European Journal of Medicinal Chemistry 51 (2012) 271–276

Contents lists available at SciVerse ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Short communication

Metal–ligand interactions: An analysis of zinc binding groups using the Protein Data Bank

Kentaro Kawai, Naoya Nagata*

Central Research Laboratories, Kaken Pharmaceutical Co., Ltd, 14, Shinomiya Minamikawara-cho, Yamashina, Kyoto 607-8042, Japan

ARTICLE INFO

Article history:
Received 19 October 2011
Received in revised form
13 February 2012
Accepted 14 February 2012
Available online 22 February 2012

Keywords:
Metal–ligand interaction
Zinc metalloproteins
Protein Data Bank
Zinc binding group

ABSTRACT

In the present study, we investigated zinc binding groups (ZBGs) using the coordinates of protein–ligand complex structures obtained from the Protein Data Bank. The distance from the zinc to the nearest ligand atom was measured to determine whether the atom was part of the ZBG. The most frequently found ZBG was carboxylate, followed by sulfonamide, hydroxamate, and phosphonate/phosphate. Because it was found that few heteroatoms, such as nitrogen, oxygen, and sulfur atoms, interacted with zinc, ideal distances between the zinc and these heteroatoms were identified. Whereas carboxylates bound to the zinc via both monodentate and bidentate interactions, the hydroxamates bound dominantly in a bidentate manner. These results will aid in the design of new inhibitors with the potential to interact with zinc in the target protein.

1. Introduction

Metalloproteins have a metal ion or ions in their structure and have a wide variety of functions in vivo. Hemoglobins, which carry oxygen in humans, and CYP enzymes, which are involved in drug metabolism, both contain iron. DNA polymerases, which are responsible for DNA duplication, include magnesium, and arginase, which is involved in the final step in the urea cycle, contains manganese. Among these metalloproteins, zinc metalloproteins contain many attractive drug targets. For example, inhibitors of angiotensin-converting enzyme (ACE), carbonic anhydrases (CAs), matrix metalloproteinases (MMPs), TNF-α converting enzyme (TACE), histone deacetylases (HDACs), and farnesyltransferase have been reported [1–7]. Many inhibitors of zinc metalloproteins contain zinc binding groups (ZBGs), which play an important role in the activity of these inhibitors. For example, MMP3 inhibitors include hydroxamate (Fig. 1a) and barbiturate (Fig. 1b) groups [8,9]. Hydroxamate-based inhibitors typically inhibit human MMP3 by forming two hydrogen bonds with Glu202 and Ala165 in addition to the zinc. The barbiturate group also interacts with both of amino acids and the zinc ion.

Protein–ligand complex structures provide useful information for drug discovery. Virtual screening utilizing protein–ligand complex structures is the most common example. When multiple cocrystal structures for a certain target are available, many different approaches can be used to find novel active compounds. Ghose et al. prepared a knowledge base of aligned kinase cocrystal structures [10]. One of their objectives was to generate new molecules by alignment of common molecular frameworks and subsequent pruning and grafting of the structural moieties. Pierce et al. reported the BREED [11] approach, in which new molecules were generated by alignment of a pair of crystal ligand structures followed by swapping of the fragments on each side of each overlapping bond. We thought that the development of a list of ZBGs would be useful because such a list would provide a lot of information on the functional groups that bind to zinc. The analysis of the previously discovered ZBGs will aid in the rapid design of drugs that target enzymes, such as MMPs, farnesyltransferase, and HDACs. In addition, the analysis of the ZBGs will increase the understanding of the metal–ligand interactions. However, it is not clear how many and what types of functional groups can act as ZBGs. In the present study, we tried to identify ZBGs using protein–ligand complexes listed in the Protein Data Bank (PDB).

2. Methods

2.1. Identification of ZBGs

In this study, the atomic coordinates in each PDB file were used to identify ZBGs. If a pair of atoms is not bonded, it is theorized that

* Corresponding author. Tel.: +81 75 594 0787; fax: +81 75 594 0790.
E-mail address: nagata_naoya@kaken.co.jp (N. Nagata).

0223-5234/$ – see front matter © 2012 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2012.02.028

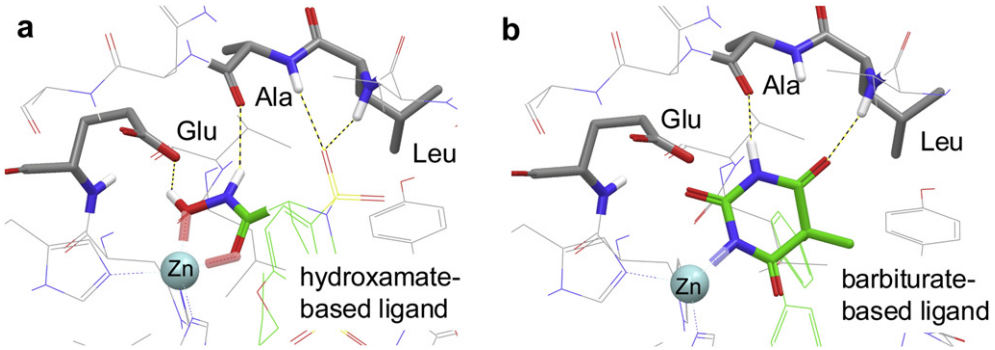

Fig. 1. Illustrative examples of ZBGs that bind a zinc ion and other amino acids in MMP3. The coordinates were taken from the PDB (1g49 and 1g4k). Dashed lines indicate hydrogen bonds. Proteins are gray, and ligands are green; only important parts of the molecules are shown in this stick representation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

these atoms can approach one another until the distance between the two is longer than the sum of van der Waals (vdW) radii of the atoms. If the pair of atoms forms a covalent bond or hydrogen bond or are linked by an attractive interaction, the distance between these two atoms is shorter than the sum of vdW radii of the corresponding elements. Therefore, we considered a ligand atom to be part of a ZBG only if the distance between the two atoms was shorter than the sum of vdW radii of the elements. For these calculations, elemental vdW radii were taken from the literature [12].

