Structure refinement with SHELX

Roland Pfoh

rpfoh@uhnres.utoronto.ca

Saskatoon Synchrotron Summer School V
June 16, 2010

SHELX History

- Originally, SHELX-76 was designed to refine small molecule and mineral structures, which usually diffract to high resolutions (better than 1.2 Å).
- The program is still widely used for this purpose. The standard SHELX reference is **cited more than 10 times every day** an has become the most highly cited scientific article of the last five years in all subjects:

  G.M. Sheldrick (2008). A short history of SHELX. *Acta Cryst.* **A64**, 112-122.
- After addition of some features SHELXL proved to be useful for the refinement of well diffracting macromolecules (Resolution should be better than 2 Å).

Purpose of SHELXL

- Small molecule crystal structures tend to have a lot of standard problems: complicated disorder, disorder involving symmetry elements, crystal twinning, different cations sharing the same site, hydrogen atoms.
- At higher resolution, those issues become visible also in proteins, and can be conveniently described in SHELXL.
- SHELXL is not suitable for resolutions worse than 2 Å. Unlike Refmac and Phenix-Refine torsion angles on side-chains are not restrained, which leads to bad rotamers at lower resolution structures. Also the solvent model and the anti-bumping restraints are a weakness in SHELXL when compared to other refinement programs. (But a new version is planned!!!). Switching between different refinement programs can be useful.
- For the refinement of macromolecules, SHELXL still “thinks” more in atoms than in aminoacid or nucleobase residues. It is very easy to address single atoms with different commands.

File handling

- Two files are needed for the refinement:

name.hkl: contains exclusively the reflection list.

name.ins: contains the instructions for the refinement and the atomic coordinates.

- The refinement is controlled by the .ins instruction file. (The program documentation is useful for getting used to the format.) This file contains also the full list of restraints (e.g. on distances and angles).
- Different format files can be written out by SHELXL (e.g. .res ore .pdb files.) The resulting maps and coordinate files can be read into Coot. Coot can also write new .ins files.
- For conversion of files (e.g. creating an .ins file from a .pdb file) the program SHELXPRO can be used.
- Information about the refinement is written to a .lst file (statistics, tables for distances and angles, disagreeable restraints, suggestions for atoms that should be split).

Two cations sharing the same site

```
...
EXYZ CA1_381 NA1_381
EADP CA1_381 NA1_381
SFAC C H O N S Na Ca
FVAR 1.0 0.6
...
RESI 381 Cal
PART 1
Ca1 7 0.37041 0.34874 0.03824 21.0 0.20937
PART 2
Na1 6 0.37041 0.34874 0.03824 -21.0 0.20937
PART 0
...
```

Coordinates and displacement parameters (ADP's) for Ca1 and Na1 (residue #381) are **constrained** to be the same.

The element Na gets #6 assigned, Ca #7.

The value of free variable #2 is 0.6, which in this case describes the occupancy of Ca1. The value can change during refinement.

The occupancy of Ca is 1.0*(free variable #2) = 0.6

The occupancy of Na is 1.0*(1-[free variable #2]) = 0.4

Adding hydrogens to a disordered side-chain

```
...
HFIX_SER 23 CB
...
RESI 38 SER
N  3 0.77141 0.92838 0.00838 11.0 0.10928
CA 1 0.78863 0.97930 0.03849 11.0 0.13845
PART 1
CB 1 0.83746 1.04938 0.05383 41.0 0.10928
OG 4 0.89383 1.00293 0.02938 41.0 0.10928
PART 2
CB 1 0.84940 1.08494 0.05389 -41.0 0.10928
OG 4 0.87494 1.00103 0.03199 -41.0 0.10928
PART 0
...
```

This command will add hydrogen atoms to CB in the next refinement. The code 23 describes the way the hydrogens are treated. HFIX 147 could be used to search for the hydrogen in the electron density around OG in the case of very good data.

Disordered components have the same atom names but different PART numbers. The sum of the two occupancies is constrained to 1.0 by the use of free variable #4 in this case.

