Introduction to single crystal X-ray analysis
V. Some Key Points of structure analysis by Rigaku's CrystalStructure

Akihito Yamano* and Mikio Yamasaki*

1. Introduction

The previous article described data collection. When data collection is done, the next step is to determine the structure. This article discusses critical points in structure analysis.

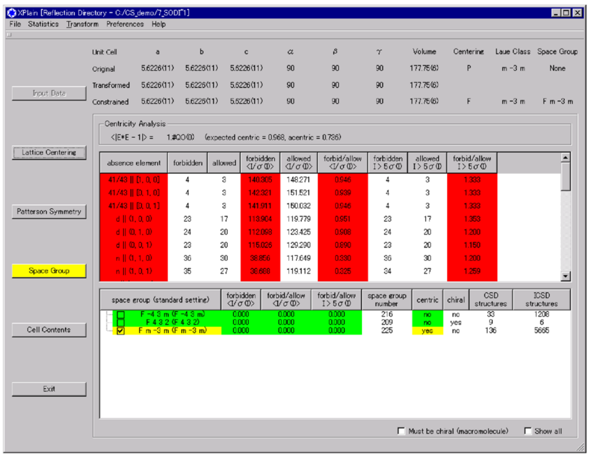
2. Determination of space group

The first step of structure analysis is to select a space group and to create a reflection file by merging and averaging equivalent reflections. CrystalStructure has two options to perform this procedure: PROCESS and XPlain. PROCESS is fully automated and derives a correct space group in most cases. However, when there is an ambiguity in space group selection due to the nature of the crystal, it sometimes selects an incorrect space group. XPlain, integrated in version 4.0 or above, may resolve this problem. XPlain has maximum flexibility allowing us to perform the space group determination process step by step: lattice type, Patterson symmetry and space group. In the final space group selection step, mean variance $\langle|E^2-1|\rangle$ (where E is the normalized structure factor) is calculated primarily to check the centricity (Fig. 1). Additionally, this mean variance gives us information about possible twinning. When $\langle|E^2-1|\rangle$ is significantly lower than the theoretical values, 0.968 and 0.736 for non-centrosymmetric and centrosymmetric crystal respectively, one should consider the possibility of merohedral twinning until the structure analysis finishes.

PROCESS automatically transforms the unit cell according to the resulting space group if necessary. However, when one needs to repeat executing PROCESS, one should be aware that it uses the transformed unit cell. Contrary, XPlain always starts from the original unit cell, therefore it is not necessary to transform the unit cell to the original one.

3. Direct methods

In the phase determination step, direct method programs are generally used for small molecule crystals. CrystalStructure is interfaced to virtually all of the existing direct method programs. Some of the programs are included as default but others needs to be obtained by the user and placed in a certain directory under 'extras'.



Unit Cell	a	b	c	α	β	γ	Volume	Centering	Laue Class	Space Group
Original	5.6226(11)	5.6226(11)	5.6226(11)	90	90	90	177.75(6)	P	m -3 m	None
Transformed	5.6226(11)	5.6226(11)	5.6226(11)	90	90	90	177.75(6)			
Constrained	5.6226(11)	5.6226(11)	5.6226(11)	90	90	90	177.75(6)	F	m -3 m	F m -3 m

Centricity Analysis

<|E*E - 1|> = 1.#QO(0) (expected centric = 0.968, acentric = 0.736)

absence element	forbidden	allowed	forbidden <I/σ(I)>	allowed <I/σ(I)>	forbid/allow <I/σ(I)>	forbidden I > 5σ(I)	allowed I > 5σ(I)	forbid/allow I > 5σ(I)
41/43 || [1, 0, 0]	4	3	140.305	148.271	0.946	4	3	1.333
41/43 || [0, 1, 0]	4	3	142.321	151.521	0.939	4	3	1.333
41/43 || [0, 0, 1]	4	3	141.911	150.032	0.946	4	3	1.333
d || (1, 0, 0)	23	17	113.904	119.779	0.951	23	17	1.353
d || (0, 1, 0)	24	20	112.098	123.425	0.908	24	20	1.200
d || (0, 0, 1)	23	20	115.026	129.290	0.890	23	20	1.150
n || (1, 1, 0)	36	30	38.856	117.649	0.330	36	30	1.200
n || (1, 0, 1)	35	27	38.688	119.112	0.325	34	27	1.259

space group (standard setting)	forbidden <I/σ(I)>	forbid/allow <I/σ(I)>	forbid/allow I > 5σ(I)	space group number	centric	chiral	CSD structures	ICSD structures
☐ F -4 3 m (F -4 3 m)	0.000	0.000	0.000	216	no	no	33	1208
☐ F 4 3 2 (F 4 3 2)	0.000	0.000	0.000	209	no	yes	9	6
☑ F m -3 m (F m -3 m)	0.000	0.000	0.000	225	yes	no	136	5665

☐ Must be chiral (macromolecule) ☐ Show all

Fig. 1. XPlain menu.

