# *SHELX*による精密化の紹介

量子科学技術研究開発機構（QST）
量子ビーム科学研究部門
東海量子ビーム応用研究センター

平野　優

# Outline

1. *SHELX*の概要

2. *SHELXL*による精密化

3. 実例（NADHシトクロム$b_5$還元酵素を例にして）

**参考文献等**

SHELX HP： http://shelx.uni-goettingen.de/

結晶構造精密化SHELXLの使い方、大場茂・植草秀裕著、三共出版

SHELXLの特徴を知ろう！、大場茂、日本結晶学会誌57, 87 (2015)

# 1. *SHELX*の概要

*SHELX*は、X線回折および中性子回折による
低分子と高分子の結晶構造解析のためのプログラム群

アカデミックフリー（Linux, Win, Mac OSXに対応）

SHELX-2017に含まれるプログラム

SHELXT：低分子の構造決定

SHELXS：低分子の構造決定（直接法）

**SHELXL：低分子と高分子の精密化**

PDB2INS：SHELXLのインプットファイル作成

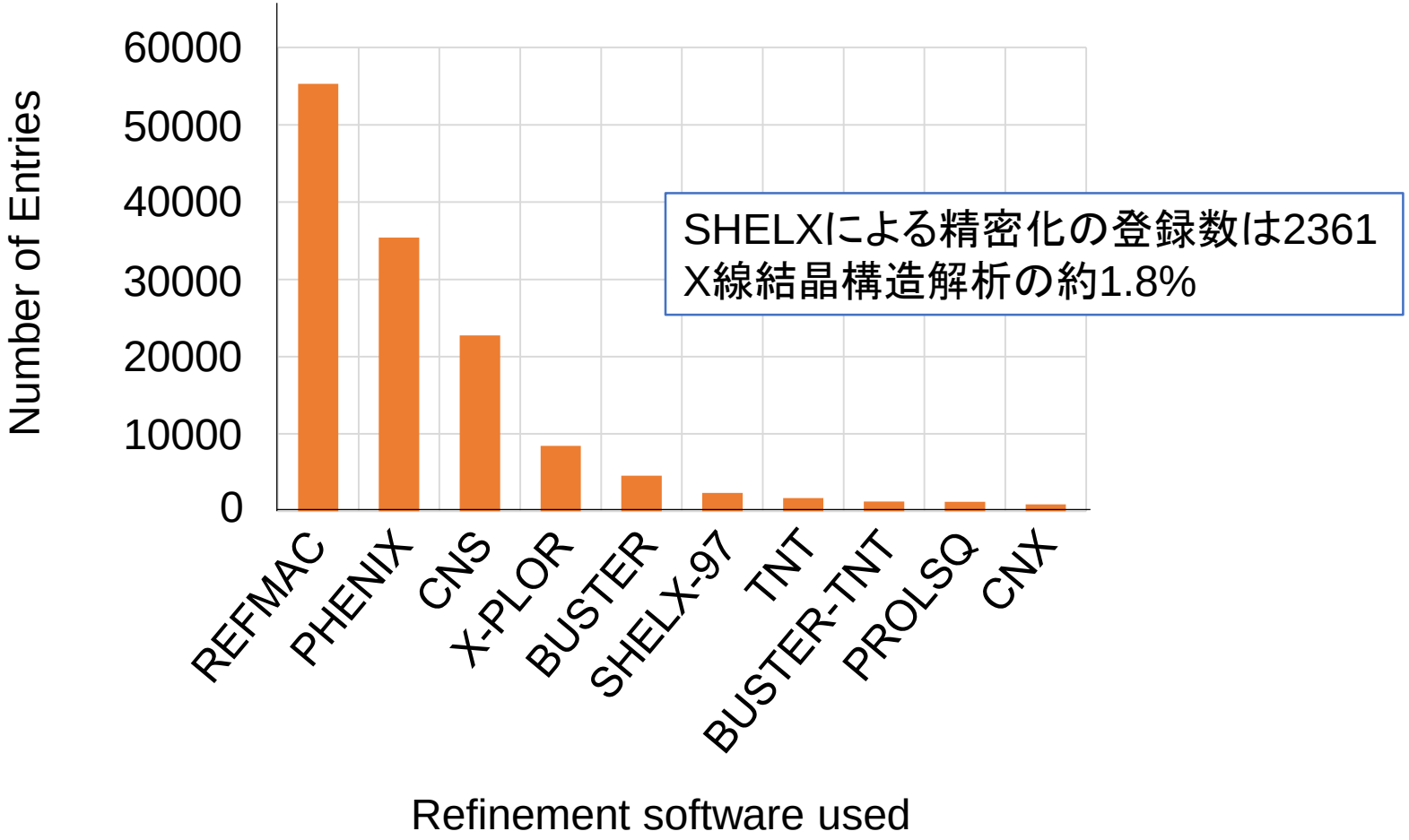
CIFTAB, ShredCIF：SHELXLのCIFファイル編集

SHELXC, SHELXD, SHELXE：高分子の位相決定 → この後の実習

AnoDe：異常散乱電子密度の作成等

SHELX HPより： http://shelx.uni-goettingen.de/

# タンパク質の精密化に用いられるプログラム




PDB HPより：https://www.rcsb.org/stats/distribution_software

# *SHELXL*による精密化の特徴

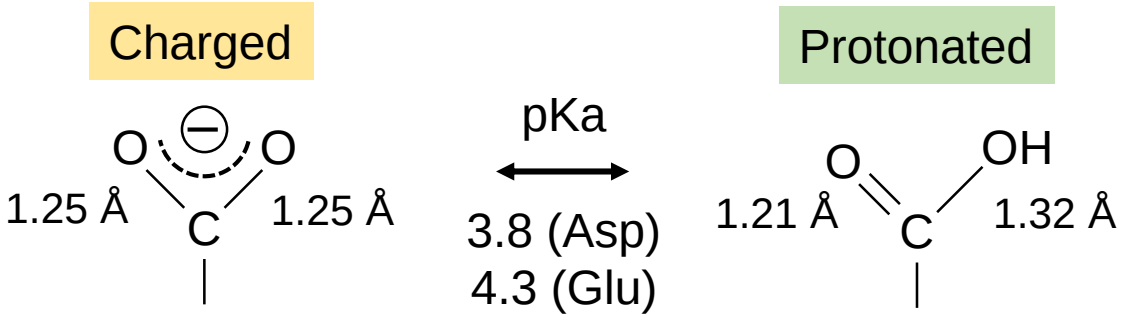
高分解能データの精密化に適している

（2 Åより高分解能で、できるだけ高分解能が望ましい）

ディスオーダー、非等方性ADP（atomic displacement parameter）、水素原子をモデルに含めた精密化が可能

結合距離、結合角度のesd値（estimated standard deviation）が得られる

利用例：Asp, Glu側鎖のプロトン化状態




C – O距離　　同じ　　異なる（~0.1 Å）

# 完全行列最小二乗法による精密化

モデル構造の全ての変数　$p_i = 10 \times \mathrm{N}$、$\mathrm{N}$ = 原子数