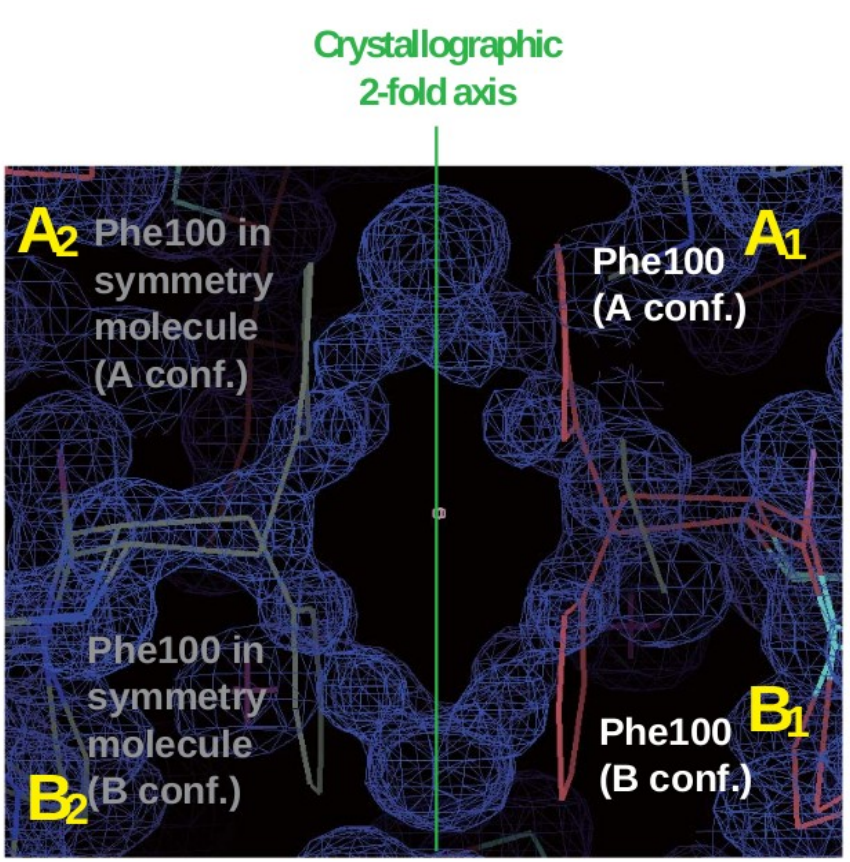
PART 0 is used for the non-disordered part of the molecule.

A threefold disordered side-chain

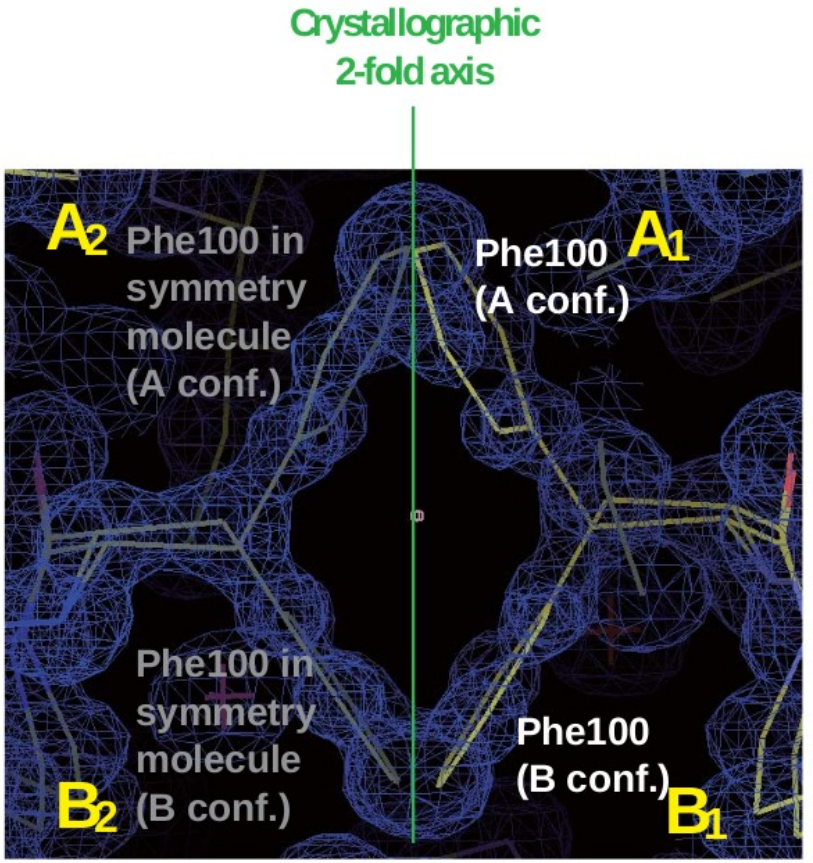
```
FVAR 1.0 0.5 0.3 0.2
SUMP 1.0 0.001 1.0 2 1.0 3 1.0 4
...
RESI 38 SER
PART 1
CB 1 0.83746 1.04938 0.05383 21.0 0.10928
OG 4 0.89383 1.00293 0.02938 21.0 0.10928
PART 2
CB 1 0.84940 1.08494 0.05389 31.0 0.10928
OG 4 0.87494 1.00103 0.03199 31.0 0.10928
PART 3
CB 1 0.84849 1.04440 0.05494 41.0 0.10928
OG 4 0.83044 1.00303 0.03303 41.0 0.10928
```

