An Approach to the Qualitative Prediction of Crystal Structures; Hierarchical Crystal Structures of Inclusion Compounds of Steroids and Alkaloids

Tsuyoshi Watabe, Kazuaki Kato, Norimitsu Tohnai, and Mikiji Miyata*

Department of Material and Life Science, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan
*e-mail: miyata@molec.mls.eng.osaka-u.ac.jp

Abstract: We describe that a hierarchical interpretation serves as a useful approach to the qualitative prediction of crystal structures of organic inclusion compounds. Crystallographic studies on natural compounds such as steroids and alkaloids led us to such an interpretation on the basis of their characteristic chirality, termed as three-directional chirality. It is considered that their hierarchical structures of the crystals consist of the following four steps: chiral molecules as primary structures, helical assemblies of the molecules as secondary structures, bundles of the helices as tertiary structures, and host-guest complexes of the bundles as quaternary structures. Chirality of the corresponding high-order assemblies can be reasonably explained starting from three-directional chirality of the host compounds.

Introduction

A question, “Are crystal structures predictable?”, has been fascinating many chemists for a long time. Gavezzotti gave a quite negative answer to it in 1994 [1], and Dunitz did a polite negative one in 2003 [2]. Many experts in the crystallographic field discussed about it [3]. Some methods seem to be hopeful, but no methods still seem to give consistently reliable predictions.

According to Kitaigorodskii’s studies [4], molecules without symmetry elements form 2_ helical assemblies predominantly, so that close packing is attainable in the space group P2_1/c, P2_1 or P2_12_1_2_. Since highly asymmetric natural compounds, such as steroids and alkaloids, have no symmetry elements, it may be assumed that the compounds form the 2_ helical assemblies related to the corresponding chiral space group. In other words, chirality of the starting compounds may enable us to predict chirality of the helices and space groups. To our knowledge, however, there seems to be no reports that clarified a correlation among their molecular structures, helical assemblies and space groups from a viewpoint of chirality.

We have reported that a hierarchical interpretation is efficient for a qualitative prediction of crystal structures of organic inclusion compounds of steroids such as bile acids and their
derivatives (Fig.1(a)) [5]. It was summarized that diverse hydrogen bonds among the molecules yield their helical molecular assemblies and bundles of the helices. Moreover, such an interpretation was extended to one of alkaloids, brucine (Fig.1(b)), which has no hydrogen bonding donor groups [6]. This result indicates that weaker interactions than usual hydrogen bonds are also efficient for understanding the crystal structures.

This paper deals with a comparative study on the hierarchical crystal structures of bile acids and brucine. It will be described that both of the molecules construct a common hierarchical structure involving the helical assemblies and bundles through non-covalent interaction.

**Experimental**

All chemicals and solvents were commercially available and used without further purification. Bile acids and brucine were recrystallized from various neutral solvents such as alcohols, esters, ketones, aromatic compounds and so on, resulting in the corresponding inclusion crystals.

Crystal structures of the resulting inclusion compounds were determined in the following way. X-ray single-crystal diffraction data were collected on a Rigaku RAXIS RAPID diffractometer with graphite-monochromatized Cu-K$_\alpha$ radiations. Lattice parameters were obtained by reflections for oscillation images for the area detector. Direct methods were employed for the structure solution based on F$^2$. The structure was refined by the full-matrix least squares procedure with the program TEXSAN [7]. All non-hydrogen atoms were refined with anisotropic displacement parameters and hydrogen atoms were placed in idealized positions.

Crystal structural data of the inclusion compounds of bile acids and brucine were quoted from Cambridge Structural Database (CSD) [8].

