

REFMAC5

Roberto A. Steiner

IFOM

Istituto FIRC di Oncologia Molecolare
Milan, Italy

Aim of this talk

Enable new users to get started with
REFMAC5

Organization

1

Refinement and overview of *REFMAC5*

- General
- Dictionary
- TLS

2

Demo

- Refinement with *REFMAC5*
- SKETCHER

GENERAL

What is *REFMAC5*?

REFMAC5 is a program for the refinement of macromolecular structures. It is distributed as part of the CCP4 suite (<http://www.ccp4.ac.uk/download.php>, http://www.ytbl.york.ac.uk/~garib/refmac/latest_refmac.html).

Some points about the program:

It is strongly based on ML and Bayesian statistics

It is highly optimized

It is very easy to use (CCP4)

It has an extensive built-in dictionary

It allows various tasks (model idealisation, rigid-body refinement, phased and non-phased restrained and unrestrained refinement)

It allows a flexible model parameterization (iso-, aniso-, mixed-ADPs, TLS, bulk solvent)

It exploits a good minimization algorithm

Selected *REFMAC5* references

[Murshudov, G.N. & al. (1997), Refinement of macromolecular structures by the maximum-likelihood method, Acta Cryst. D53, 240-255]

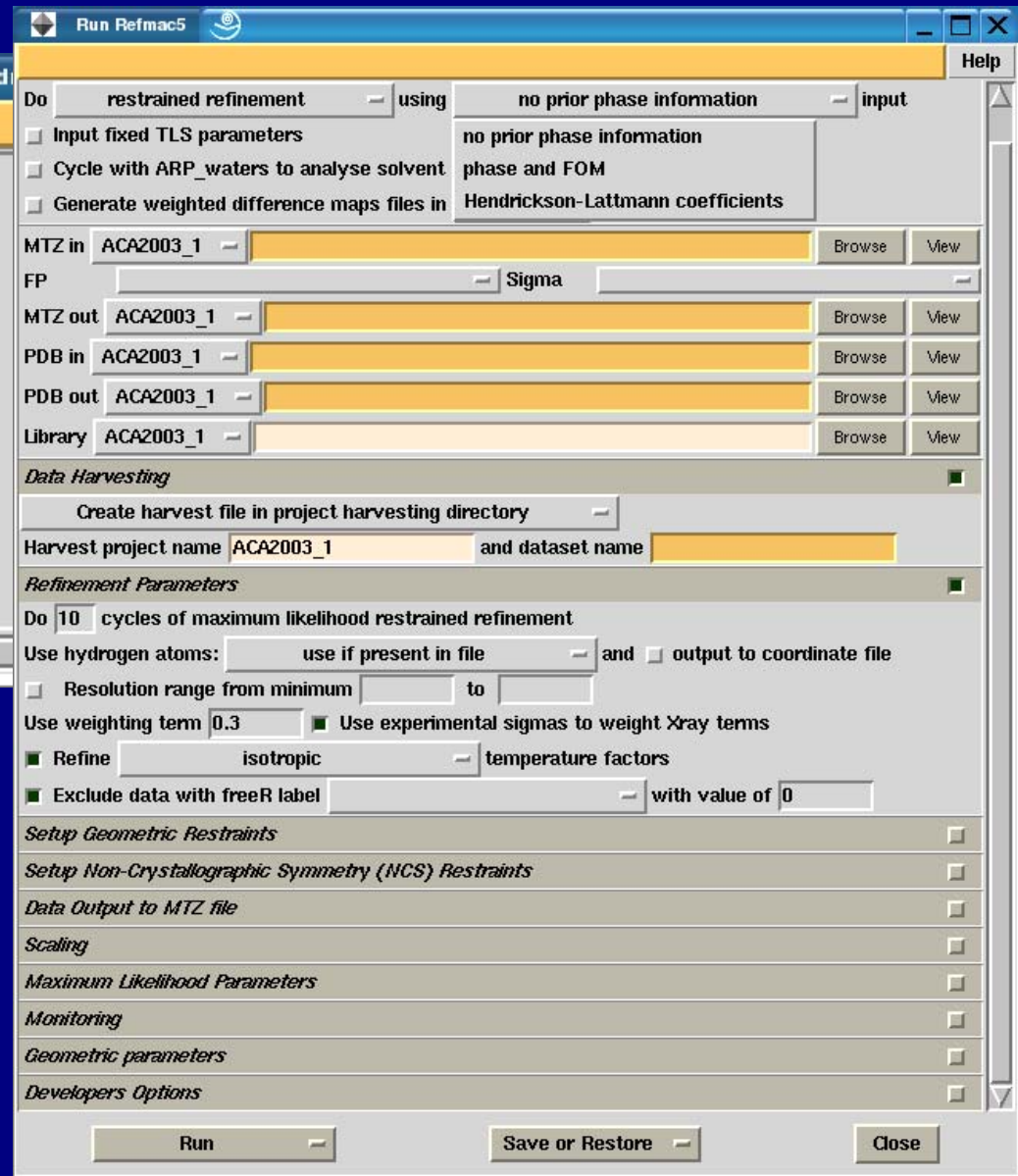
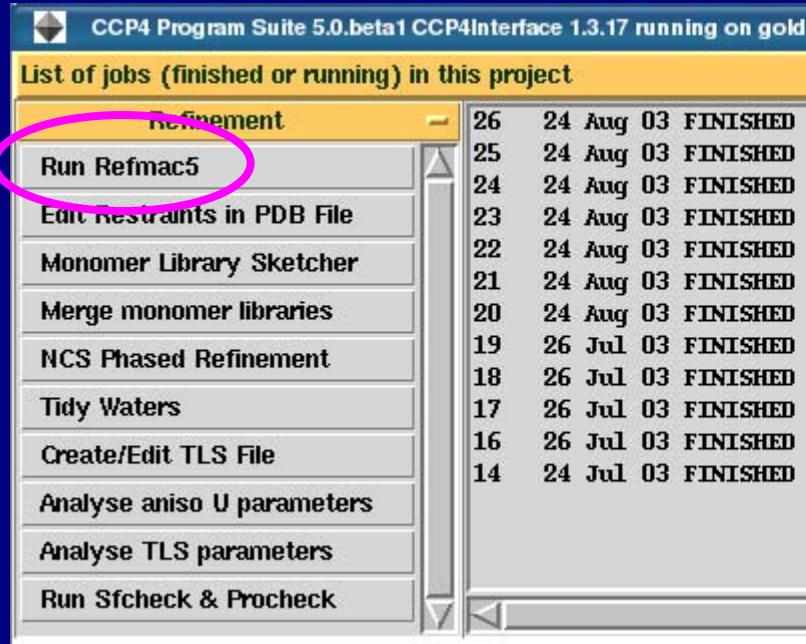
[Murshudov, G.N. & al. (1999), Efficient anisotropic refinement of macromolecular structures using FFT, Acta Cryst. D55, 247-255]