The SUMP command restrains the sum of the three occupancies to 1.0 (standard deviation 0.001) by the use of three free variables

Disorder close to symmetry elements



PART 1
...
PART 2

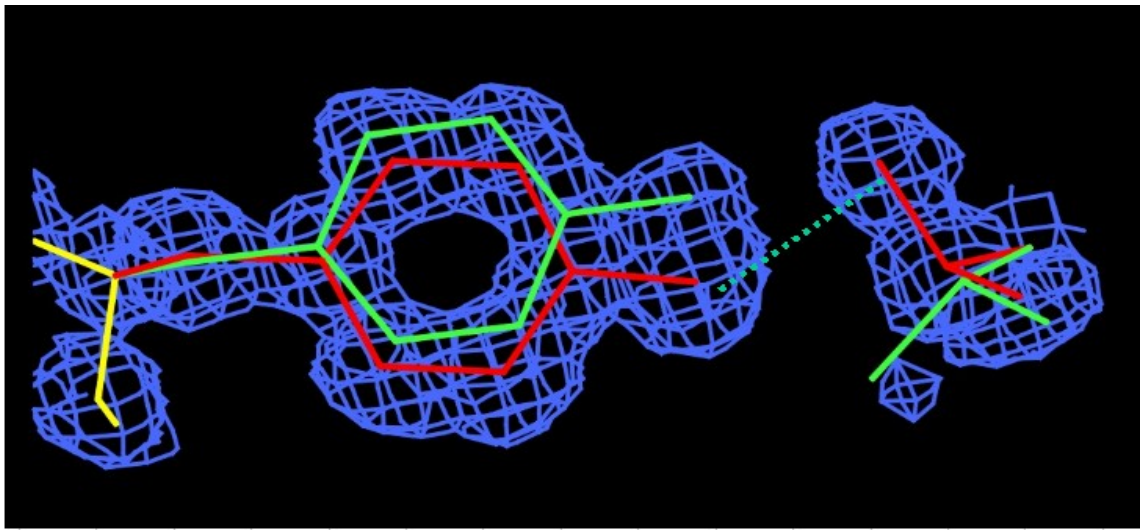


PART -1
...
PART -2

Masahiro Fujihashi, Kyoto

If disordered components have the same PART numbers close to symmetry elements, anti bumping restraints make them repel from each other. In this case, negative PART numbers can be used.

