ELSEVIER

Journal of Molecular Structure: THEOCHEM 823 (2007) 16–27

THEO CHEM

www.elsevier.com/locate/theochem

Identification of the bioactive conformation for mucin epitope peptides

Adrian Kalaszi, Gabor Imre, Imre Jakli, Odon Farkas *

Laboratory of Chemical Informatics, Institute of Chemistry, Eötvös Loránd University, 1/A Pázmány Péter sétány, 1117 Budapest, Hungary

Received 28 June 2007; received in revised form 30 July 2007; accepted 9 August 2007
Available online 29 August 2007

Abstract

Most of the drug molecules exhibit their biological activity through binding to the target protein. When the 3D structure of the binding site is unknown, pure ligand-based approaches are often used to perceive the 3D pharmacophore. However, dealing with conformational flexibility of ligands in such methods is still in the frontline of the current research. The special thermodynamic properties of the binding of flexible molecules, as derived here, show that the probability of the bioactive conformations in solution can determine the likelihood of binding. The binding activities can be obtained experimentally, while the probability of conformations in solution can be computed. Our present paper discusses the thermodynamic basis of performing 3D QSAR studies on molecules, with considerable conformational flexibility. In addition, we supply an algorithm to locate the bioactive conformations. The work is initiated to find the binding conformation of the therapeutically promising mucin epitope pentapeptides.

Keywords: Conformational flexibility; Binding free energy; Molecular dynamics; Conformational search; Drug design; Bioactive structures; Rational library design

1. Introduction

The mucins, a family of high molecular weight glycoproteins, may be overexpressed and partially underglycosylated in carcinomas of gastrointestinal cancer [1]. Where the carbohydrate coating is deficient, the protein core exposes. This property of mucins of tumor tissues may serve as a future marker of such disease. Among other MUC2 protein specific antibodies, the mAb 994 was developed against the VTPTPTPTGTQTPT peptide sequence. This sequence was previously identified in the uncovered protein core. Binding experiments revealed that mAb 994 recognizes pentapeptides, which are comprised of a common structural TX^1TX^2T motif, as the minimal epitopes [2]. A systematic binding study was performed using combinatorial library and direct synthetic approach, where all natural aminoacids except Cys was tested as "X" [3]. The most effective binders were identified in the TQTXT sub-library. Besides the TQTPT sequence, which presents in the native protein, five other peptides were proven active. These peptides are the TQTAT, TQTYT, TQTWT, TQTFT TQTST in decreasing order of activity. It is interesting to note that while TQTAT had considerable binding affinity, the TQTGT did not bind at all, and while TQTPT was the best binder, the amide C-terminal containing TQTPT-NH_2 was also inactive. Conformational effects are hypothesized to be the cause of the difference in binding activity of the different molecules. Our goal was to explain this effect and to build up a quantitative relationship between structure and activity for the congeneric pentapeptides above, which may improve the results of a future therapeutic vaccine design.

This series is rather challenging from a SAR viewpoint because of the low number of molecules and their high conformational flexibility, as the examined TQTXT motif contains 18 rotatable bonds. Numerous methods for structure–activity relationship and prediction that incorporates the conformational flexibility were developed during the past decade. Unfortunately, the applicability of most of

* Corresponding author. Tel.: +36 30 2553111; fax: +36 1 3722620.
E-mail addresses: adrian@chem.elte.hu (A. Kalaszi), farkas@chem.elte.hu (O. Farkas).
URL: http://organ.elte.hu/farkas (O. Farkas).

0166-1280/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.theochem.2007.08.016

the available 4D QSAR methods [4–13] were demonstrated on molecules with a few rotatable bonds. One of the first 4D QSAR methods was developed by Hopfinger and coworkers [4]. The fourth dimension stands for the conformational distribution, sampled by molecular dynamics, of each ligand molecules in the set. The algorithm fits each conformation by previously selected pharmacophore points, calculates the grid cell occupancy values, and uses PLS to create the model. This procedure promptly raises the question what happens if the molecule is 'too flexible' and some conformations cannot be readily overlaid due to too large differences in spatial arrangements. Vedani and coworkers designed a 5D QSAR method that mimics the induced fit and can take into consideration multiple conformations [6–9]. The authors usually report excellent internal and external validation values. However, 'the interactions of all conformations, orientations, and protonation-states are calculated toward all members of the receptor-model family' to locate the bioactive conformation and bioactive orientation. This concept can be efficiently used for molecules having limited number of flexible bonds, but may lead to combinatorial explosion for molecules that can only be represented by large number of diverse conformations and can adopt large number of common orientations.