[Winn, M.D. & al. (2001), Use of TLS parameters to model anisotropic displacement parameters, Acta Cryst. D57, 122-133]

[Steiner, R.A. & al. (2003), Fisher's information in maximum-likelihood macromolecular crystallographic refinement, Acta Cryst. D59, 2114-2124]

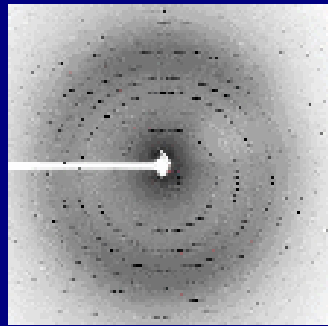
[Vagin, A.A. & al. (2004), REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use, Acta Cryst. D60, 2184-2195]

CCP4*i*



*[Potterton, E. & al. (2003),
A graphical user interface to
the CCP4 program suite,
Acta Cryst. D59, 1131-1137]*

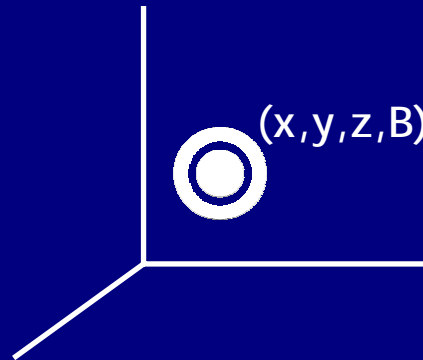
Crystallographic refinement



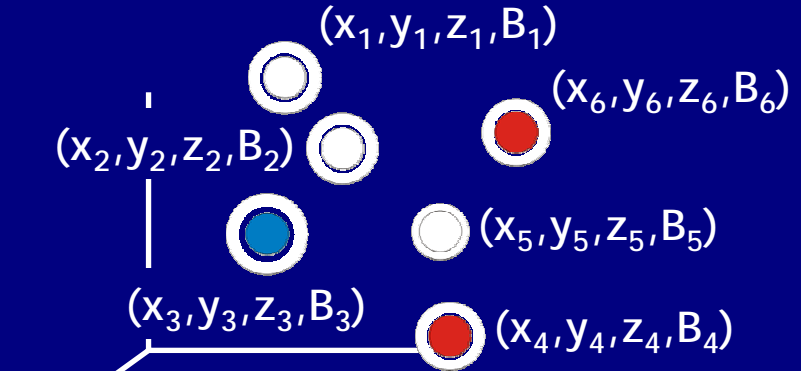
Data

$$\{ |F|_{\underline{h}_0}, \sigma_{\underline{h}_0} \}$$

Given



Model



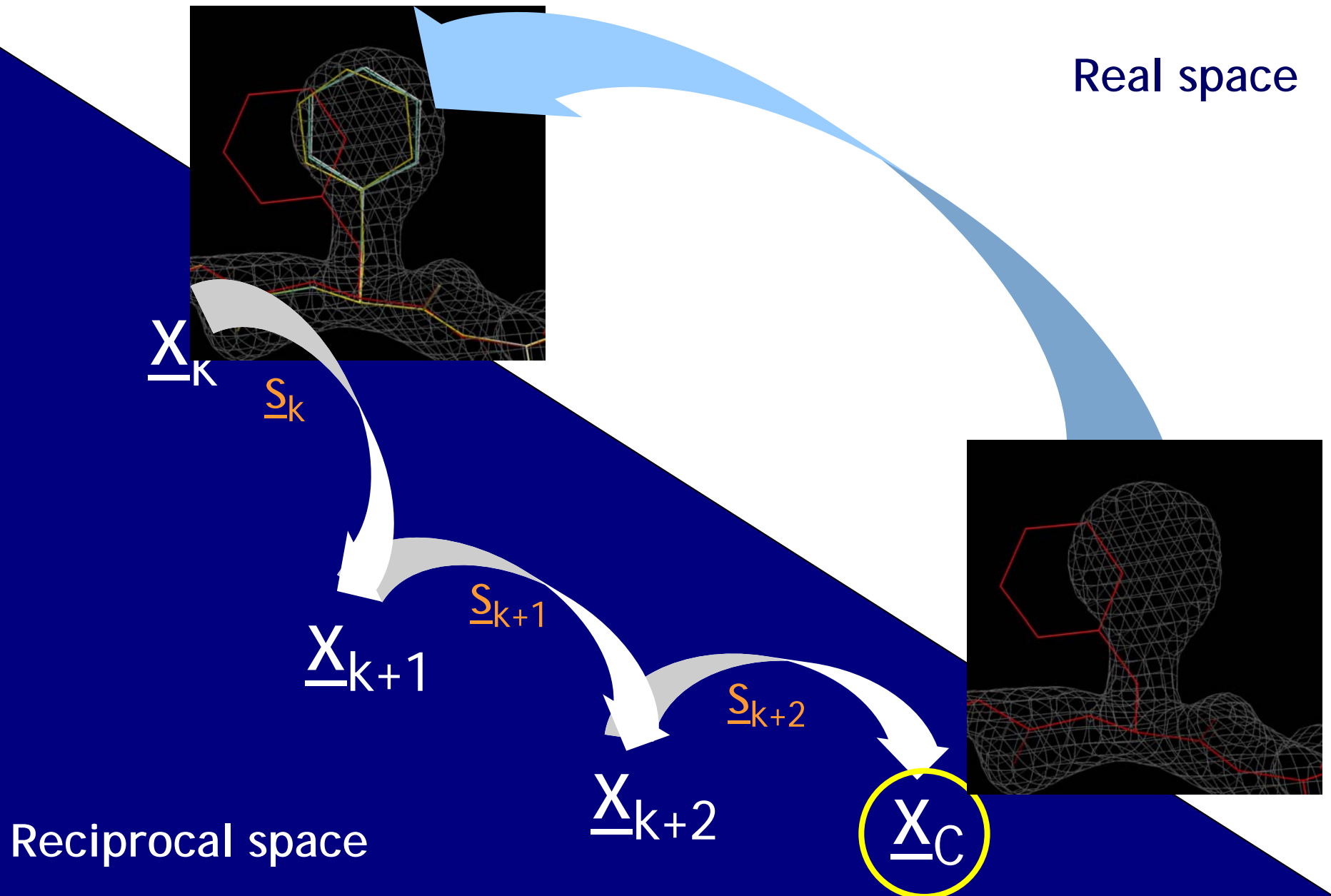
Initial values

Find

The model which fits the data best

The accuracy of such a model

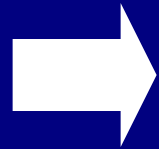
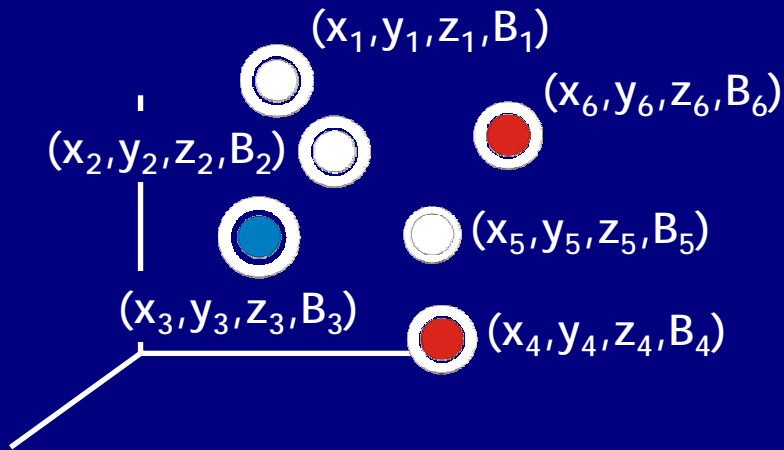
Crystallographic refinement in practice



Key aspects of refinement

- Objective function
- Prior knowledge
- Model-parametrization
- Optimization

Least-squares refinement