（座標×3、異方性温度因子×6、占有率×1）

$$R = \sum_{hkl} w(hkl)\{|F_o(hkl)|^2 - |F_c(hkl)|^2\}^2$$ を最小にする条件 $$\frac{\partial R}{\partial p_i} = 0$$

Taylor展開によって近似して連立一次方程式（normal equation）を求める

$$\sum_i \varepsilon_i \underbrace{\left[\sum_{hkl} w(hkl)\left(\frac{\partial|F_c(hkl)|^2}{\partial p_i}\right)\left(\frac{\partial|F_c(hkl)|^2}{\partial p_j}\right)\right]}_{a_{ij}} = \underbrace{\sum_{hkl} w(hkl)\{|F_o(hkl)|^2 - |F_c(hkl)|^2\}\left(\frac{\partial|F_c|^2}{\partial p_j}\right)}_{b_j}$$

$$\Rightarrow \quad \sum_i \varepsilon_i a_{ij} = b_j$$

（$\varepsilon_i$ : $p_i$のシフト量）　全ての行列要素について計算

$p_j$の標準偏差　$$\sigma(p_j) = \sqrt{\frac{a_{ij}^{-1} R}{N_o - N_p}}$$

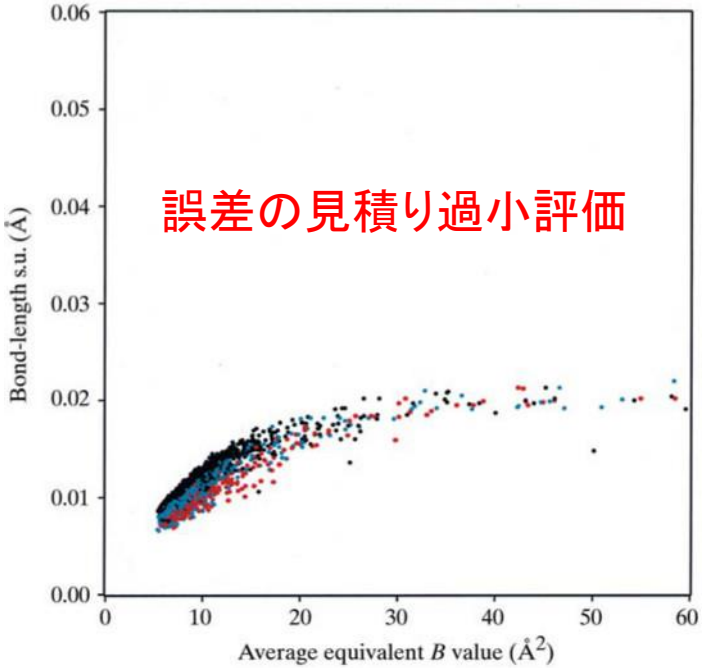
（$N_o, N_p$ : データ数，パラメータ数）

（A–B間距離 $l$ については　$\sigma(l) = \sqrt{\sigma_A{}^2 + \sigma_B{}^2}$ ）

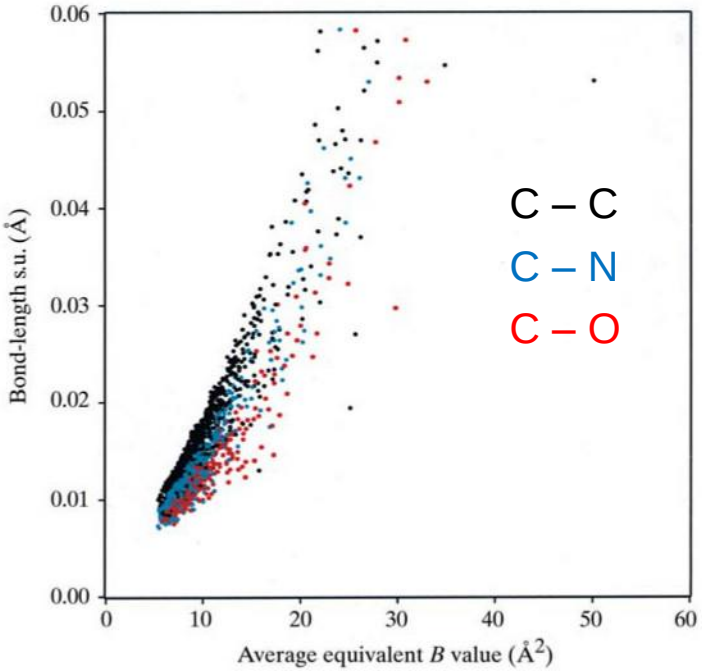
# 結合距離のesdに対する制約の影響

Concanavalin A（0.94 Å分解能）における完全行列最小二乗法による精密化

Restrained




Unrestrained




Restrainedあり精密化の際の結合距離のσ

$$1/\sigma_{res}^2(l) = \underbrace{1/\sigma_{diff}^2(l)}_{\text{制約なし}} + \underbrace{1/\sigma_{geom}^2(l)}_{\text{結合距離の制約}}$$

