

## Special Feature: Pharmaceutical Analysis (2)

## Drug discovery by single crystal X-ray structure analysis

Akihito Yamano\*

## 1. Introduction

In the early stages of new drug discovery, the structure of small molecule hits resulting from compound screening is modified to derive leads (putative drug molecules) and then the leads are optimized intensely to achieve a molecule with a higher affinity to the target protein. The atomic resolution protein structure indispensable to this optimization step is determined by X-ray crystallography. The structure of the lead itself including its absolute configuration is also important because optical isomers often have completely different effects on living organisms. Single crystal structure analysis is essentially the sole method for determining the absolute structure of a molecule.

The principle of single crystal analysis is quite simple and common in both small molecule and macromolecule crystallography (Fig. 1). When a single crystal is irradiated with X-rays, they bounce off the electrons and interact to produce a set of diffraction spots, which researchers transform Fourier analysis to obtain the electron density inside the crystal. By assuming atoms reside at electron density peak positions, one can reconstruct the molecule. Experimental methods such as multiple isomorphous replacement method (MIR) provide phases necessary to calculate a Fourier transform in the case of macromolecule crystals. When a structure highly homologous to the target molecule is known, the phasing process becomes much simpler because the molecular replacement method can be used. This is often the case with single crystal analysis as used in drug development.

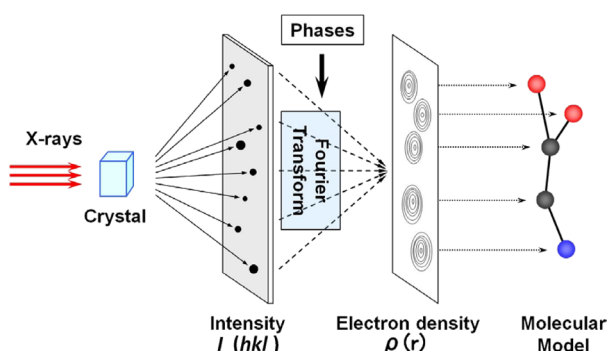


Fig. 1. The principle of single crystal X-ray structure analysis.

## 2. Structure Based Drug Design (SBDD)

There are various approaches used in the initial step of new drug development. The lead discovery based on the precise three-dimensional structure of the target protein is commonly called structure based drug design (SBDD). SBDD utilizes the shape, hydrophobicity, hydrophilicity and charge distributions of the binding pocket to optimize the lead compounds selected most commonly by high-throughput screening (HTS) against pre-existing small molecule libraries. The leads are iteratively examined for their binding modes by computer calculation as well as the co-crystal structure to get directions for improvements of their properties required for a drug (Fig. 2).

This implies that once the structure of a key protein of a disease is determined by X-rays, the door to the new drug development is opened. The latest and one of the most notable examples is the prokaryotic ribosome. The ribosome is a huge cytoplasmic organelle (MW=ca. 2.5MDa) consisting of 3 ribosomal RNA (rRNA) and 55 proteins that form a protein factory. A number of questions on the ribosome were answered when the atomic resolution structure of the ribosome was determined. The action mechanism of an antibiotic, paromomycin (Fig. 3)<sup>(1)</sup> is one of the outstanding examples. Paromomycin is an anti-intestinal infection treatment and binds to the decoding site of 16S rRNA of the 30S subunit. The “S” is the unit describing the sedimentation speed in a centrifuge and an index of

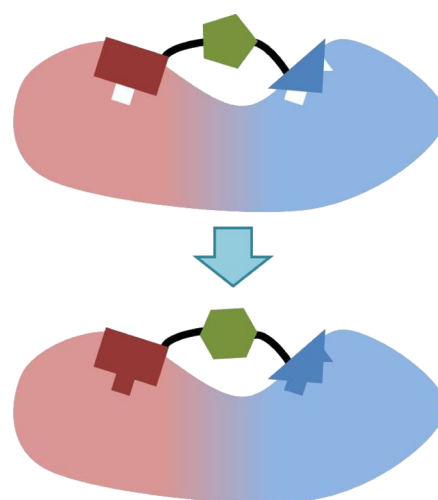
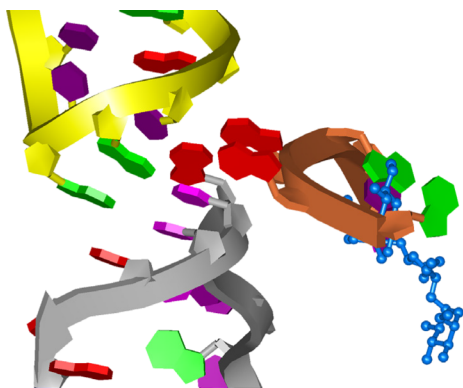


Fig. 2. A schematic diagram of SBDD. SBDD modifies the compound structure to maximize the affinity to the target protein.