$$F_c(\underline{s}) = \sum_j f_j \exp(2\pi i \underline{r}_j \cdot \underline{s})$$

$$f = \sum_{\underline{h}} W(\underline{h}) (|F_o| - |F_c|)^2$$

+

$$\sum_b W(b) (Q_o - Q_c)^2$$

Least-squares
crystallographic function

Least-squares
restraints function

Bayesian flavor

Bayesian approach

The best model is the one which has the highest probability given a set of observations and a certain **prior knowledge**.

Bayes' theorem

$$P(M;O) = P(M)P(O;M)/P(O)$$

Maximum likelihood residual (posterior)

$$P(M;O) = P(M) \frac{P(O;M)/P(O)}{L(O;M)} = P(M)L(O;M)$$

$$\max P(M;O) \Leftrightarrow \min -\log P(M;O) = \min [-\log P(M) - \log L(O;M)]$$

[Pannu, N.S. & Read, R.J. (1997) Improved structure refinement through maximum-likelihood, Acta Cryst, A52, 659-668]

[Bricogne, G. (1997) Bayesian statistical viewpoint on structure determination: basic concepts and examples, Methods in Enzymology, 276]

[Murshudov, G.N. & al. (1997), Refinement of macromolecular structures by the maximum-likelihood method, Acta Cryst. D53, 240-255]

[McCoy, A.J. (2004) Linking likelihood, Acta Cryst. D60, 2169-2183]