International Tables for Crystallography (2012). Vol. F, p. 503, Figure 18.5.4.2を改変

# 2. *SHELXL*による精密化

*Refmac*, *PHENIX*などのプログラムによる精密化
（~1.5 Å分解能、等方性温度因子）

↓

*SHELXL*へ移行（Rigid body refinement）

↓

分解能の拡張（共役勾配法による精密化）
（マルチコンフォメーションの追加、非等方性温度因子、制約の除外）

↓

水素原子をモデルに付加して精密化

↓

完全行列最小二乗法によるesdの計算

# 精密化において必要なファイル

## 1. インプットファイル（ins）

```
TITL Rubredoxin in P1 (from 6RXN in PDB)
CELL 1.54178 24.920 17.790 19.720 101.00 83.40 104.50
ZERR       1  0.025  0.018  0.020   0.05  0.05   0.05
LATT -1
SFAC  C   H   N  O   S  FE
UNIT  224 498 55 136 6  1

DEFS 0.02 0.2 0.01 0.05
                        ⋮
RESI   1  MET
C1      1  -0.01633   0.35547   0.44703  11.00000   0.11817
O1      4   0.01012   0.32681   0.48491  11.00000   0.17896
N       3   0.00712   0.35446   0.37983  11.00000   0.11863
                        ⋮

HKLF  4
END
```

コマンド

原子座標、占有率、温度因子

## 2. 反射データファイル（hkl）

```
  -1   0   0  273.70    0.87   1
   1   0   0  272.48    0.87   1
  -2   0   0  585.06    1.80   1
   2   0   0  587.29    1.80   1
  -3   0   0  471.76    1.48   1
   3   0   0  475.22    1.50  -1
                ⋮
```

$(h\ k\ l\ I_o\ \sigma(I_o)\ \mathrm{flag})$

# 精密化に必要なファイル作成方法

## 1. insファイル

PDB2INSを利用　　pdb2ins ↵

```
⋮
Create .hkl file from structure factor file (cif) or PDB code? (y or n) [N]: N ↵
Enter name of PDB file to read (To download a PDB file enter '@<PDBCODE>'): start-model.pdb ↵
Enter HKLF code (3 for F, 4 for F-squared) [4]:4 ↵
Please enter Z (number of molecules per cell):4 ↵
No wavelength found in file. Enter wavelength in Angstroms [1.54178]: 0.7 ↵
Reset water occupancy to unity? (y or n) [Y]: Y ↵
Enter .ins filename to be created [start-model.ins]: ↵
⋮
```