\* Application Laboratories, Rigaku Corporation.



**Fig. 3.** The structure in the vicinity of the codon-anticodon pairing site in the 16S rRNA of the 30S subunit (brown), mRNA (gray), tRNA (yellow) and paromomycin (blue) complex. Adenine, guanine, cytosine and uracil are in red, green, pink and purple, respectively. The side chain of A1492 and A1493 residues proofreading the codon-anticodon matching are protruding toward A-C-C and U-G-G due to the binding of the paromomycin molecule.

the size of the molecule. It had been known that the precision of translation from messenger RNA (mRNA) to protein was deteriorated when paromomycin binds to the ribosome. When a correct transfer RNA (tRNA) binds to mRNA, adenines 1492 and 1493 change their conformation to participate in hydrogen bonds between a codon-anticodon recognition to proofread. However, when paromomycin is bound, those residues were fixed to their proofreading position. As a result, a similar but different tRNA can form hydrogen bonds and the precision of translation is deteriorated (Fig. 3). This knowledge led to the drug development that has potency against drug resistant bacteria.

The first step of SBDD is the determination of the three dimensional structure of the target protein. Today, protein structure analysis is getting divided in to two extreme divisions. The structure determination of a novel membrane protein or a huge protein complex is quite difficult even using modern facilities and technologies. Even when the difficulty of crystallization is overcome, there can still be additional obstacles: the limited resolution due to the low crystallinity and the rapid decay of the crystal by X-rays. One must overcome these difficulties one by one by making the full use of crystallographic knowledge and experience. On the contrary, X-ray structure analysis used in SBDD is the other extreme. The co-crystal structure of a known protein and a ligand can be viewed as a structure very similar to that of the protein alone. In this case, the structure analysis becomes very simple. There is no phase problem and there is no need for chain tracing. One can finish the structure analysis by the simple molecular replacement and the successive refinement. Some people believe that the protein data collection is possible only at a beamline. This may be true in such cases when the crystal is smaller than 50  $\mu\text{m}$  and it has unit cell exceeding 500 Å and multiple anomalous



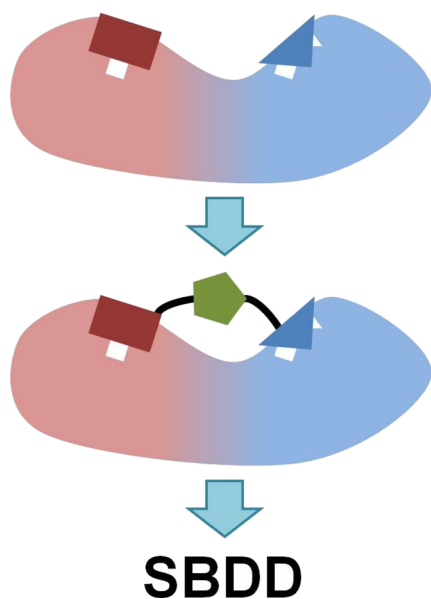
**Fig. 4.** An XtaLAB P200 system consists of a hybrid pixel array detector (PILATUS 200K), a dual wavelength microfocus X-ray generator (MicroMax007HF-DW) and a multilayer optics (VariMax DW).

dispersion (MAD) phasing is required. However, a laboratory system usually suffices in the structure analysis used in SBDD where the molecular replacement method is the major phasing method. Figure 4 shows an example of an in house system that can be used for the structure analysis of biological macromolecules. The combination of the multilayer focusing optics and the micro focus X-ray generator dramatically increases the number of photons delivered to the crystal. The hybrid pixel array detector PILATUS 200K from DECTRIS that enables the true shutterless data collection is employed as the detector. This system can finish the data collection required for SBDD within 20 minutes.