DICTIONARY

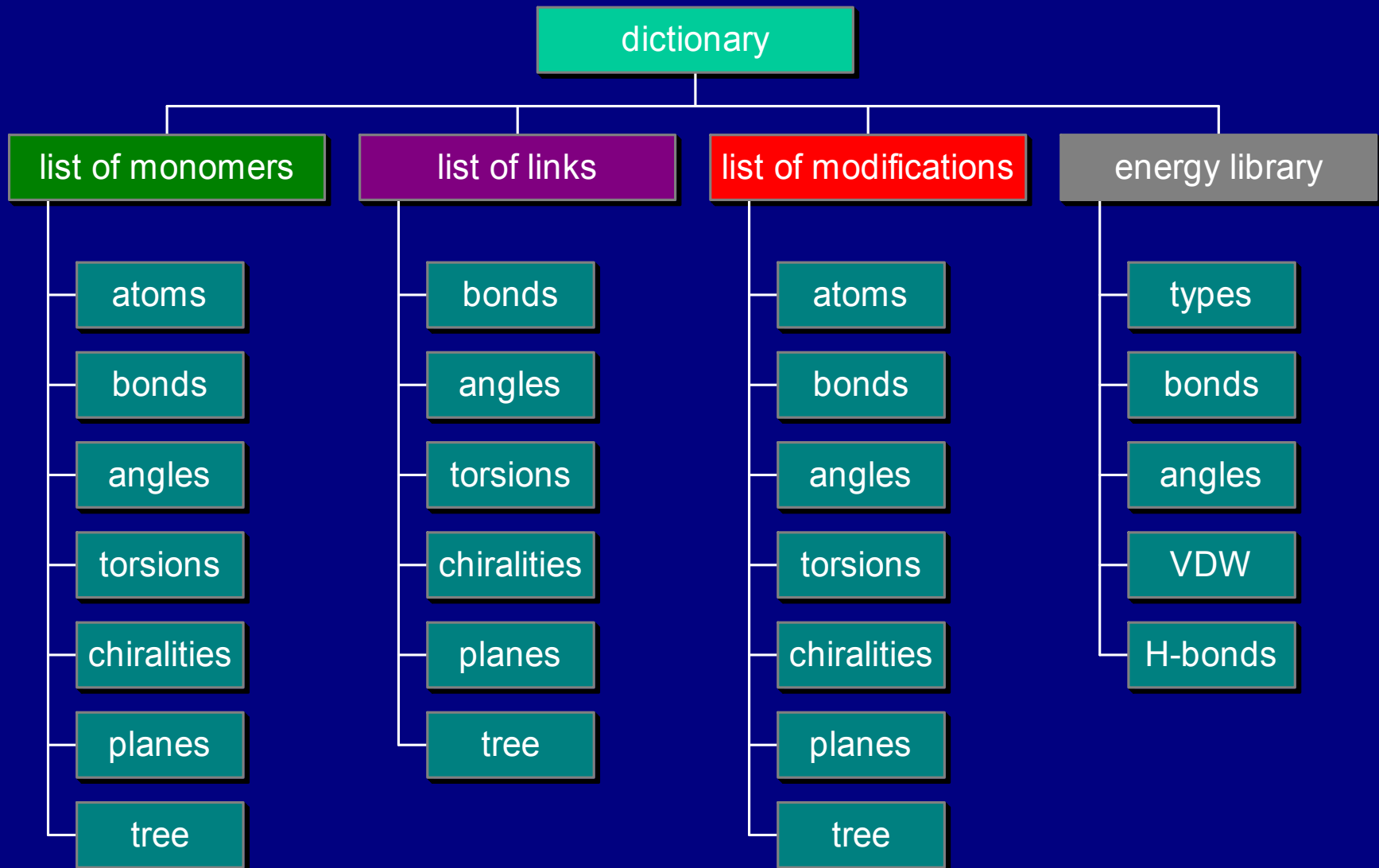
Dictionary

The use of prior knowledge requires its organized storage.

[\\$CCP4/html/mon_lib.html](#)

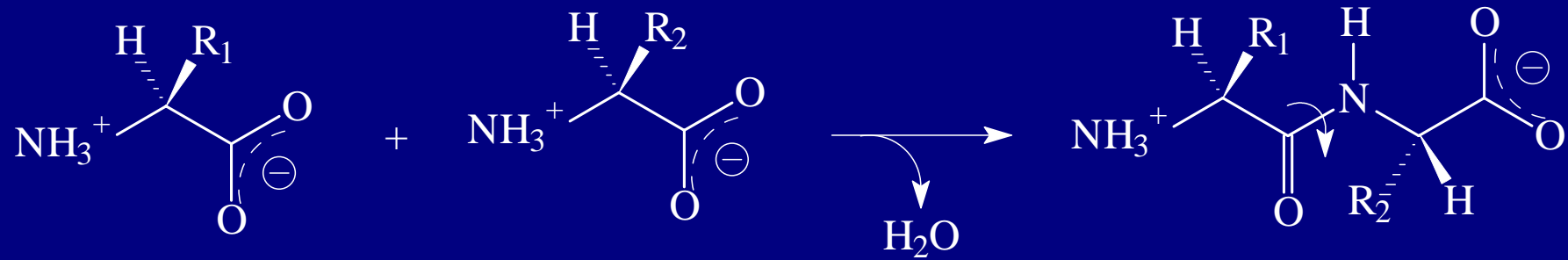
<http://www.ysbl.york.ac.uk/~alexei/dictionary.html>

Organization of dictionary

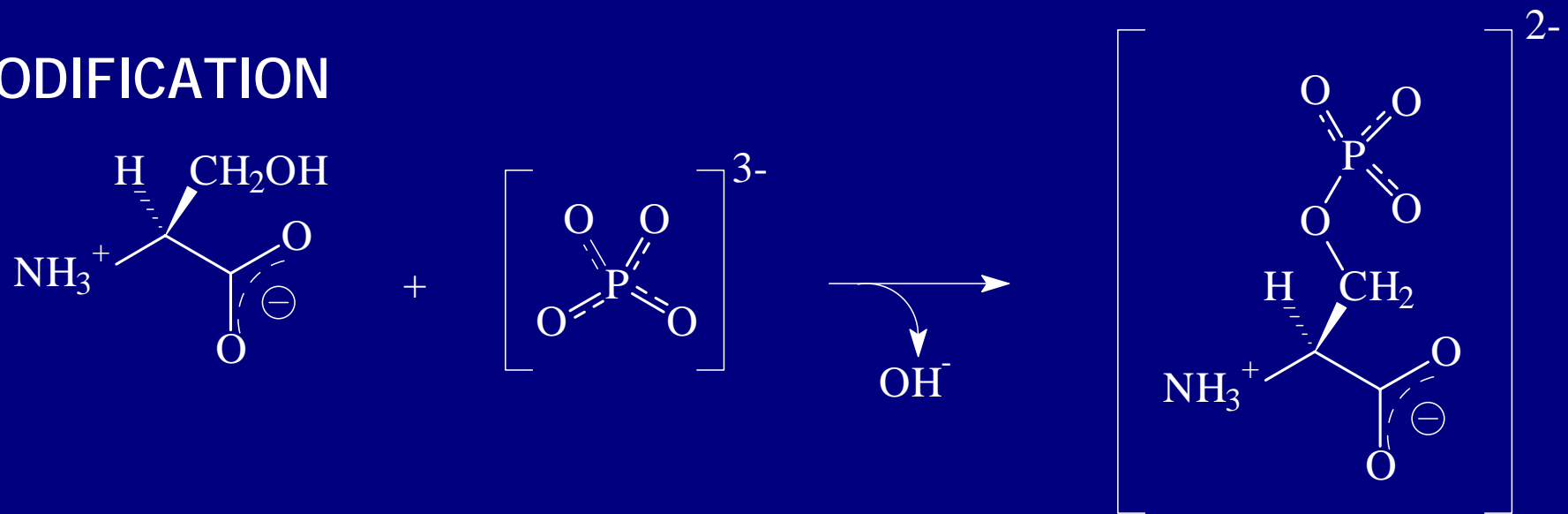


Links and Modifications

LINK



MODIFICATION



Monomer library

`$CCP4/lib/data/monomers/`

<code>ener_lib.cif</code>	definition of atom types
<code>mon_lib_list.html</code>	info
<code>0/,1/,...a/,b/,...</code>	definition of various monomers

Description of monomers

In the files:

a/A##.cif

Monomers are described by the following categories:

- _chem_comp
- _chem_comp_atom
- _chem_comp_bond
- _chem_comp_angle
- _chem_comp_tor
- _chem_comp_chir
- _chem_comp_plane_atom

Monomer library (`_chem_comp`)

```
loop_  
_chem_comp.id  
_chem_comp.three_letter_code  
_chem_comp.name  
_chem_comp.group  
_chem_comp.number_atoms_all  
_chem_comp.number_atoms_nh  
_chem_comp.desc_level
```

```
ALA    ALA    'ALANINE'    L-peptide    10    5    ○
```