The concept of 3 + 3D descriptors [14,15], introduced by Martinek et al. They examined the distance distribution of a priori selected pharmacophores of flexible ligands, e.g. ligands of the μ-opiate receptor. Contrary to ignoring the conditional probability of the separate distances at describing the conformation, they could reproduce the bound conformation and also present acceptable statistical validation. Bernard and coworkers presented the conformationally sampled pharmacophores [12,13], which are the 2D distribution of two selected internal coordinates, and used them to explain the biological activity of flexible opioid ligands. In both cases, the conformational distribution is modeled as a canonical ensemble, and the region for bioactive conformation is determined in the space of previously defined internal coordinates. While Bernard and coworkers examine the overlap of these distributions, Martinek and coworkers locate the region where the conformational density correlates to the binding affinity.

We have elaborated a new methodology that is nearly related to the latter two procedures, in order to handle structure–activity relationships for molecules with high flexibility. In Section 2, we describe the thermodynamics basis and consequences to build up the conformation – binding activity relationship. The algorithm, which provides the binding conformation, is presented in Section 3.

2. Theory

Consider the general process of binding of a flexible molecule to the active site of an enzyme or receptor. The binding free energy, $\Delta G_{\text{bind-tot}}$, is independent of the actual reaction path, so we may split it into two parts [16]. The first part can be the formation of the bioactive conformations (BC) in solution, denoted as ΔG_{conf}. The bioactive conformations are the possible conformations of the molecule while it is bound to the active site. The second part is the "rigid" binding of the BC to the active site, denoted as ΔG_{bind} (Fig. 1). Hence we may write Eq. (1):

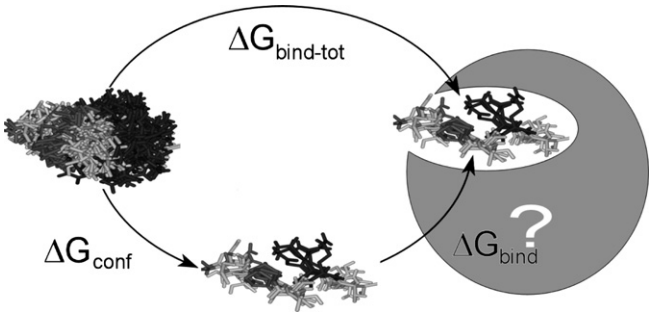

Fig. 1. From left to right: all possible conformations of a flexible molecule in solution, bioactive conformations (BC) in solution, complex of the protein and the small organic molecule.

$$\Delta G_{\text{bind-tot}} = \Delta G_{\text{conf}} + \Delta G_{\text{bind}}, \tag{1}$$

where ΔG_{conf} is the free energy required to form the BC, an assembly of conformations in solution, and ΔG_{bind} is the free energy required for binding the collection of bioactive conformations to the active site.

The concentration of the bound ligand $[l_{\text{bound}}]$ depends on the free energy change related to the complete binding process, while the concentration of BC in solution, $[l_{\text{BC}}]$ depends only on the ΔG_{conf} ((2) and (3)), where $[l_0]$ is the concentration of the most stable conformer in solution and β is the thermodynamic beta.

$$[l_{\text{bound}}] = [l_0]\mathrm{e}^{-\beta(\Delta G_{\text{conf}}+\Delta G_{\text{bind}})} \tag{2}$$

$$[l_{\text{BC}}] = [l_0]\mathrm{e}^{-\beta\Delta G_{\text{conf}}}. \tag{3}$$

The fraction of $[l_{\text{bound}}]$ and $[l_{\text{BC}}]$ depends only on the ΔG_{bind} value (4). If the quotient is expanded by the total concentration of the ligand in the solution, l_{tot}, the relative concentration or molfraction of the bound conformer is obtained in the numerator. We may replace the relative concentration of the bioactive conformation in solution, the denominator, by the proportion of the conformations in a simulated trajectory, as the quotient of the frequency of BC (n_{BC}) and the total number of steps, $n_{\text{traj-tot}}$, in a trajectory, which represents the full conformational distribution of the flexible molecule in solution.

$$\frac{[l_{\text{bound}}]}{[l_{\text{BC}}]} = \mathrm{e}^{-\beta\Delta G_{\text{bind}}} = \frac{[l_{\text{bound}}]/l_{\text{tot}}}{[l_{\text{BC}}]/l_{\text{tot}}} = \frac{[l_{\text{bound}}]/l_{\text{tot}}}{n_{\text{BC}}/n_{\text{traj-tot}}}. \tag{4}$$

Supposedly, we can collect a library of flexible molecules, where the members have very close ΔG_{bind} values.

$$\Delta G^i_{\text{bind}} \approx \Delta G^j_{\text{bind}}. \tag{5}$$

In such a case, as it follows from (4) and (5), the ratios of the relative concentrations of bound molecules for any