Disorder networks

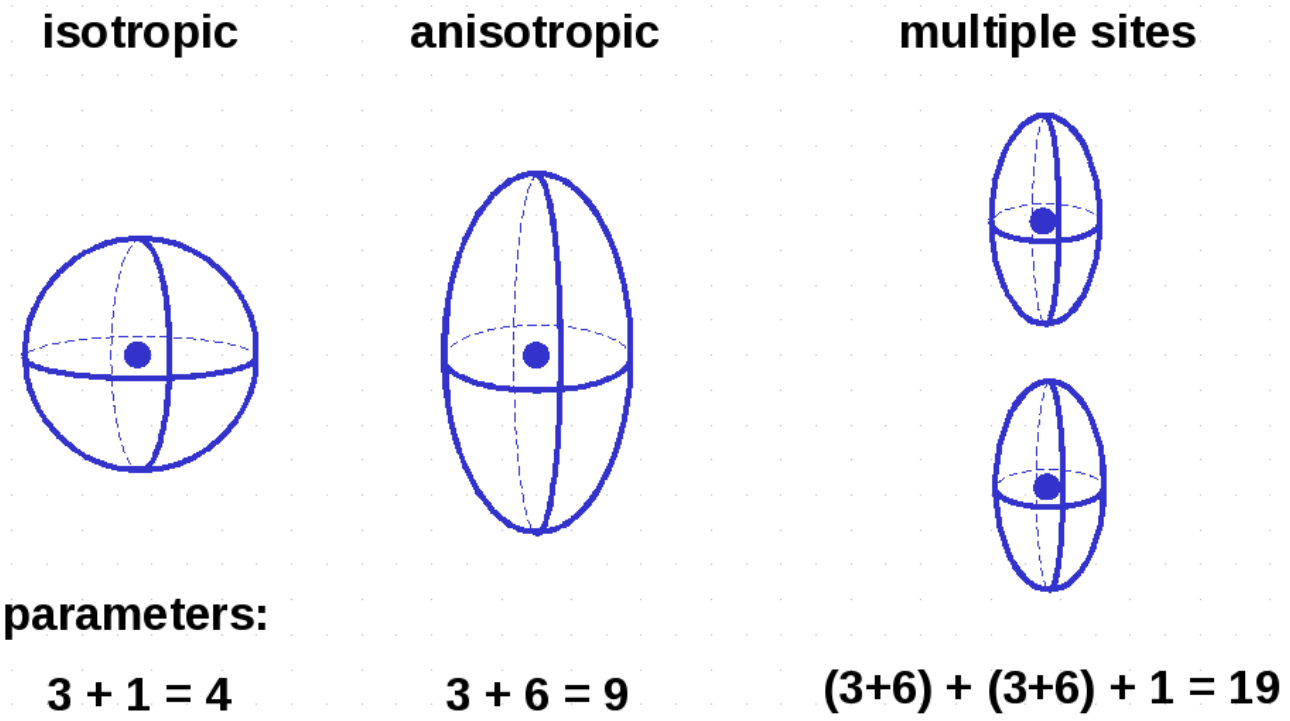


```
RESI 233 TYR
...
PART 1 31.0
...
PART 2 -31.0
...
PART 0
```

```
RESI 123 THR
...
PART 1 31.0
...
PART 2 -31.0
...
PART 0
```

1 parameter (free variable #3) describes the occupancies of 22 atoms.

B-factors and multiple sites



Thomas R. Schneider, Hamburg

Anisotropic displacement parameters are refined by using the command ANIS. It should only be used if the amount of data justifies a refinement of additional parameters. Single residues or atoms can be addressed separately with the ANIS command.

Restraints and Constraints

- Constraints are exact mathematical conditions that lead to a reduction in the number of parameters.
- Restraints have a standard deviation and can be regarded as additional observations (data).
- Both have a positive effect on the data:parameter ratio. (Data would be a measured intensity, a parameter would be a B-factor, for example.)
- If a constraint/restraint makes physically sense, it should be used (e.g. most hydrogen atoms can be placed at any resolution, riding hydrogen atoms cost no extra positional parameter.)

Types of restraint in SHELXL-97

DFIX, DANG and SADI - distances and ‘angle distances’