To determine whether the ligand includes a ZBG, the Euclidean distances between all pairs of zinc ions and ligand atoms were measured, and then, the shortest distance was identified. An illustrative example is shown in Fig. 2. In this case, the shortest Euclidean distance was 2.44 Å, which is shorter than the sum of the vdW radii of oxygen and zinc. Therefore, the oxygen atom of this ligand (a hydroxamate moiety) was considered to be a part of the ZBG.

2.2. Development of the list of ZBGs

The workflow of the study is shown in Fig. 3. As of April 12, 2011, the number of entries in the PDB was 72386, among which 63011 entries were X-ray crystal structures and 6467 entries included zinc in their structures. First, we obtained 5741 X-ray complex structures that contained zinc using the advanced search function of the PDB [13]. The query used here was "Chemical ID = Zn and Experimental method = X-ray." Next, the structures containing ZBGs were identified by measuring the Euclidean distances between the zinc ions and the ligand atoms. The coordinates of the ligand molecules including zinc were exported in both 2D and 3D forms. Then, irrelevant structures such as salts (e.g., ZnCl, ZnBr, etc.) and unknown molecules (e.g., those assigned as UNK or UNL in the hetID field of the PDB file) were removed. In this study, only structures with a relatively high resolution (≤2.5 Å) were used for the analysis because this resolution reveals defined conformations of amino acid side chains [14]. The aforementioned process gave a set of 1215 structures.

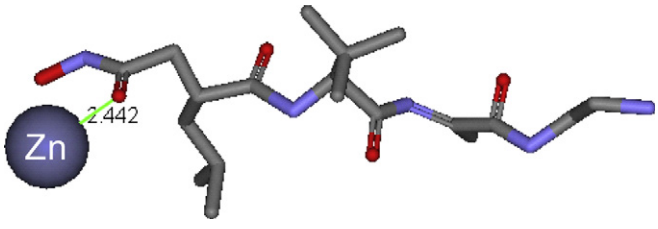

Fig. 2. Identification of ZBG atoms by measuring the distance between the ligand and the zinc ion.

A ZBG database of 2D structures was created by ISIS/Base (Symyx technologies, inc.) because many medicinal chemists are familiar with simple 2D sketched structures. This database includes ligand structures, PDB headers, PDB titles, resolutions, journal information, and the identities of the nearest ligand atoms and their respective distances from the zinc ion. To make the recognition of an interaction between the zinc ion and a ligand intuitive, an explicit single bond was formed between the ligand and the zinc ion. This explicit bond allowed us to search substructures and subsequently analyze the ZBGs. The 3D coordinates of the zinc and ligand atoms were exported for the purpose of visual inspection of the positional relationships between ZBGs and zinc. Finally, the list of ZBGs was created based on the substructure-based analysis of the obtained database. All the references of the 1215 structures are provided as supplementary material (see Table S1).

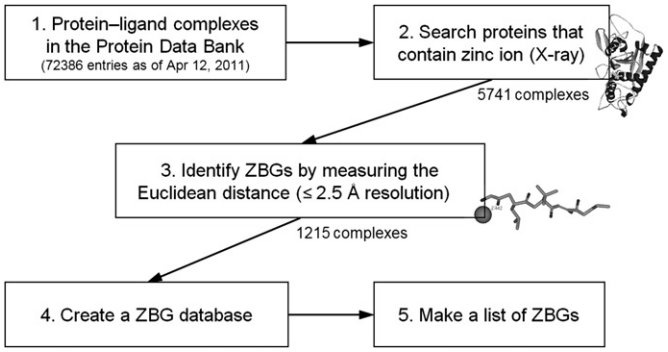

Fig. 3. Workflow of the present study.

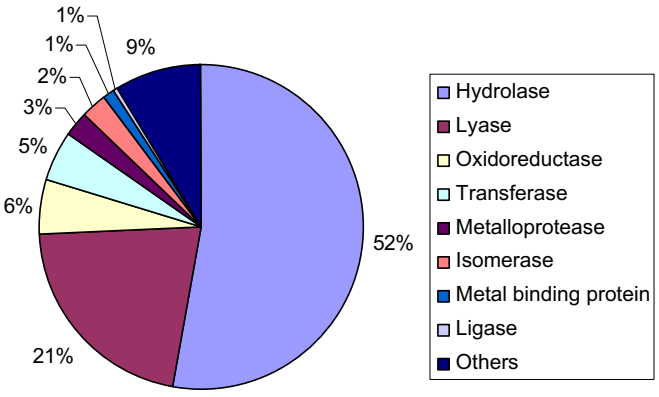

Fig. 4. Family distribution determined from the PDB headers.