Generally, newer programs have better features dealing with more difficult structures. Each program has its own special features; therefore having a variety of programs improves the possibility to obtain the initial structure. It is a common practice that one program can readily derive initial phases when other programs fail.

In the model building step, the key is how well one can assign atoms to the initial E-map peaks. When the actual structure is largely different from the expected structure, it is often difficult to trace the connectivity because the initial phases greatly depends on the accuracy of the E values and the chemical formula is used in the normalization process which produces the E values. In such a case, it may be a good idea to assign heavy atoms to higher peaks tentatively. If it is an organic molecule, you should identify and assign characteristic modules such as 5 or 6-membered rings. The difference Fourier peaks calculated by using improved phases at the end of a refinement may represent peaks that can be used to complete the structure.

CrystalStructure 4.0 has an interface for Superflip. Superflip is also a direct method program but the algorithm to improve the phases is largely different from that of conventional direct method programs. It uses 'charge flipping' (density modification) to improve phases. One of the outstanding features of this software is that it can be executed without chemical composition. Superflip should be the choice when the chemical formula is uncertain.

---

* Application Laboratories, Rigaku Corporation.

4. Refinement

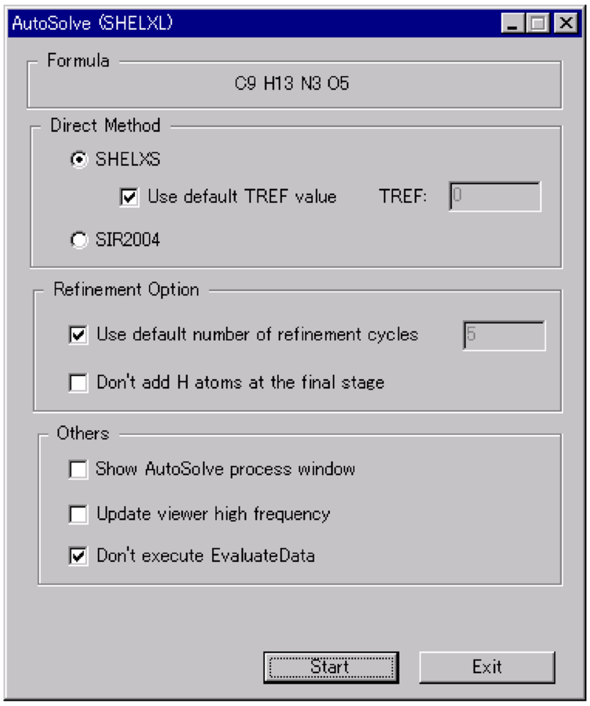
CrystalStructure has interfaces for two structure refinement programs: Crystals and SHELXL(1). CrystalStructure 4.0 gives the priority to SHELXL when it is installed on the computer because SHELXL is the *de facto* standard of refinement program.

4.1. AutoSolve

AutoSolve (Fig. 2) is enhanced in CrystalStructure 4.0. When a correct formula is given, AutoSolve can reach the final structure most of the time in the case of organic compounds. The success rate is approximately 60% with metal complexes(2). To take the advantage of AutoSolve, SHELXL must be installed because it employs SHELXL to refine the structure.

4.2. Automated hydrogen assignment

Earlier versions of CrystalStructure had a function to semi-automatically add hydrogen atoms when 'Locate in Fourier automatically' was selected. This function works only when difference Fourier peaks corresponding to hydrogen atoms are present at geometrically consistent positions. There also is another function to add hydrogen atoms manually. First one specifies the type of geometry and then the routine generates hydrogen atoms in geometrically correct positions. When the structure is large, this procedure becomes tedious because one must iterate this process for the number of parent atoms. The algorithm developed for AutoSolve is now available as a standalone module to add hydrogen atoms automatically. This function, 'Generate all hydrogen atoms geometrically', calculates all hydrogen atoms when the button is pressed (Fig. 3).



Formula: C9 H13 N3 O5

Direct Method
● SHELXS
☑ Use default TREF value TREF: 0
○ SIR2004

Refinement Option
☑ Use default number of refinement cycles 5
☐ Don't add H atoms at the final stage

Others
☐ Show AutoSolve process window
☐ Update viewer high frequency
☑ Don't execute EvaluateData

Fig. 2. AutoSolve menu.