FLAT and CHIV - planarity and chiral volumes

BUMP - antibumping

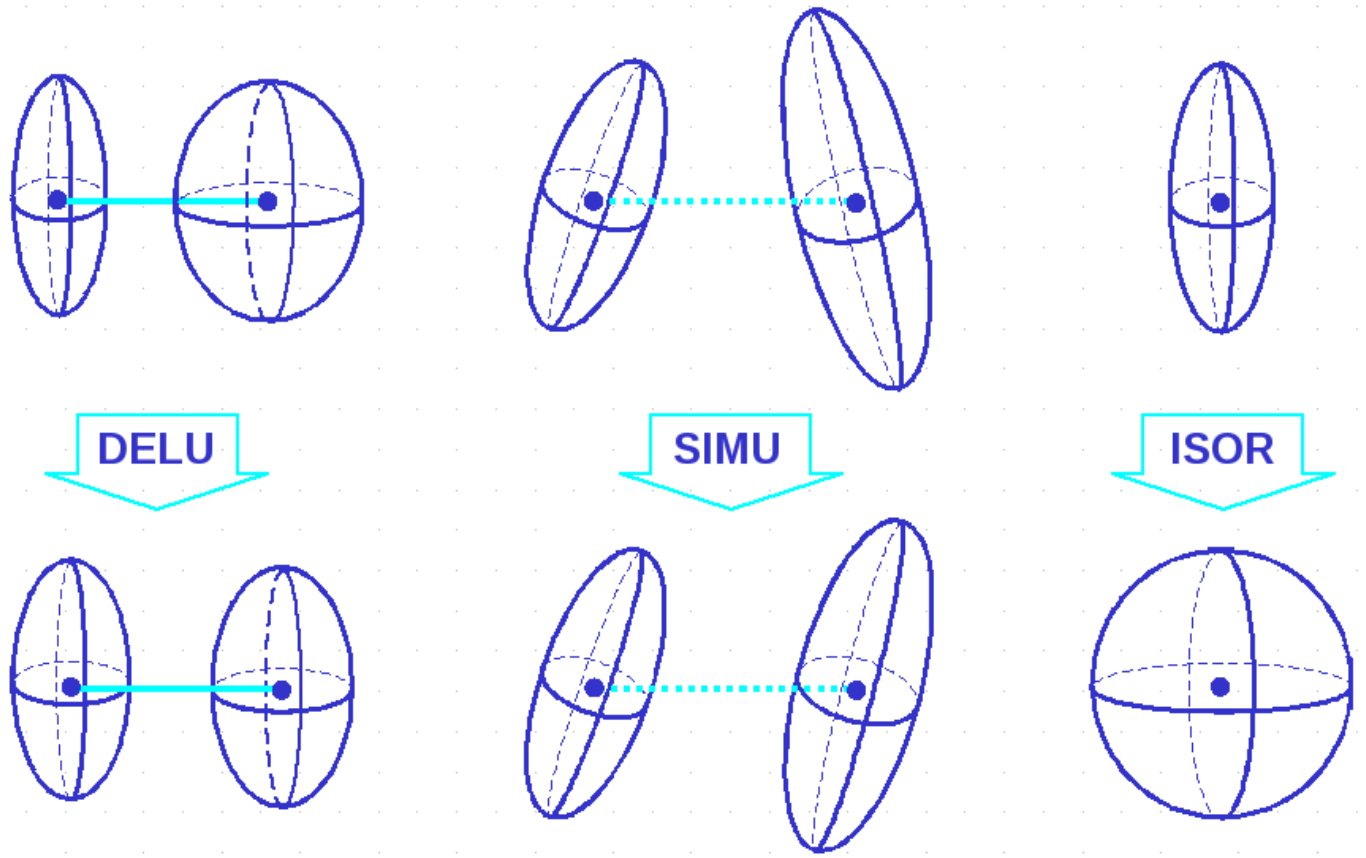
NCSY - non-crystallographic symmetry (NCS)

DELU, SIMU and ISOR - (an)isotropic displacements

SUMP - general ‘free variable’ restraint (e.g. for the sum of occupancies of side-chains with three disorder components)

DEFS sets default restraint esds and SAME can generate SADI restraints. CHIV, BUMP, NCSY and DELU make use of the connectivity array.

Restraints on ADP's



Thomas R. Schneider, Hamburg

It is useful to supplement this restraint by SIMU or ISOR.

Useful for disorder, six restraints.

Useful to get rid of non positive ADP's.

Similar distance restraints

Similar distance restraints assume that distances are equal, but without target values.