ins作成時"N"
初期構造
4（default）選択
Z値
測定波長
水分子の占有率
insファイル名

## 2. hklファイル

mtz2hklを利用　　http://shelx.uni-ac.gwdg.de/~tg/research/programs/conv/
よりダウンロード可能

mtz2hkl␣data1.mtz ↵　　hklファイル出力（HKLF 4フォーマット）

CCP4のmtz2variousを利用

Format: SHELX,　Label: $I$, $\sigma(I)$, FreeR-flag

# 精密化における入出力の流れ

1. 入力

insファイル：精密化のコマンド、原子座標データ

hklファイル：反射データ

2. 実行

insファイル、hklファイルの拡張子前部分を揃える

（例： refine-1.ins, refine-1.hkl）

shelxl␣refine-1 ↵ （並列計算利用の場合 –t オプション、”-t4”など）

3. 出力

resファイル：insファイルと同じ形式、精密化後の原子座標データ

lstファイル：精密化のログ、結合距離・角度（esd）なども書き出される

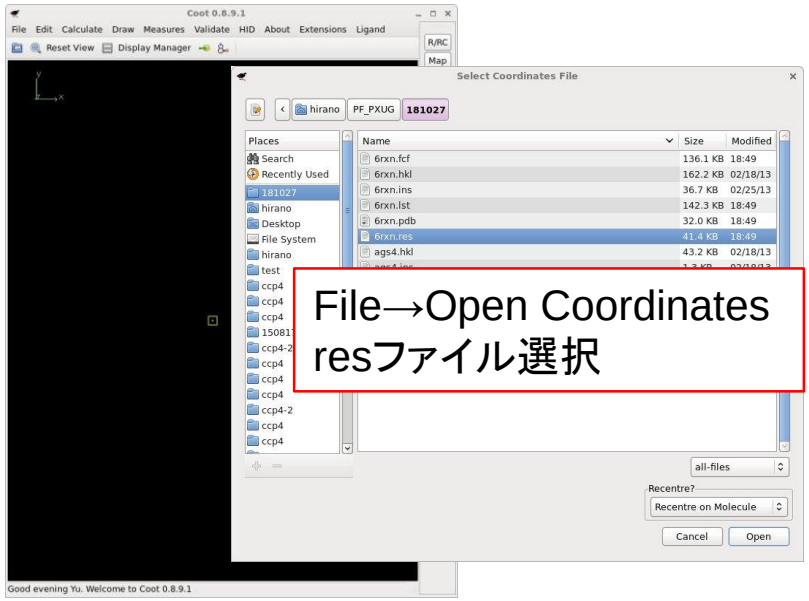
pdbファイル：精密化後のpdb

fcfファイル：精密化後の反射リスト（Cootで電子密度図の描写に利用）

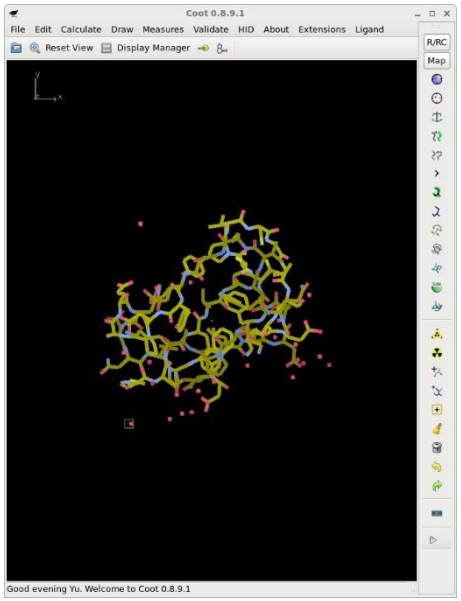
（例： refine-1.res, refine-1.lst, refine-1.pdb, refine-1.fcf）

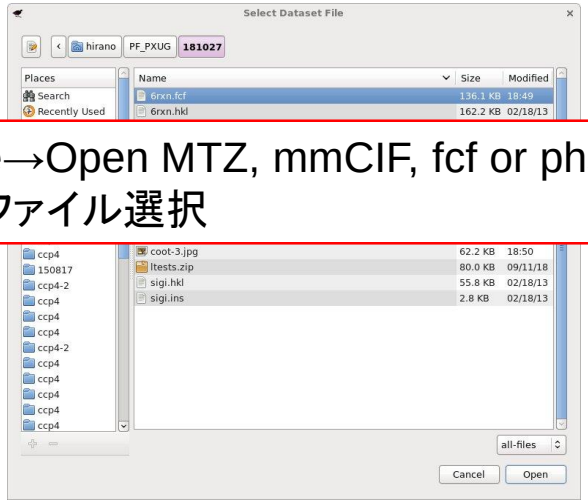
Cootで読み込み

# Cootにおけるファイルの読み込み

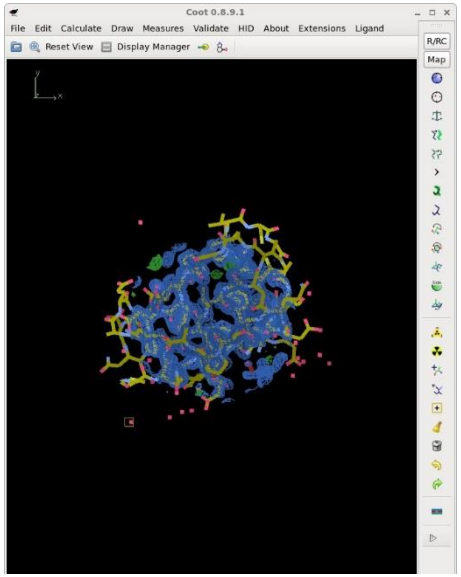



File→Open Coordinates
resファイル選択




File→Open MTZ, mmCIF, fcf or phs…
fcfファイル選択




モデル修正後、File→Save Coordinates…、で新しいinsファイルが保存される

# insファイルの内容　1

コマンドの種類約80
コマンド　＋　引数　という形式になっている

```
TITL converted from file start-model.pdb
CELL 0.70000  48.628  72.233  85.080   90.00   90.00   90.00
ZERR       4   0.049   0.072   0.085    0.00    0.00    0.00
```

TITL　タイトル（76文字以内）
CELL　$\lambda$, $a$, $b$, $c$, $\alpha$, $\beta$, $\gamma$
ZERR　Z, esd($a$), esd($b$), esd($c$), esd($\alpha$), esd($\beta$), esd($\gamma$)

```
REM Space group P 21 21 21

LATT -1
SYMM x+0.5, -y+0.5, -z
SYMM -x, y+0.5, -z+0.5
SYMM -x+0.5, -y, z+0.5
```

REM　コメント行

LATT　空間格子の型、単純格子は“1”、対象心の無い場合（タンパク質）は負
SYMM　対象操作

# insファイルの内容 2

```
SFAC  C H N O P S
DISP $C     0.00328    0.00164     10.97565
DISP $H    -0.00002    0.00000      0.66864
                        ⋮
UNIT  6248 0 1664 4084 8 48
```

SFAC　モデルに含まれる原子種
DISP　$f'$, $f''$, 吸収断面積（Cu $K\alpha$, Mo $K\alpha$など使用時は省略可）
UNIT　結晶格子中の原子数