**Results and Discussion**

**Overview of hierarchical structures of the inclusion crystals** Bile acids are chiral compounds with multiple asymmetric carbons and hydrogen bonding groups, as shown in Fig. 1(a). It is known that they form many inclusion compounds with various organic substances [9]. On the other hand, brucine (Fig. 1(b)) is a chiral compound with many asymmetric carbons and no hydrogen-bonding donor groups. Although its diastereomeric salts are so far well-known [10], we obtained many inclusion crystals of brucine with neutral organic substances [6]. We carried out a comparative study on the crystal structures of the inclusion compounds of bile acids and their derivatives as well as brucine. It was ascertained that most of the crystals have a common framework in spite of their guest-dependent polymorphism. This fact induced the idea that their crystals have a hierarchical structure in common as mentioned below.

Fig. 2 shows that these molecules form characteristic inclusion crystals *via* the following four steps. First, the molecules (Fig. 2(a)) are regarded as primary structures for making the subsequent molecular architecture. Second, such molecules associate together to produce bimolecular assemblies (Fig. 2(b)) followed by helical tape-like assemblies with three distinguishable directions as secondary structures (Fig. 2(c)). Third, the helical tapes
are combined in a parallel or anti-parallel fashion to produce bundles of the helices as tertiary structures (Fig. 2(d)). Fourth, the bundles leave cavities for accommodating guest components to yield host-guest complexes as quaternary structures (Fig. 2(e)). Such complexes are characterized by molecular recognition naturally. It is noteworthy that chirality of the starting molecules dominates chirality of the assemblies, bundles and complexes in the corresponding steps.

**Figure 1.** Structural formulae of bile acids (a) and brucine (b). R₁=R₂=R₃=OH; cholic acid, R₁=R₃=OH, R₂=H; deoxycholic acid, R₁=OH, R₂=R₃=H; lithocholic acid.

**Figure 2.** Overview of a hierarchical structure of inclusion crystals of cholic acid. A chiral molecule (a), a bimolecular assembly (b), a helical tape-like assembly (c), a bundle of the assemblies (d), and host-guest complexes of the bundles (e).

**Primary structures: host molecules with three-directional chirality** Our hands serve as a good tool for recognizing chirality of facial substances. This is because our hands exhibit facial shape with three different directions. Likewise, the host compounds such as steroids and alkaloids have facial shape, since they consist of polycyclic and planar skeletons with many asymmetric carbons. When we rotate steric models of the molecules, we notice that the molecules are alike vertebrate animals with three distinguishable directions. In order to express the directions of animal bodies in our daily life, we use the words; head and
tail (or leg), right and left, and belly and back. Therefore, we propose that such facial asymmetry of the molecules is termed as three-directional chirality, and to use these words for defining the directions [9].

The accompanying problem is how to define three directions of the molecules. We get a suitable idea from their steric models, but not from their structural formulae in Fig. 1. Figs. 3(a) to 3(d) exemplify our stereochemical views for the molecules of bile acids and brucine, where the three directions are designated with the six words mentioned above.

Figs. 3(a) and 3(b) show a front-side and a left-side view of bile acids, respectively. First, the side chain is recognized as a tail, while a part with a hydroxyl group at position 3 as a head. Another direction depends on polarity of the faces of steroidal skeletons. The hydrophilic face with hydrogen bonding groups is assigned as a belly side, while the other lipophilic face with methyl groups as a back side. The final direction is spontaneously determined. That is, a part with a hydroxyl group at position 7 is a left side, while one at position 12 is a right side.

Figs. 3(c) and 3(d) show a front-side and a left-side view of brucine, respectively, which was more difficult to assign the directions than bile acids. This is because the molecule does not have a clear tail and facial polarity. We focus on a convex shape around tertiary nitrogen at position 19. It would be suitable to assign the part as a right side, while a part with an amide group at positions 9 and 10 as a left side. The indol part with two methoxy groups at positions 4 and 5 is assigned as a leg side.

**Figure 3.** Steric views of chiral molecules; (a) a front view and (b) a left-side view of cholic acid, (c) a front view and (d) a side view of brucine. The three-directional chirality of the facial molecules is designated with six words; head and tail (leg), belly and back, right and left.