### 3. Fragment Based Drug Development (FBDD)

Until the 1990's, there were some doubts about the effectiveness of SBDD. These days, there are numerous successful examples as represented by HIV protease, therefore there is no room to question the validity of the method itself. However there still is a drawback to SBDD. The molecules integrated in the compound library often have the molecular weight similar to that of the final drug. This means that the space to grow the lead is limited. Only replacements such as a six-membered ring with a five-membered ring, a double bond with a single bond and of functional groups are allowed. This makes it difficult to reach a drug candidate deviating significantly from the starting compound and having a drastic improvement in the activity. It is FBDD (Fragment Based Drug Discovery) that was developed to overcome the weakness of the conventional drug development path of HTS and SBDD. FBDD starts from searching for candidates of components of a drug molecule and either two to three candidates are joined or a candidate is extended to obtain a lead subjected to the structural optimization by SBDD (Fig. 5).

FBDD is expected to be an outstanding method that brings a paradigm shift in the derivation of a lead compound. It is said that FBDD can reach a compound having higher activity and originality. Additionally, the



**Fig. 5.** A schematic diagram of FBDD. In this particular case, two hits were joined to create a starting compound.

number of compounds of a typical HTS library may sum up to a few million compounds, but that of the fragment based screening (FBS) library is only several tens of thousands of compounds at most. FBDD can cover larger chemical space with a much smaller number of compounds compared to HTS. The smaller number of compounds is critical because it means the library is easy to build and inexpensive to maintain.

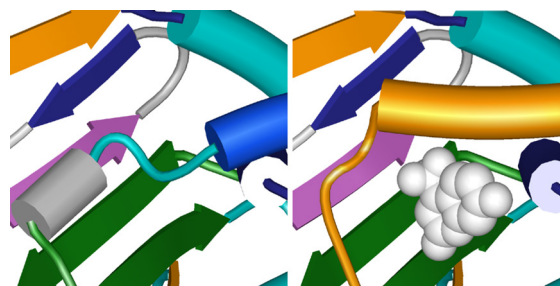
The step of selecting fragment molecules is called fragment based screening (FBS). The first analytical method applied to FBS was NMR but soon X-ray structure analysis was implemented intensely by researchers from Abbott Laboratories<sup>(2)</sup>. They automated the structure analysis thoroughly from data collection to refinement upon introduction of single crystal X-ray analysis<sup>(3)</sup>. Abbot's automated system provided the starting point of the high-throughput system used in pharmaceutical companies around the world (Fig. 6)<sup>(4),(5)</sup>. 60 data sets can be measured without human intervention<sup>(6)</sup>.

The protein structure analysis in FBS is even easier than that of SBDD. In SBDD, compounds may need to be added in the crystallization process and this sometimes necessitates exploring crystallization conditions. In FBS, compounds are usually introduced by soaking crystals prepared beforehand because the chemicals are relatively small. The soaking eliminates the possibility of changing crystallization conditions and assures the isomorphism to the native crystals. The successful data collection is almost equivalent to the completion of the structure analysis and little crystallographic knowledge is required.

The most outstanding feature of FBS by X-rays is the capability of visualizing how the fragment is bound to the target protein. The position and orientation of



**Fig. 6.** A high-throughput single crystal X-ray structure analysis system. A sample changer, automated crystal centering and evaluation are integrated. This system gives us options such as data collection in the descending order of the score and elimination of crystals below a threshold.



**Fig. 7.** An example of the structural change induced upon the binding of a compound. The loop observed at the center of the left picture becomes a helix in the right picture.

the fragment and atoms involved in the interaction can directly be identified. Additionally, remaining spaces and directions are identified. These give us tangible guidelines to improve the compound. When hits can be grouped by their locations, possible connections among fragments can also be predicted. Since the binding of fragments is directly judged by looking at the structure, non-specific binding is automatically eliminated.

There are cases where the relatively large structural change occurs upon binding of a compound (Fig. 7). At the initial stage of the compound screening, a cocktail of compounds is used to raise the efficiency. Sometimes multiple fragments bind to the protein simultaneously. FBS by X-rays can deal with all these cases as long as the crystallinity of the co-crystal is maintained.