Level of description
.
= COMPLETE
M
= MINIMAL



Monomer library (`_chem_comp_atom`)

```
loop_  
_chem_comp_atom.comp_id  
_chem_comp_atom.atom_id  
_chem_comp_atom.type_symbol  
_chem_comp_atom.type_energy  
_chem_comp_atom.partial_charge  
ALA      N      N      NH1      -0.204  
ALA      H      H      HNH1     0.204  
ALA      CA     C      CH1      0.058  
ALA      HA     H      HCH1     0.046  
ALA      CB     C      CH3      -0.120  
ALA      HB1    H      HCH3     0.040  
ALA      HB2    H      HCH3     0.040  
ALA      HB3    H      HCH3     0.040  
ALA      C      C      C        0.318  
ALA      O      O      O        -0.422
```

Monomer library (`_chem_comp_bond`)

```
loop_  
_chem_comp_bond.comp_id  
_chem_comp_bond.atom_id_1  
_chem_comp_bond.atom_id_2  
_chem_comp_bond.type  
_chem_comp_bond.value_dist  
_chem_comp_bond.value_dist_esd  
ALA      N      H      single      0.860      0.020  
ALA      N      CA     single      1.458      0.019  
ALA      CA     HA     single      0.980      0.020  
ALA      CA     CB     single      1.521      0.033  
ALA      CB     HB1    single      0.960      0.020  
ALA      CB     HB2    single      0.960      0.020  
ALA      CB     HB3    single      0.960      0.020  
ALA      CA     C      single      1.525      0.021  
ALA      C      O      double      1.231      0.020
```

Monomer library (`_chem_comp_chir`)

```
loop_  
_chem_comp_chir.comp_id  
_chem_comp_chir.id  
_chem_comp_chir.atom_id_centre  
_chem_comp_chir.atom_id_1  
_chem_comp_chir.atom_id_2  
_chem_comp_chir.atom_id_3  
_chem_comp_chir.volume_sign  
ALA chir_01 CA N CB C negativ
```

positiv, negativ, both, anomer

What happens when you run REFMAC5

You have a monomer for which there is a complete description

the program carries on and takes everything from the dictionary. Currently, there are about 1000 ligands with a complete description in the *REFMAC5* library. Cis-peptides, S-S bridges, sugar-, DNA-, RNA-links are automatically recognised.

You have a monomer for which there is only a minimal description or no description

No description or minimal description

In the case you have monomer(s) in your coordinate file for which there is no description (or minimal description) *REFMAC5* generates for you a complete library description (**monomer.cif**) and then it stops so you can check the result.

If you are satisfied you can use **monomer.cif** for refinement. The description generated in this way is good only if your coordinates are good (**CSD, EBI, any program that can do energy minimization**).

A more general approach for description generation requires the use of the graphical program **SKETCHER** from CCP4*i*. **SKETCHER** is a graphical interface to **LIBCHECK** which creates new monomer library descriptions
<http://www.ytbl.york.ac.uk/~alexei/libcheck.html>

Alternatively, you can use the **PRODRG2** server
<http://davapc1.bioch.dundee.ac.uk/programs/prodrg/prodrg.html>

SKETCHER

CCP4 Program Suite 5.0.beta

Monomer Library Sketcher

List of jobs (finished or running)

Refinement

- Run Refmac5
- Edit Restraints in PDB File
- Monomer Library Sketcher**
- Merge monomer libraries
- NCS Phased Refinement
- Tidy Waters
- Create/Edit TLS File
- Analyse aniso U parameters
- Analyse TLS parameters
- Run Sfcheck & Procheck

File Edit Help

MOUSE BUTTONS Left:rotate Right:drag Control-left:zoom Control-right:Select active atom
Shift-left:Click close to active atom to add fragment Shift-right:Click bond to change bond type