Example: A structure contains six phosphate ions which should be refined as regular tetrahedra with equal bond lengths, but we don't know what target value to use (it will be affected by pH and by libration):

```
SADI_PO4 P O1 P O2 P O3 P O4

SADI_PO4 O1 O2 O1 O3 O1 O4 O2 O3 O2 O4 O3 O4
```

The first line restraints all P-O distances to be equal, the second all O···O distances.

Sulfate ion on a twofold axis

```
FVAR 1.0 0.32
SADI_305 S O1 S O2
EQIV $1 2-x, 2-y, z
SADI_305 O1 O2 O1 O1_$1 O1 O2_$1 O2 O2_$1
RESI 305 SO4
S          ...    20.5
O1         ...    21.0
O2         ...    21.0
```

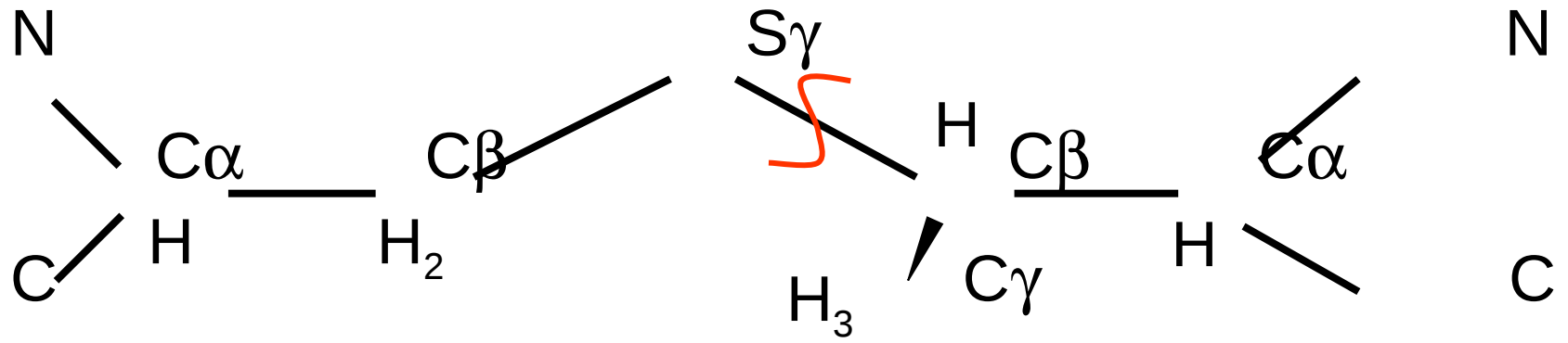
The EQIV command makes it possible to address atoms from molecules created by crystal symmetry.

Occupancy of Sulfur: 0.5 * (free_variable 2) = 0.16

Only the two independent S-O distances have to be restrained, the other two will then be equal by symmetry. In order to restrain the distances between O1 and its symmetry mate the symmetry operation $1 has to be defined by EQIV.

Example of similar distance restraints

Twinned crystals have a lower effective data to parameter ratio (especially if k is close to 0.5), so restraints are often required. Mersacidin has 6 molecules in the asymmetric unit and each contains 3 copies of the unusual monosulfide bridge:

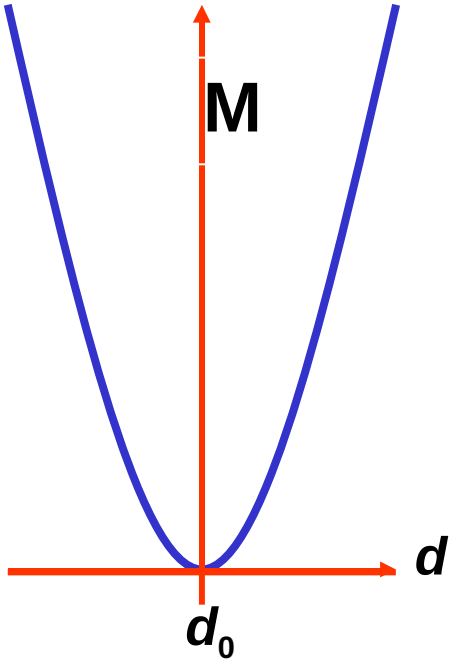


This is ideal for the application of SADI restraints to the 1,2- and 1,3-distances (equal distances, no target value), e.g.

```
SADI_Aba CA CB
SADI_Aba CA CG
SADI CB_2 SG_1  CB_4 SG_12  CB_13 SG_18  etc.
```

