but the cross product of the two axes is orthogonal to both axes and defined as *Z*-axis. The real *Y*-axis was then defined as a cross product of the *Z*- and *X*-axes (Fig. 1). In this way, the environment of the same steroid molecule in different proteins was set to the common coordinate system and could be compared. A regular grid of 3 Å interval in three-dimension was placed so that the axes of the *XYZ* coordinates were either parallel or perpendicular to the grid. We have tested several sizes for the interval and found that 3 Å was the best for the analyses here. Each atom in the space was then assigned to one of the nearest cubicles built by the grid. A digit 1 was assigned to a cubicle with one or more atoms and zero otherwise. All the cubicles were, then, numbered systematically in one dimension, hence the grid space could be transformed to vector v. The similarity of two grid spaces i and j, derived from two different protein–steroid interactions, was measured with score s calculated by the following equation:

$$S_{ij} = \sum_{k=1} (v_k^i \otimes v_k^j), \quad (1)$$

where $\otimes$ indicates the exclusive nor operation, which means that when v_k^i and v_k^j are the same, the operation returns 1, and 0