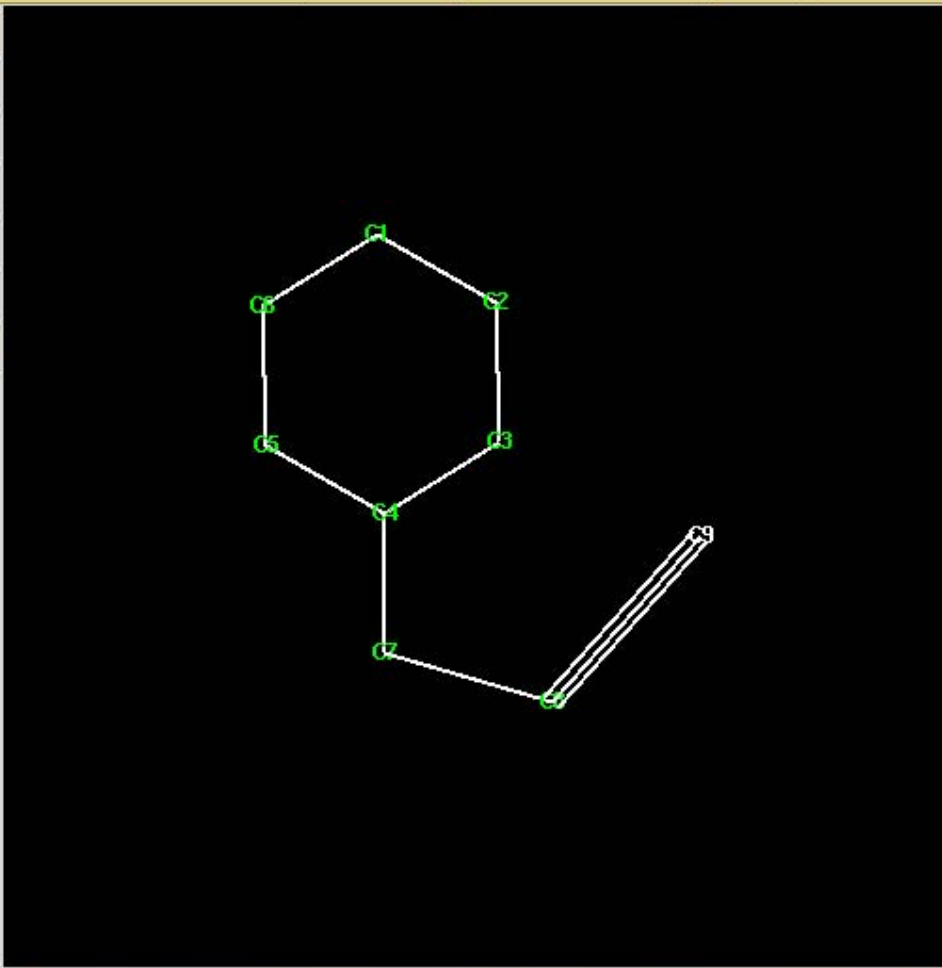
Do nothing

Undo last edit

Recentre View

Mouse mode

- Edit Monomer
- Move Fragment

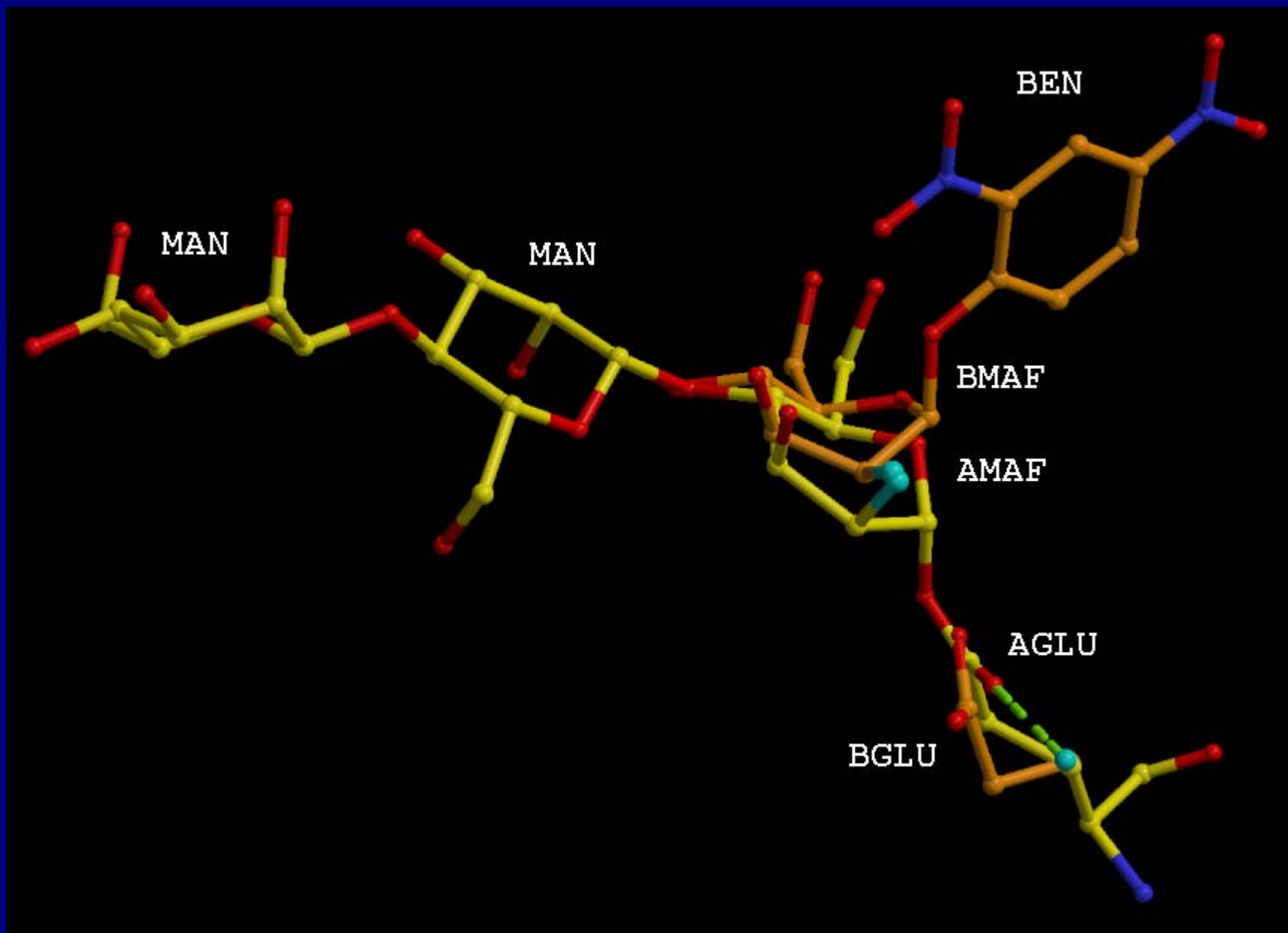


Element	Name	Ox
C	C1	0
C	C2	0
C	C3	0
C	C4	0
C	C5	0
C	C6	0
C	C7	0
C	C8	0
C	C9	0

Centre Sign B/3 F/4 1/5 2/6

Edit Table Add Row

REFMAC5 can handle complex descriptions



Links and Modifications in practice

```

0          1          2          3          4          5          6          7
123456789012345678901234567890123456789012345678901234567890123456789
LINK          C6  BBEN B   1          O1  BMAF S   2          BEN-MAF
LINK          OE2  GLU A   67          1.895  ZN      ZN R   5          GLU-ZN

LINK          GLY H  127          GLY H  133          gap

LINK          MAF S   2          MAN S   3          BETA1-4

SSBOND       1 CYS A   298          CYS A   298          4555

MODRES       MAN S   3  MAN-b-D          RENAME

```

TLS

TLS

ADPs are an important component of a macromolecule
Proper parameterisation
Biological significance

Displacements are likely anisotropic, but rarely we have the luxury of refining individual aniso- U . Instead iso- U are used.