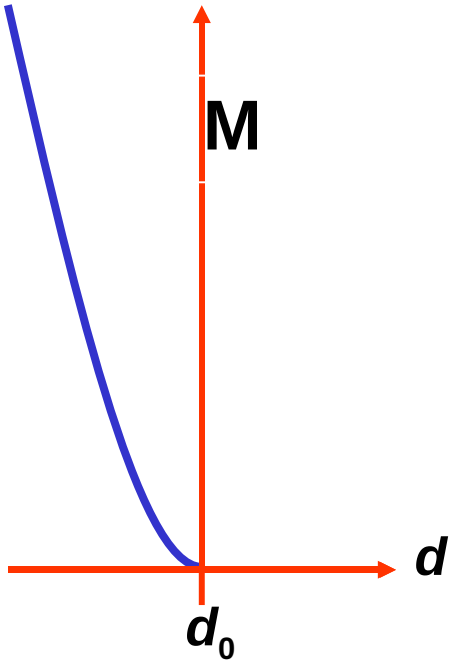
George M. Sheldrick, Göttingen

Distance and antibumping restraints



DFIX
DANG } $+ \Sigma\ (1/\sigma^2)\ (d - d_0)^2$

SADI $+ \Sigma\ (1/\sigma^2)\ (d_1 - d_2)^2$



BUMP

$+ \Sigma\ (1/\sigma^2)\ (d - d_0)^2$

if $d < d_0$

Symmetry equivalent molecules

If the structure contains a disulfide bond across a twofold axis, we need to add restraints for the distances across the symmetry element, and (important !) switch off the antibumping involving symmetry transformed molecules:

```
EQIV  $2  -x, y, -z
DFIX_61  2.031  SG SG_$2
DANG_61  3.035  CB SG_$2
BUMP  -0.02
```

With the negative sign before the esd symmetry equivalent atoms are regarded as connected.

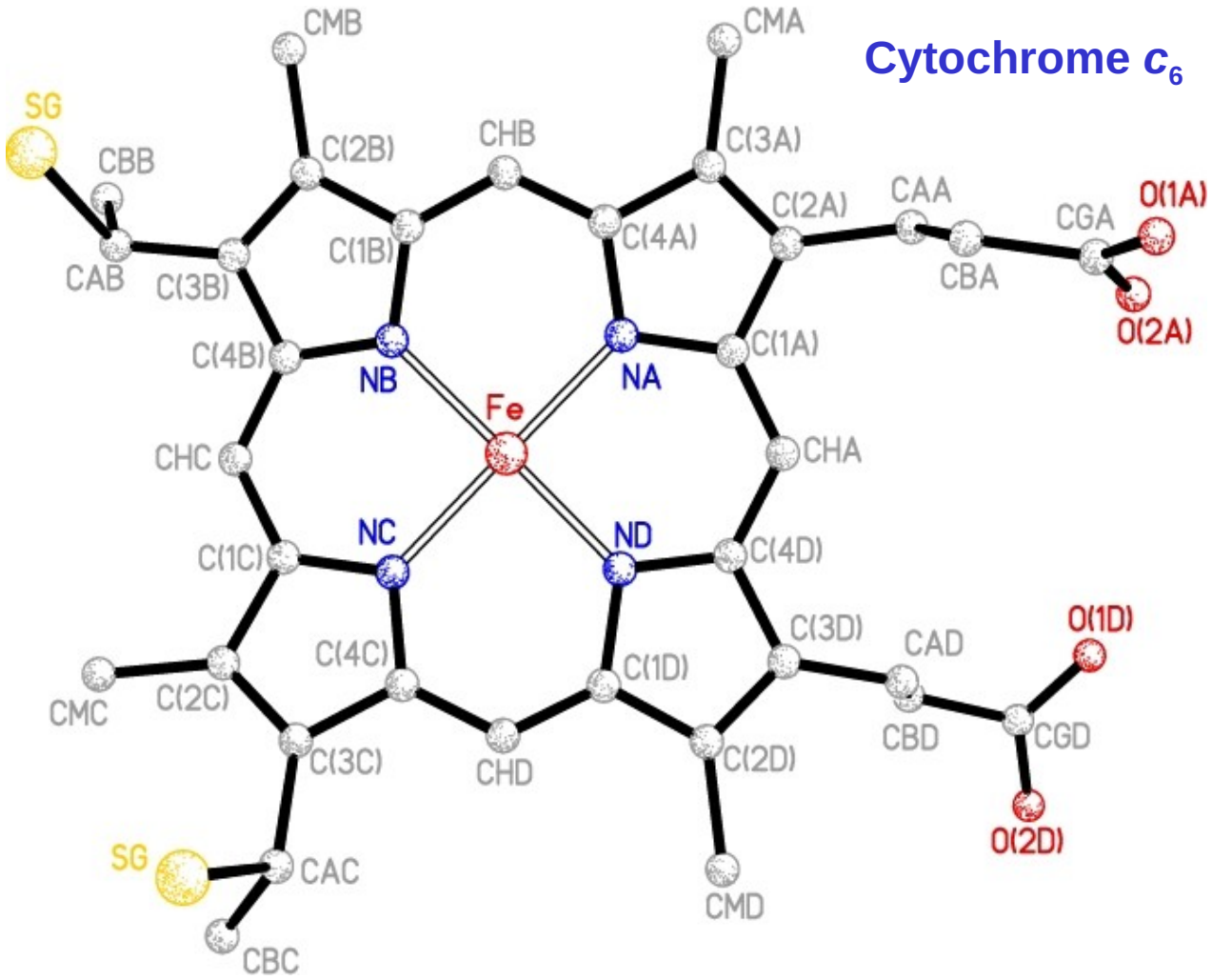
If the bond does not involve the same residue number in the two molecules (unlikely, but might be needed to model disorder), use has to be made of the rule that a local residue number has priority over a global one:

```
DFIX_134  2.031  SG_78 SG_$2
```