FBS by X-rays can deal with the widest range of ligand concentration compared to other analytical methods. In X-ray structure analysis, there is a technique called a difference Fourier calculation. It produces an electron density map by using the differences between the calculated and observed structure amplitudes ( $F_o - F_c$ ) as the coefficients. The difference Fourier map is used to elucidate the deficiency of the current molecular model. When the data quality is sufficient,

**Table 1.** The summary of the FBS experiment targeting HSP90.

Items	Results
Total # of compounds	384 (4 compounds×96)
# of consumed crystals	146
# of successful crystals	96
Success rate	65.8%
Total measurement time	~ 100 hours
Structure analysis time	~ 20 hours
FBS total time	Total 120 hours (~ 5 days)
# of Hits	10
Hit rate	2.6%

compounds low in occupancy may possibly be detected.

Table 1 summarizes the results of FBS performed on heat shock protein 90 (HSP90). HSP90 is a 90 kDa molecular chaperon contributing to the stabilization of various kinases and transcription factors involved in the signal transduction, and duplication and differentiation of cells<sup>(6),(7)</sup>. HSP90 is a target of cancer drugs because the growth of tumor is restrained when the HSP90 is inhibited. The development of HSP90 inhibitor has been carried out intensely<sup>(8),(9)</sup>. FBS on 384 compounds were carried out against HSP90 and resulted in 10 hits. Figure 7 shows examples of these hits.

#### 4. X-ray structure analysis in generic drugs

The most notable feature of generic drugs is the lower price than that of the brand-name drugs. This is possible because generic drug manufacturers can eliminate the development cost. Some generic drug manufacturers try to raise the yield of chemical synthesis for additional cost reduction. However, when the synthetic path is changed, the three dimensional structure including the absolute configuration of the final compound must be checked. APIs (Active Pharmaceutical Ingredients) often have chiral carbons and optical isomers must be distinguished because they have strikingly different effects on the human body. Single crystal X-ray structure analysis is practically the sole method to determine the absolute configuration of the optically active compound. With the latest equipment, a 10 μm

cubic crystal is sometimes enough to obtain diffraction data for the structure determination along with the absolute structure.

In recent years, the development of biomedicine is becoming popular. The real substance of the biomedicine is a protein itself. A familiar example of biomedicines is the antibody drug specific for the antirheumatism treatment called tocilizumab (Actemra) from Chugai Pharmaceutical Co., Ltd. The generic form of an antibody drug is called a biosimilar (similar drug) because in principle it is believed to be impossible to make an identical copy of a protein. The antibody drug itself is easily produced once the amino acid sequence is determined, however the three dimensional structure must be confirmed by single crystal structure analysis. The structure analysis in this case is as simple as those in SBDD and FBDD because the structure is known, therefore the molecular replacement method can be used. When the structure cannot be determined by the molecular replacement method, it indicates that the protein may not have the correct folding. It is expected that when the development of biosimilar drugs increases, generic drug manufacturers will need the capability to perform protein structure analysis.

#### References

- (1) V. Ramakrishnan and P. B. Moore: *Curr. Opin. Struct. Biol.*, **10** (2001), 144–154.
- (2) V. L. Nienaber, P. L. Richardson, V. Klighofer, J. J. Bouska, V. L. Giranda and J. Greer: *Nature. Biotechnol.*, **18** (2000), 1105–1108.
- (3) S. W. Muchmore, J. Olson, R. Jones, J. Pan, M. Blum, J. Greer, S. M. Merrick, P. Magdalinos and V. L. Nienaber: *Struct. Old. Des.*, **8** (2000), R243–R246.
- (4) J. W. Pflugrath, R. Athay, D. P. Edwards, T. J. Niemeyer, T. L. Hendrixson, A. R. Criswell, C. Yang, G. K. Crane, J. D. Ferrara, T. Nienaber, W. Robertson and R. Schafer: *Acta Cryst.*, **A58**, c72 (2002).
- (5) C. Abad-Zapatero: *Acta. Cryst.*, **D61** (2005), 1432–1435.
- (6) A. J. Sharff: *The Rigaku Journal (English version)*, **19/20** (2003), No. 2/1, 5–10.
- (7) W. Xu and L. Neckers: *Clin. Cancer Res.*, **13** (2007), 1625–1629.
- (8) M. V. Powers and P. Workman: *Endocrine-Related Cancer*, **13** (2006), S125–S135.
- (9) M. Congreve, G. Chessari, D. Tisi and A. J. Woodhead: *J. Med. Chem.*, **51** (2008), 3661–3680.