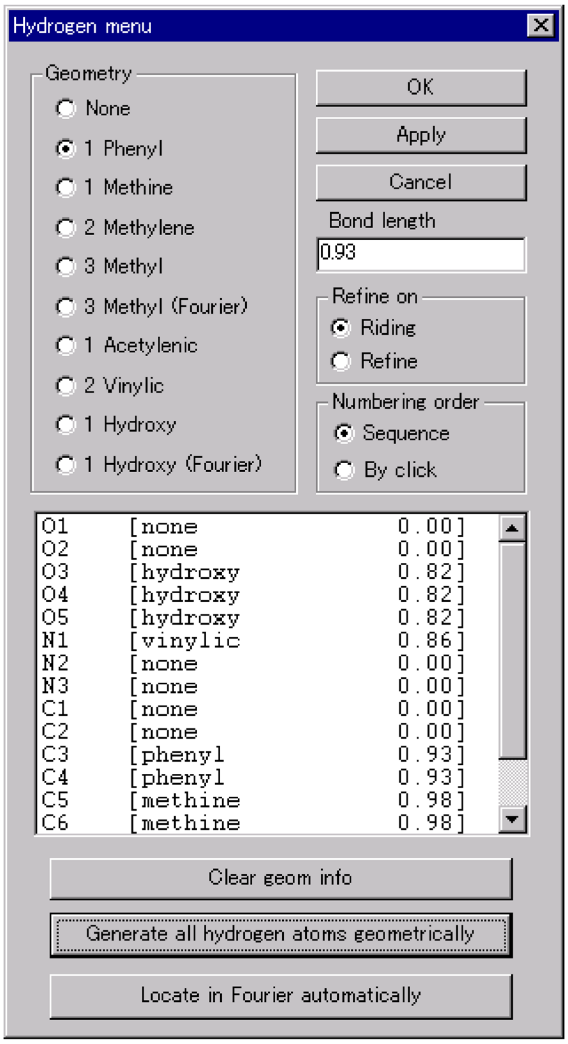
4.3. Check for Acta for confirmation of refinement

To decide when to stop the refinement, one can use the CIF check service provided by IUCr. However, it is not a good idea to abuse the IUCr service by submitting every time after a refinement. It is also time consuming because a CIF must be created and submitted to have it checked by the service.

CrystalStructure checks 8 common and most critical items at the end of a refinement. The result is shown in a popup dialogue 'Check for Acta'. It adopts the criteria set by the CIF check service CheckCIF/PLATON (Fig. 4). The first column indicates the alert level. An Alert A is displayed with a red font because it is a fatal error. Alert A must be addressed and proper action needs to be taken. One must either eliminate the cause of the problem or provide a rational excuse when the alert is unavoidable.

(a) Mu×R

This is the criterion for determining the appropriate type of absorption correction to be applied. There are basically two types of absorption corrections: empirical and numerical correction. Empirical absorption correction utilizes symmetry equivalent reflections to calculate a virtual absorption correction surface. Numerical absorption corrections are calculated from the actual X-ray



Geometry
○ None
● 1 Phenyl
○ 1 Methine
○ 2 Methylene
○ 3 Methyl
○ 3 Methyl (Fourier)
○ 1 Acetylenic
○ 2 Vinylic
○ 1 Hydroxy
○ 1 Hydroxy (Fourier)

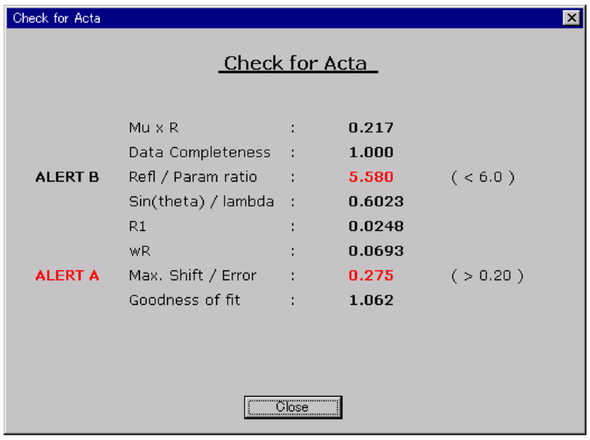
Bond length: 0.93

Refine on
● Riding
○ Refine

Numbering order
● Sequence
○ By click

```
O1    [none        0.00]
O2    [none        0.00]
O3    [hydroxy     0.82]
O4    [hydroxy     0.82]
O5    [hydroxy     0.82]
N1    [vinylic     0.86]
N2    [none        0.00]
N3    [none        0.00]
C1    [none        0.00]
C2    [none        0.00]
C3    [phenyl      0.93]
C4    [phenyl      0.93]
C5    [methine     0.98]
C6    [methine     0.98]
```

Fig. 3. Hydrogen menu.