**Secondary structures: helical assemblies with three-directional chirality** We do not discriminate three directions of 2 helical assemblies composed of the molecules with no symmetry elements. However, the molecules with three-directional chirality, such as bile acids and brucine, make it possible to define three-directional chirality of their helical assemblies in the following way.

First, we discuss about bimolecular assemblies starting from the molecules with three-directional chirality [11]. Fig. 4 shows various modes of bimolecular assemblies when a belly side of the molecule meets a belly side of another molecule. For example, Fig. 4(a) depicts an association of head-to-head, left-to-right, tail-to-tail, and right-to-left type. Fig. 4(b) shows an association of head-to-head, left-to-left, tail-to-tail, and right-to-right type.
4(b) illustrates the case of clockwise rotation of the upper molecule by 90 degrees to give head-to-left, left-to-head, tail-to-right, and right-to-tail type. Further rotation and sliding of the upper molecule yield another types, as shown in Figs. 4(c) to 4(g). A combination of the rotation and sliding may yield an arbitrary position (Fig. 4(h)).

Second, it is necessary to examine chirality of the helices closely, since facial molecules form tape-like helical assemblies. Fig. 5 shows three-directional chirality of such helices. It can be seen that each of the right-handed and left-handed helices has two sides; in and out, up and down. Therefore, we have four types of the tape-like helices, as shown in Figs. 5(a) to 5(d). For example, Fig. 5(a) depicts a right-handed helix with the belly-inside and back-outside, head-upside and tail-downside, while Figure 5(b) does a left-handed helix with the same sides.

On the basis of such consideration, we can explain how to form a helical assembly starting from the molecules with three-directional chirality. Figs. 6(a) to 6(d) illustrate the way for deoxycholic acid. This molecule has two hydroxyl groups at positions 3 and 12 (Fig. 6(a)), so that the molecule faces parallel to a helical axis with its belly-side (Fig. 6(b)). The other molecule attaches the original one according to 2₁ symmetry through intermolecular hydrogen bonds (Fig. 6(c)). Simple simulation suggests that the original molecule leans a little to the right to acquire suitable hydrogen-bonding distances. Further attachment of the molecules leads to a right-handed helical assembly (Fig. 6(d)).

Figs. 6(e) to 6(h) show the way for brucine. Since this molecule has a convex part at the right side around nitrogen at position 19 (Fig. 6(e)), its neighboring concave part can function for accommodating two methoxy groups at positions 2 and 3 of the other molecule. So, the molecule leans quite right to a helical axis with the belly side (Fig. 6(f)). One more molecule attaches the original one according to 2₁ symmetry through weak CH/O and van
der Waals interactions (Fig. 6(g)). Further attachment of the brucine molecules leads to a right-handed helical assembly (Fig. 6(h)).

**Figure 5.** Helices with three-directional chirality; (a) a right-handed helix with belly-inside and back-outside, head-upside and tail(leg)-downside, (b) a left-handed with the same sides as those of (a), (c) a right-handed helix with back-inside and belly-outside, head-upside and tail(leg)-downside, (d) a left-handed helix with the same sides as those of (c).

**Figure 6.** Formation of chiral tape-like assemblies. Molecular formulae (a) and (e), its steric back views (b) and (f), its bimolecular assemblies (c) and (g), and its right-handed helical assemblies (d) and (h) for deoxycholic acid and brucine, respectively.
Tertiary structures; bundles of the helical assemblies. Next, we consider how to assemble the $2_1$ helical assemblies mentioned above. There are two kinds of alignment of the helices in a parallel and antiparallel fashion, as shown in Figs. 7(a) and 7(b), respectively. A typical hexagonal packing gives bundles as tertiary structures, where one helical assembly is surrounded by six other ones. Although it seems to recognize such bundles readily, the crystal structures look like complicated because of flat and directional helices.