TLS parameterisation allows an intermediate description

T = translation

L = libration

S = screw-motion

[Schomaker & Trueblood (1968) On the rigid-body motion of molecules in crystals Acta Cryst. B24, 63-76]

[Winn & al. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement Acta Cryst. D57, 122-133]

Decomposition of ADPs

$$U = U_{\text{cryst}} + U_{\text{TLS}} + U_{\text{int}} + U_{\text{atom}}$$

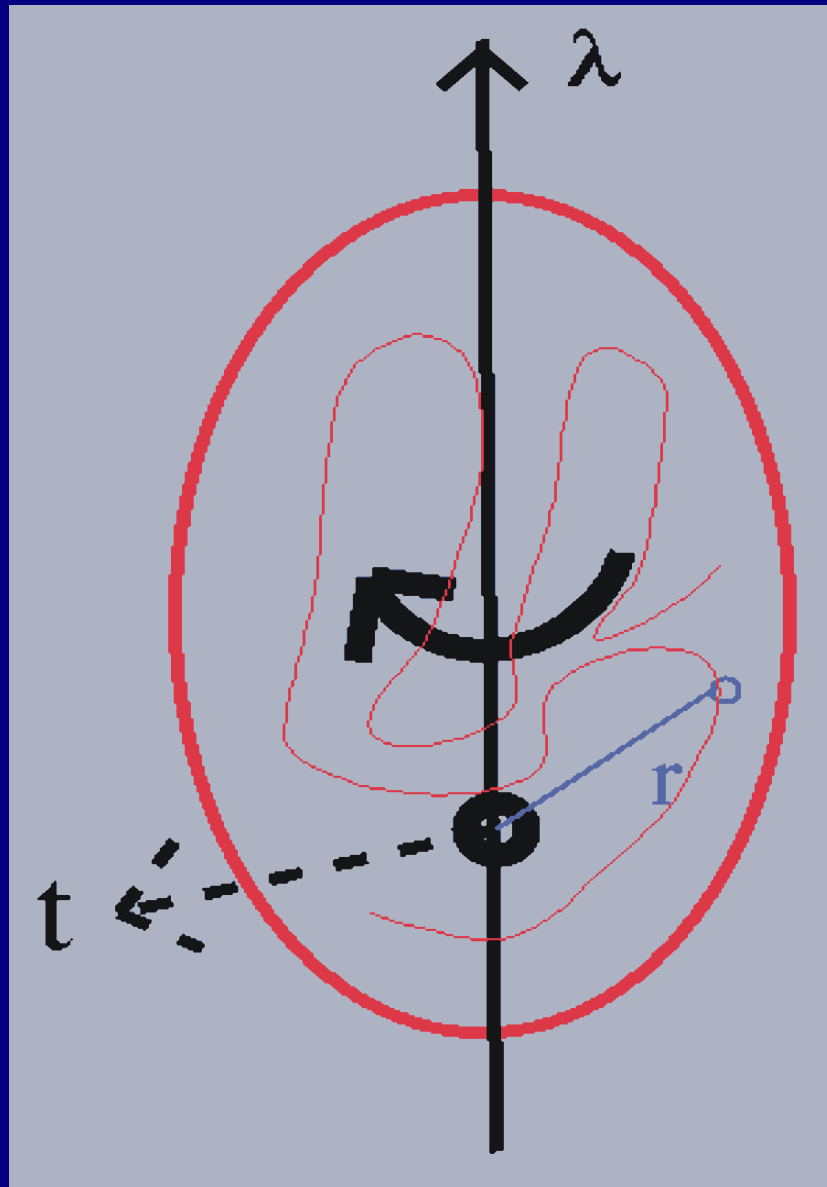
U_{cryst} : overall anisotropy of the crystal

U_{TLS} : TLS motions of pseudo-rigid bodies

U_{int} : collective torsional librations or internal normal modes

U_{atom} : individual atomic motions

Rigid-body motion



General displacement of a rigid-body point can be described as a rotation along an axis passing through a fixed point together with a translation of that fixed point.

$$\underline{u} = \underline{t} + D\underline{r}$$

for small librations

$$\underline{u} \approx \underline{t} + \underline{\lambda} \times \underline{r}$$

D = rotation matrix

$\underline{\lambda}$ = vector along the rotation axis of magnitude equal to the angle of rotation

TLS parameters

Dyad product:

$$\underline{uu}^T = \underline{tt}^T + \underline{t\lambda}^T \times \underline{r}^T - \underline{r} \times \underline{\lambda t}^T - \underline{r} \times \underline{\lambda\lambda}^T \times \underline{r}^T$$

ADPs are the time and space average

$$U_{\text{TLS}} = \langle \underline{uu}^T \rangle = \mathbf{T} + \mathbf{S}^T \times \underline{r}^T - \underline{r} \times \mathbf{S} - \underline{r} \times \mathbf{L} \times \underline{r}^T$$

$$\mathbf{T} = \langle \underline{tt}^T \rangle$$

6 parameters, **TRANSLATION**

$$\mathbf{L} = \langle \underline{\lambda\lambda}^T \rangle$$

6 parameters, **LIBRATION**

$$\mathbf{S} = \langle \underline{\lambda t}^T \rangle$$

8 parameters, **SCREW-ROTATION**

Use of TLS

analysis: given individual aniso-ADPs fit TLS parameters

[Harata, K. & Kanai, R., (2002) Crystallographic dissection of the thermal motion of protein-sugar complex, Proteins, 48, 53-62]

[Wilson, M.A. & Brunger, A.T., (2000) The 1.0 Å crystal structure of Ca(2+)-bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity, J. Mol. Biol. 301, 1237-1256]

[Harata, K. et al., (1999) Crystallographic evaluation of internal motion of human α -lactalbumin refined by full-matrix least-squares method, J. Mol. Biol., 26, 347-358]

refinement: TLS as refinement parameters

[Winn et al., (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions Methods Enzymol., 374, 300-321]

[Papiz, M.Z. et al., (2003) The structure and thermal motion of the B800-850 LH2 complex fromJ. Mol. Biol., 326, 1523-1538]

[Howlin et al., (1989) Segmented anisotropic refinement of bovine ribonuclease A by the application of the rigid-body TLS model, Acta Cryst., A45, 851-861]