Check for Acta

	Mu x R	:	0.217	
	Data Completeness	:	1.000	
ALERT B	Refl / Param ratio	:	5.580	(< 6.0)
	Sin(theta) / lambda	:	0.6023	
	R1	:	0.0248	
	wR	:	0.0693	
ALERT A	Max. Shift / Error	:	0.275	(> 0.20)
	Goodness of fit	:	1.062	

Fig. 4. Check for Acta menu.

path and the reduction in intensity due to the absorption by the crystal. The type of appropriate absorption correction is defined by the linear absorption coefficient μ (Mu) and the median of height, width and depth of the crystal.

- $\mathrm{Mu}\times\mathrm{R}<0.1$: no absorption correction
- $0.1<\mathrm{Mu}\times\mathrm{R}<1$: empirical absorption correction
- $1<\mathrm{Mu}\times\mathrm{R}<3$: numerical or empirical absorption correction
- $\mathrm{Mu}\times\mathrm{R}>3$: numerical absorption correction

Note that to apply a numerical absorption correction, the outer shape of the crystal must be obtained during the data collection. To collect data on an inorganic crystal having a large linear absorption coefficient, one should either use a regular sized crystal and apply the numerical absorption correction, or use a crystal small enough to give $\mathrm{Mu}\times\mathrm{R}$ smaller than 1. This item has been considered as obsolete by some lately and no longer checked directly. However, it is a good practice to follow the criteria to minimize the systematic errors caused by strong absorption.

(b) Data Completeness

Data completeness tells you the ratio of number of collected reflections against potentially collectable reflections. It is expected to be very close to 100% and must be 98% at lowest. Data collection software should incorporate a routine capable to generate a data collection strategy to satisfy this criterion.

(c) Refl/Parm ratio

For a crystal having an inversion center, the reflection to parameter ratio must be larger than 10. For a crystal without an inversion center, this threshold is reduced to 8. A small reflection to parameter ratio may make least squares refinement unstable.

(d) Sin(theta)/lamda

Sin(theta)/lamda is to check the diffraction resolution. This value is required to be larger than 0.6. The 2θ angle should be larger than 50 and 134 degree for Mo$K\alpha$ and Cu$K\alpha$, respectively.

(e) R1

R1 is the most common index showing the quality of single crystal analysis. It indicates the agreement between the measured and calculated structure amplitudes. The R1 value is calculated using reflections $|Fo|>4\sigma(|Fo|)$ and needs to be below 10%.

(f) wR

wR is a weighted R factor and needs to be smaller than 25%.

(g) Max. Shift/Error

Max. Shift/Error indicates the conversion of refinement. It must be within the range of ±0.02.

(h) Goodness of fit (GOF)

GOF is to judge whether the weighting scheme employed in the refinement is appropriate. GOF is ideally 1, therefore it should be in the range of 0.8 to 2. When the GOF is far from 1, the weighting scheme should be adjusted until the GOF becomes close to 1 even if the new weight gives higher R1 value.

(i) Residual electron density

The largest peak should be within the range of the atomic number of the heaviest atom x±0.075. When there are peaks above the threshold, unassigned solvent molecule or a part of the structure may not have been modeled properly.

5. Report creation and evaluation of the result

When the refinement is finished, one can produce a report and a CIF file.

Bond lengths and angles should be checked to make sure they are within the acceptable values. If there is an unusual value, the structure must carefully be checked.

6. Summary

Points in single crystal structure analysis are summarized here.

Quality data is the most important point in structure analysis. R_{merge} is to assess the data quality and it must be lower than 10%. Accurate high angle reflections are also important to obtain good initial phases by direct methods. The chemical formula is also critical because it is used to calculate normalized structure factors that gives the basis to statistical treatment in direct methods.

Ambiguity in space group is one of the most common problems that undermine direct methods and successful refinement. When there is something unusual in a structure, such as inconsistent temperature factors, one should suspect that the space group has been assigned incorrectly. An insufficient number of observed data, twinning and disorder are also frequent problems.

If the structure is used internally for checking the atomic connectivity, it is not necessary to resolve all alerts. However, to publish the result, one needs to address all alerts. Even if there remain unavoidable problems, one can use 'Validation reply formula' or 'vrf' to give a rational explanation. CheckCIF/PLATON excludes trivial mistakes, but it is the researcher who is ultimately responsible for the structure. Patience is the key to overcome alerts.

References

(1) G. M. Sheldrick: *Acta Cryst.*, **A64** (2008), 112–122.
(2) H. Kanda, T. Higashi and M. Maeyama: *Acta Cryst.*, **A67** (2011), C599.