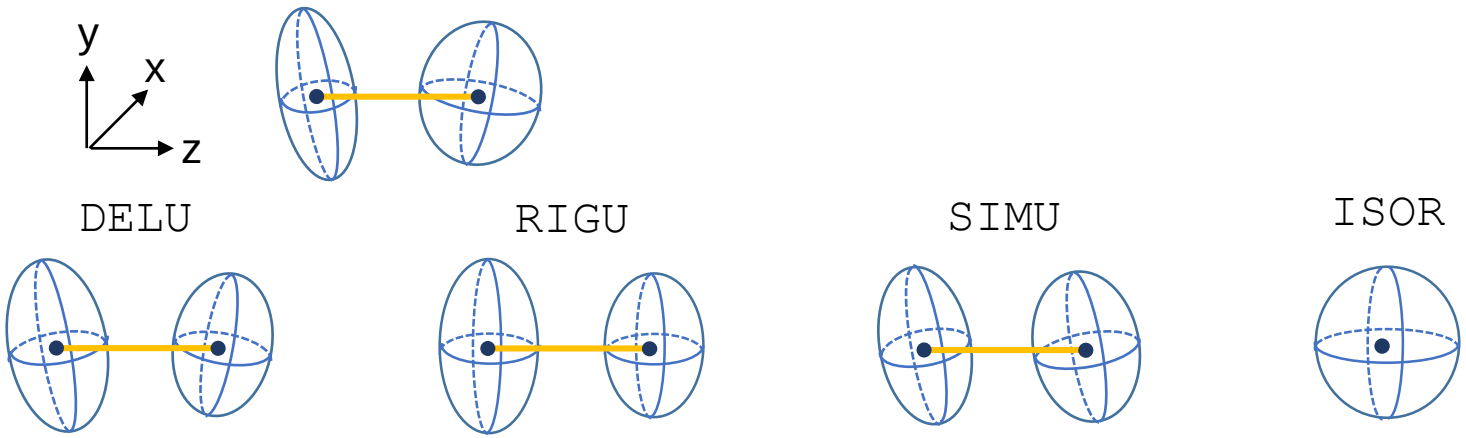
```
DEFS 0.02 0.1 0.01 0.04
CGLS 20 -1
SHEL 50 1.5
ANIS
FMAP 2
PLAN 200 2.3
LIST 6
WPDB 2
```

DEFS　制約の標準偏差のdefault値（DFIX, CHIV, DELU, SIMU）
CGLS　共役勾配法による精密化（サイクル数、-1の反射を$R_{free}$の計算に使用）
SHEL　分解能範囲
ANIS　非等方性温度因子の適用
FMAP　フーリエ合成の種類（“2”は$F_o$-$F_c$）
PLAN　フーリエ合成で探すピーク数、O, N等の原子からの距離（2.3 Å<$d$<4.0 Å）
LIST　反射リスト（.fcf）の種類（“6”の時Cootで読込み可）
WPDB　“1”（等方性）、“2”（非等方性）、負の時水素をPDBファイルに書出し

# insファイルの内容　3

```
RIGU
DELU $C_* $N_* $O_* $S_* $P_*
SIMU 0.1 $C_* $N_* $O_* $S_* $P_*
XNPD 0.001
SWAT
MERG 4
ISOR_HOH 0.1 $O
CONN_HOH 0 O
```

RIGU,DELU,SIMU,ISOR　　温度因子の制約
XNPD　　温度因子が非正定値にならないように下限値を設定
SWAT　　バルク溶媒補正
MERG　　反射データの平均化、“4”の時フリーデル対を含め平均化
CONN　　非水素原子間の結合を設定、上記の場合水分子とタンパク質間の結合を回避



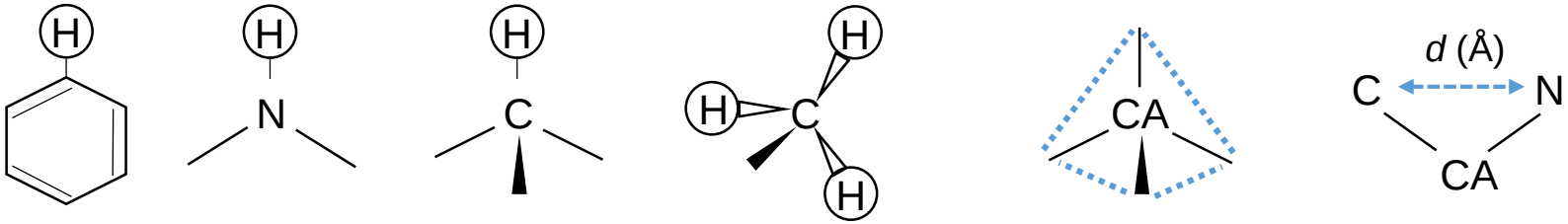

$\Delta U_{zz} = 0$　　$\Delta U_{zz} = 0, \Delta U_{yz} = 0, \Delta U_{xz} = 0$

# insファイルの内容 4

```
RTAB_* Omeg CA_+ N_+ C CA
RTAB_* Phi C_- N CA C
RTAB_* Psi N CA C N_+
            ⋮
REM HFIX_ALA 43 N
REM HFIX_ALA 13 CA
REM HFIX_ALA 33 CB
            ⋮
CHIV_ALA C
CHIV_ALA 2.477 CA
DFIX_ALA 1.231 C O
            ⋮
DANG_ALA 2.462 C N
            ⋮
```

| | |
|---|---|
| RTAB | 指定した原子間の距離、角度等をlstファイルに出力 |
| HFIX | 水素原子の発生、一桁目の3はriding model使用時 |
| CHIV | キラル体積（指定した原子に結合する4原子が形成する四面体の体積）の制約 |
| DFIX | 結合距離の制約 |
| DANG | 結合角度の制約 |

HFIX 43　　HFIX 13　　HFIX 33　　CHIV　　DANG



# insファイルの内容　5

```
WGHT     0.100000
FVAR     0.05337   0.5000  0.5000 0.5000
            ⋮
```

WGHT　　反射の重み付け係数

$$w_h = 1/\{\sigma^2(I_o) + (aP)^2 + bP\}$$ （default $a$ = 0.1, $b$ = 0）

$$P = f\{\max(0 \text{ or } I_o)\} + (1-f)I_c$$ （default $f$ = 1/3）

FVAR　　スケール因子（1番目）、自由変数（1番目以外）

```
PART 2
CA   1   0.327029  0.216338  0.083769  -21.0000  0.10552  0.09058 =
         0.08658    0.01954    0.02162    0.02402
            ⋮
PART 0
C    1   0.335856  0.236547  0.084662  11.00000  0.08949  0.08826 =
         0.10989    0.03225    0.01479    0.02792
            ⋮
HKLF 4
END
```