Anti-bumping restraints can also be generated by hand using DFIX instructions with negative distances d.

Use of free variables to obtain mean distances with esds

The following input refines fv 2, 3 and 4 to be the mean Fe-N, N-C and N…CH distances. Because of the 4- and 8-fold redundancy, accurate values are obtained that can be used as restraints.




```
FVAR 1.0 1.8 1.4 2.4
DFIX_HEM 21 Fe NA Fe NB Fe NC Fe ND
DFIX_HEM 31 NA C1A NA C4A NB C1B NB C4B NC C1C NC C4C ND C1D ND C4D
DFIX_HEM 41 NA CHA NA CHB NB CHB NB CHC NC CHC NC CHD ND CHD ND CHA
etc...
```

George M. Sheldrick, Göttingen

NCS restraints

- Non-crystallographic symmetry usually occurs when there is more than one molecule of the same kind in the asymmetric unit.
- In SHELXL the NCS-related 1,4-distances are restrained to be equal. This is similar to restraining torsion angles but does not distinguish between ± gauche positions.

```
NCXY 1000 N_1001 > OT2_1109
NCSY 2000 N_1001 > OT2_1109
```

would specify threefold NCS, where the three chains are numbered 1001-1109, 2001-2109 and 3001-3109.

How to find where to improve the model

- List of disagreeable restraints: SIMU / DELU
- High *B*-values
- Negative F_o–F_c density close to atoms
- Positive F_o–F_c density: check all peaks higher than 5σ
- 'May be split' messages
- 'Non-positive-definite' messages
- Ramachandran and Kleywegt (NCS) plots (SHELXPRO)

One should never forget that the structure determined by X-ray diffraction is based on the electron density averaged over ca. 10^{15} unit-cells and a few hours data collection time (radiation damage !). Our attempts to model it are still rather rudimentary; the disagreement between F_o and F_c can primarily be attributed to errors in the model, not in the X-ray data!

Acknowledgments

- Prof. George M. Sheldrick (University of Göttingen, Germany).

- Prof. Emil Pai (University of Toronto) and his group.

UNIVERSITY OF TORONTO

HFSP