Simple alignments of the bundles are introduced below. As shown in Figure 7(c), further parallel alignment of the helices forms a layer assembly as a tertiary structure. A parallel stacking of the layers yield a crystal structure with monoclinic, $P2_1$ space group in the case of cholic acid (Fig. 7(d)), while an antiparallel stacking does one with orthogonal, $P2_12_12_1$ in the case of deoxycholic acid (Fig. 7(e)).

Figure 7. Bundles of the helical assemblies. Parallel gathering (a) and antiparallel gathering (b), a layer composed of parallel gathering (c), a parallel gathering of the layers (d), and an antiparallel gathering of the layers (e).

Quaternary structures: host-guest complexes characterized by molecular recognition. The bundles leave cavities among the helices, where other organic substances are accommodated. Flexible combinations of the helices enable the host components to accommodate the guest components in various ways, resulting in guest-dependent polymorphism of the inclusion compounds. Fig. 8 shows such polymorphism in the inclusion crystals of cholic acid. It can be seen that the polymorphic crystals are based on common helical assemblies but classified by different bundle modes, such as sheet, crossing,
and tape, as shown in Figs. 8(a) to 8(h). It should be noted that each mode accompanies the corresponding hydrogen bonds among the host and guest components.

**Figure 8.** Polymorphic crystals of cholic acid with various bundle structures, termed as sheet I (a) to (d), sheet II (e) and (f), crossing (g), and tape (h).

**Figure 9.** Proposed models for chiral recognition. Conventional three-point model (a) and four-point model in a channel of inclusion compounds (b).

By the way, the cavities can accommodate a limited range of guest components in size, shape, polarity and chirality, which is termed as molecular recognition. We introduce a
novel viewpoint of chiral recognition in inclusion compounds of bile acids. Conventionally, chiral recognition has been explained on the basis of the three-point model, as shown in Fig. 9(a) [12]. However, in the case of channel-type inclusion compounds of bile acids and their derivatives, we encountered several examples that this model is not suitable for explaining the resulting chiral recognition.

For example, Fig. 10 depicts a sectional view of the crystal structure of the inclusion compound of lithocholamide with 2-pentanol [13]. It can be seen that the channel is occupied by three substituents attached to asymmetric carbon; hydroxyl, methyl, and \( n \)-propyl groups, indicating that the condition for the three point model is satisfied. But chiral recognition is not effective according to our enantioresolution study (13 % e.e.). In order to explain this discrepancy it would be necessary to employ the four-point model instead of the three point model (Fig. 9(b)) [14], although this model was recently proposed in enzymic systems [15]. It is considered that the fourth substituent at the asymmetric carbon, that is, hydrogen, may lie in both up-side and down-side toward a plain composed of the other three substituents. When the hydrogen is attracted to either of the sides through any weak interactions, chiral recognition would become more efficient. This attraction may be accomplished by cooperative work through various weak interactions such as CH/O, CH/π, van der Waals and so on. For example, CH/O interaction is effective in the case of the inclusion compound of 3-epicholic acid with 3-methyl-2-pentanol [16].

![Figure 10](image.png)

**Figure 10.** A sectional view of a channel occupied by a guest molecule. This view shows the case of the inclusion compound of lithocholamide with 2-pentanol.

**Conclusion**

It was demonstrated that the hierarchical analysis of molecular assemblies serves as a useful approach to interpret the crystal structures of inclusion compounds of bile acids and brucine. Such molecules form characteristic 21 helical assemblies, which bundle together to give their crystal structures. We introduced a novel concept, three-directional chirality, for the facial molecules, giving us a chance to understand chiral helical assemblies in high-order structures. It is anticipated that this approach would be efficient for the prediction of crystal structures of highly chiral compounds. Finally, it is noteworthy that the hierarchical interpretation of the crystal structures closely relates to the concept of molecular information and expression on the basis of highly chiral molecules such as proteins and steroids [9, 17, 18].
References