# insファイルの内容　5

```
WGHT     0.100000
FVAR     0.05337   0.5000  0.5000 0.5000
            ⋮
RESI A:20   ASP
N    3   0.309465  0.212651  0.097415  11.00000  0.11078  0.07261 =
         0.09419    0.01200    0.02966    0.02546
```

RESI　　アミノ酸残基（リガンド、水分子等）の指定
“A:20”　Chain A, 20番目の残基（Chain IDはSHELX-2017で対応）
→　Coot v.0.8.9で読み込み可能
以前のバージョンのSHELXでは1020（chain Aは+1000）などとしていた

原子名、SFAC番号、$x$, $y$, $z$ (fractional)、占有率、$U_{11}$（iso のとき$U_{iso}$）, $U_{22}$, $U_{33}$, $U_{23}$, $U_{13}$, $U_{12}$
“=”は改行後の行も同一行とみなす
占有率の“11.00000”は1に固定を示す

```
C    1   0.335856  0.236547  0.084662  11.00000  0.08949  0.08826 =
         0.10989    0.03225    0.01479    0.02792
            ⋮
HKLF 4
END
```

# insファイルの内容　5

```
WGHT    0.100000
FVAR    0.05337   0.5000  0.5000 0.5000
            ⋮
RESI A:20   ASP
N    3   0.309465  0.212651  0.097415  11.00000  0.11078  0.07261 =
         0.09419    0.01200    0.02966    0.02546
AFIX  43
H0   2   0.301141  0.221550  0.102154  11.00000 -1.20000
AFIX   0
```

AFIX 43~AFIX 0 直前の原子に対する水素原子の取り扱いを指定

原子名、SFAC番号、*x*, *y*, *z* (fractional)、占有率、$U_{iso}$

$U_{iso}$の“-1.20000”は直前の原子の$U_{eq}$の1.2倍となるような制約

OHや$CH_3$などでは1.5倍とする

$$U_{eq} = \frac{1}{3}\{U_{11}(aa^*)^2 + U_{22}(bb^*)^2 + U_{33}(cc^*)^2 + 2U_{12}a^*b^*\cos\gamma + 2U_{13}a^*c^*\cos\beta + 2U_{23}b^*c^*\cos\alpha\}$$

```
            ⋮
HKLF 4
END
```

# insファイルの内容　5

```
WGHT     0.100000
FVAR     0.05337    0.5000   0.5000 0.5000
            ⋮
RESI A:20    ASP
N     3    0.309465  0.212651  0.097415  11.00000  0.11078  0.07261 =
           0.09419    0.01200    0.02966    0.02546
AFIX   43
H0    2    0.301141  0.221550  0.102154  11.00000 -1.20000
AFIX    0
PART 1
CA    1    0.325963  0.216598  0.083393  21.0000  0.10418  0.08829 =
           0.09001    0.02210    0.02400    0.02804
            ⋮
PART 2
CA    1    0.327029  0.216338  0.083769  -21.0000  0.10552  0.09058 =
           0.08658    0.01954    0.02162    0.02402
            ⋮
PART 0
```

`PART 1,PART 2,PART 0`　ディスオーダー（マルチコンフォメーション）部分の指定

原子の占有率“21.0000”は`FVAR`の2番目の値を参照することを示す
“-21.0000”は（1-`FVAR(2)`）を示し、占有率の合計が1となるようにしている

# insファイルの内容　5

```
WGHT    0.100000
FVAR    0.05337   0.5000  0.5000 0.5000
           ⋮
RESI A:20   ASP
N    3   0.309465  0.212651  0.097415  11.00000  0.11078  0.07261 =
         0.09419    0.01200    0.02966    0.02546
AFIX  43
H0   2   0.301141  0.221550  0.102154  11.00000 -1.20000
AFIX   0
PART 1
CA   1   0.325963  0.216598  0.083393  21.0000  0.10418  0.08829 =
         0.09001    0.02210    0.02400    0.02804
           ⋮
PART 2
CA   1   0.327029  0.216338  0.083769  -21.0000  0.10552  0.09058 =
```

HKLF　反射データファイルフォーマットの指定
“4”の時は（$h\ k\ l\ I_o\ \sigma(I_o)$ flag）

```
           ⋮
HKLF 4
END
```

# lstファイルの内容

読み込んだinsファイルの内容

原子間結合のリスト

最小二乗法の各サイクルのまとめ

最終サイクル後の座標、温度因子

- 最終サイクル後の*R*値，S値等の結果

結合距離、角度等のリスト

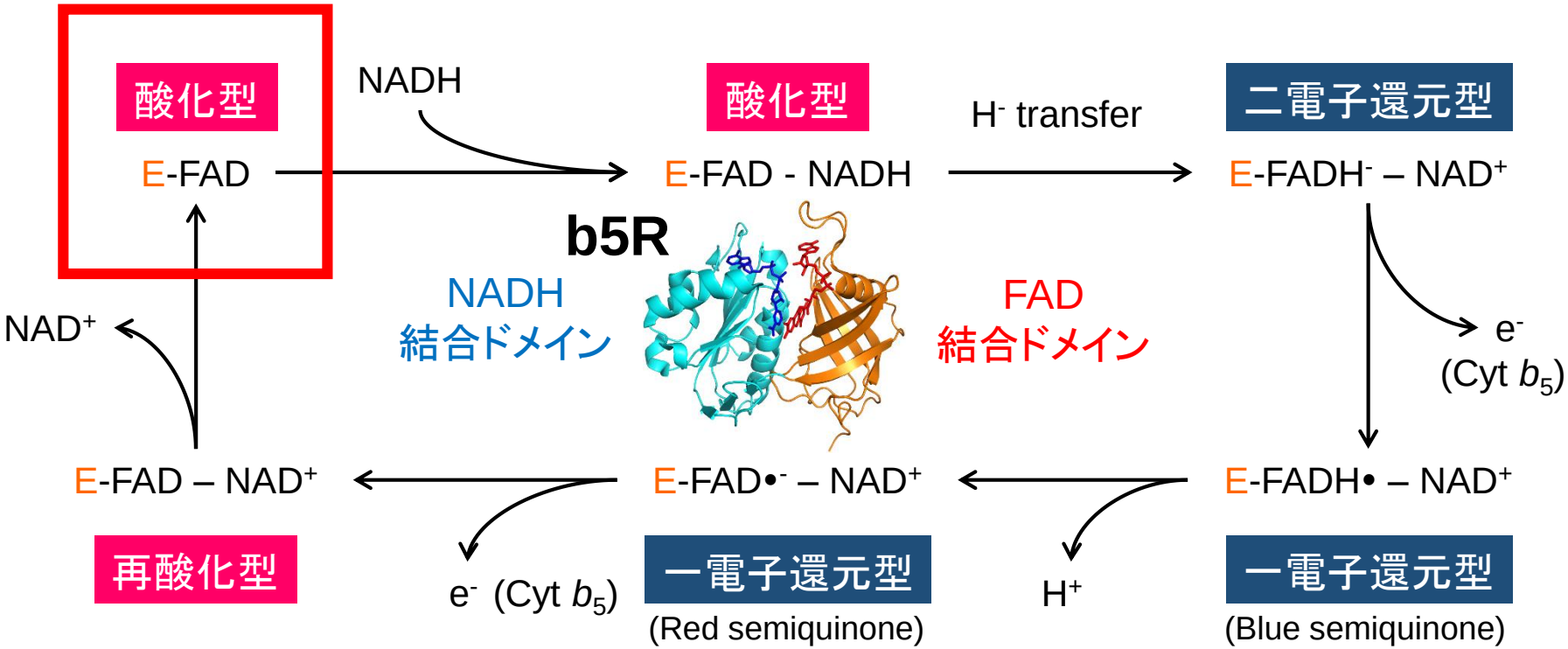
- フーリエ合成で検出したピークのリスト

resファイルの末尾にも出力

S値（Goodness of fit）

$$S = \sqrt{\frac{\sum_{hkl} w(hkl)\,\{|F_o(hkl)|^2 - |F_c(hkl)|^2\}^2}{N_o - N_p}}$$

# 3. 実例（NADHシトクロム$b_5$還元酵素を例にして）



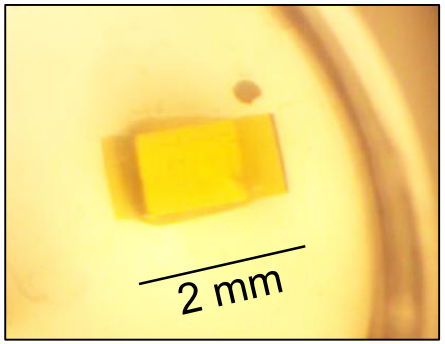

アミノ酸残基数：272
酸化型はFADを結合
非対称単位中に1分子

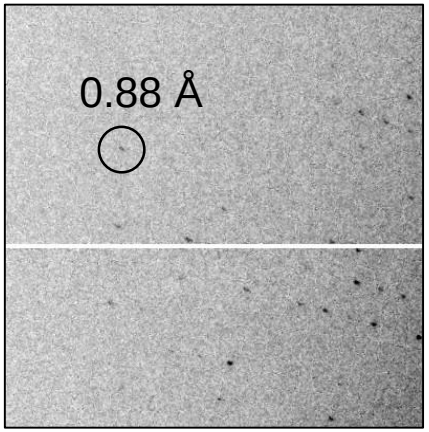
<u>X-ray crystal structures of b5R from porcine liver</u>

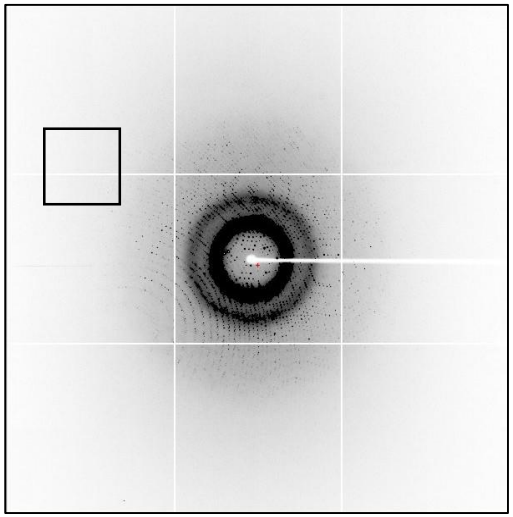
M. Yamada *et al.*, *J. Mol. Biol.*, 425, 4295-4306 (2013). Oxidized @0.78 Å, Fully-Reduced @1.68 Å

K. Takaba *et al.*, *Sci. Rep.*, 7, 43162 (2017). Oxidized @0.78 Å

# b5RのX線回折データ



PF BL-5において回折データ収集
Wavelength　0.76 Å
Detector　ADSC Q315
Distance　100 mm

統計値

| | | |
|---|---|---|
| Resolution range (Å) | | 50-0.87 |
| | | (0.89-0.87) |
| Space group | | $P2_12_12_1$ |
| Cell parameters (Å) | *a* | 48.6 |
| | *b* | 72.1 |
| | *c* | 84.8 |
| Completeness (%) | | 100 (100) |
| $I/\sigma(I)$ | | 76.1 (4.4) |
| $R_{meas}$ (%) | | 6.0 (45.5) |

# b5Rの精密化の流れ

1. *PHENIX*プログラムによる精密化
   （50-1.5 Å分解能、等方性温度因子）

2. *SHELXL*へ移行 $R_{work}$ = 18.7%, $R_{free}$ = 23.0%
   （ 50-1.5 Å分解能、Rigid body refinement）
   剛体の範囲を`AFIX 6 ~ AFIX 0`で囲む

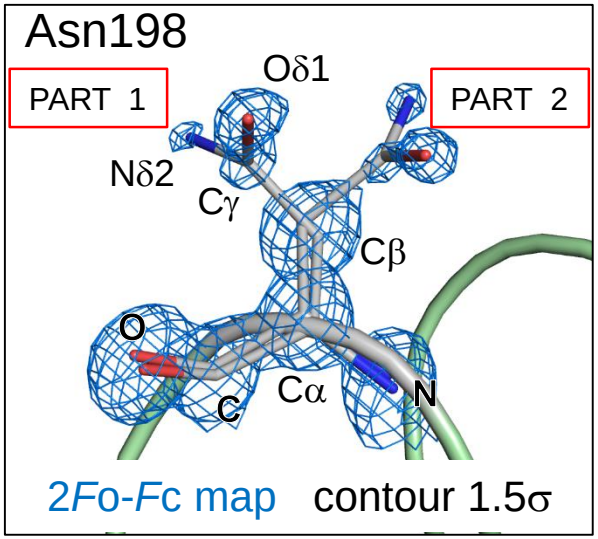
3. 分解能の拡張、マルチコンフォメーションの追加 $R_{work}$ = 17.9%, $R_{free}$ = 20.4%
   （ 50-0.9 Å分解能、等方性温度因子）

4. 分解能の拡張、非等方性温度因子の適用 $R_{work}$ = 13.9%, $R_{free}$ = 16.9%
   （ 50-0.87 Å分解能）

5. 制約の除外、水素原子の導入 $R_{work}$ = 13.1%, $R_{free}$ = 15.8%
   （ 50-0.87 Å分解能）

6. 完全行列最小二乗法によるesdの計算
   （ 50-0.87 Å分解能）

# マルチコンフォメーションと水素原子の取り扱い

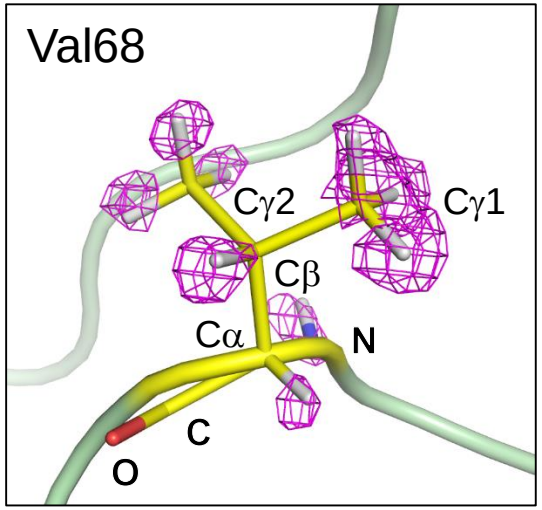



2*Fo-Fc* map　contour 1.5σ

```
RESI A:198   ASN
PART    1
N   3   0.128026   0.442065 -0.157955 251.00000   0.08620   0.06016 =
                ⋮
PART    2
N   3   0.130257   0.440773 -0.159804 -251.00000   0.12621   0.08844 =
                ⋮
PART    0
```

FVARの25番目の値　0.52104

➡ PART 1の占有率:0.521、PART 2の占有率:0.479




*Fo-Fc* H omit map　contour 2.5σ

```
RESI A:68   VAL
                ⋮
CB   1   0.169112   0.362774   0.029087   11.00000   0.07826   0.07741 =
          0.07281     0.01071     0.00458     0.00521
AFIX   13
HB   2   0.164677   0.362215   0.017819   11.00000   -1.20000
AFIX    0
CG1  1   0.183985   0.380865   0.031791   11.00000   0.08496   0.07917 =
          0.08095     0.02153     0.00553     0.00089
AFIX   33
HG1A 2   0.171917   0.391006   0.029328   11.00000   -1.50000
HG1B 2   0.189550   0.381682   0.042636   11.00000   -1.50000
HG1C 2   0.199952   0.381420   0.025130   11.00000   -1.50000
AFIX    0
                ⋮
```

# 完全行列最小二乗法によるesdの計算

1. 最小二乗法のコマンド　　CGLS → L.S.
   L.S.　1　（サイクル数1とする）

2. 変数をシフトさせず、標準偏差のみを見積もる
   DAMP　0　0
   （パラメータ数多い時はBLOC 1も使用し原子座標の標準偏差のみ計算）

3. 全ての制約を除外する
   （結合距離、温度因子に関するものなど全て）

4. 結合距離（角度）とesdの出力コマンド　　BOND
   （水素原子を含む結合については BOND $H）

5. *SHELXL*の実行

   shelxl␣refine-ls␣-b30000␣-t4 ↵

   -bオプション：パラメータ数（defaultは3000）　　$Param \times \sqrt{thred} < 65500$

   b5R計算時間　　約1時間20分
   （Core i7-4930k CPU @3.4GHz、パラメータ数: 28020、thred数：4）

# おわりに

高分解能データ（$d_{min}$ ~ 1 Å）が得られた際には、

*SHELXL*を用いた精密化を検討してみてはいかがでしょうか

謝辞

データ収集でお世話になったPFビームラインスタッフの皆さまに感謝いたします