Choice of TLS groups and resolution

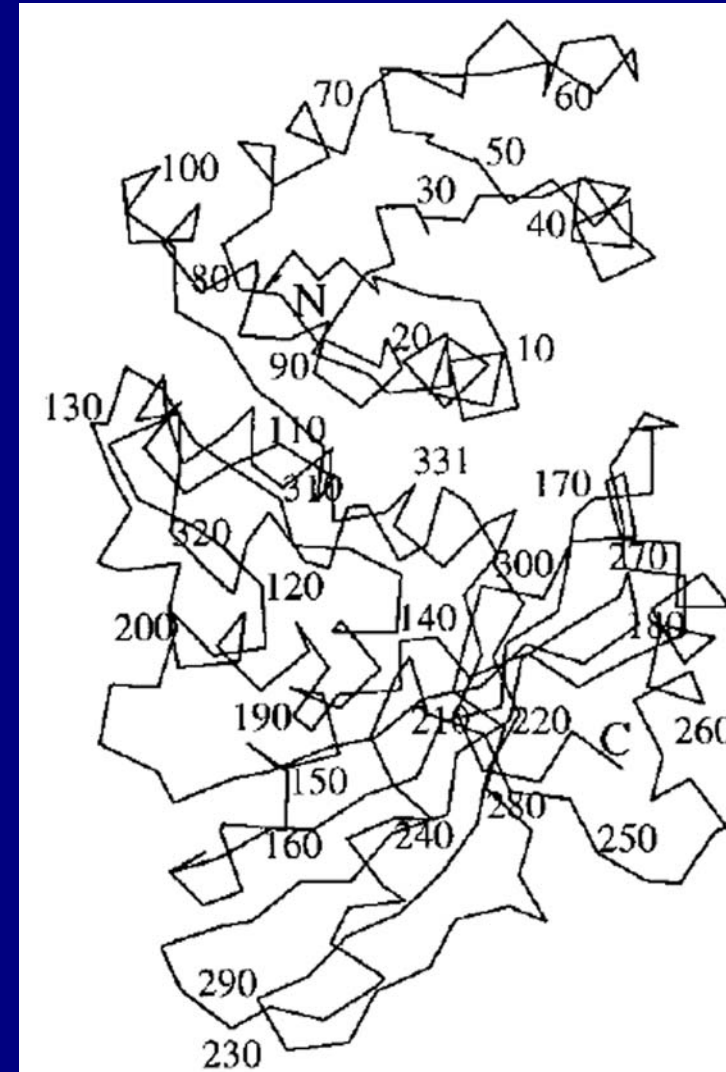
Choice chains, domains, secondary structure, elements,...

Resolution not a big problem. There are only 20 more parameters per TLS group

Thioredoxin reductase 3.0 Å [*Sandalova, T. & al., (2001)*
3D-structure of a mammalian thioredoxin reductase: implications for mechanism and evolution of a selenocysteine-dependent enzyme, PNAS, 98, 9533-9538]
6 TLS groups (1 for each of 6 monomers in asu)

Example GAPDH

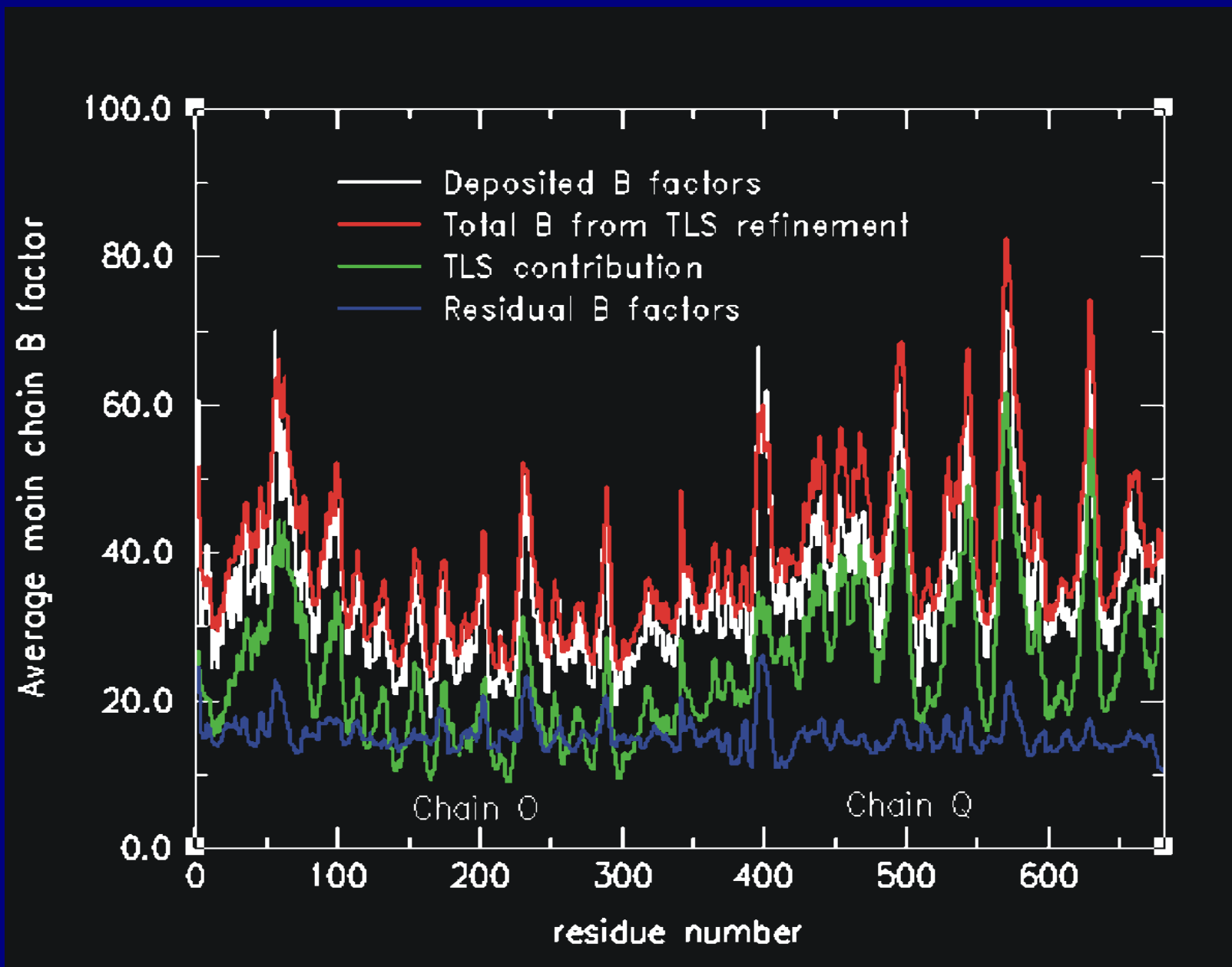
- Glyceraldehyde-3-phosphate dehydrogenase from *Sulfolobus solfataricus* [Isupov, M. & al. (1999), Crystal structure of the glyceraldehyde-3-phosphate dehydrogenase from *Sulfolobus solfataricus*, *J. Mol. Biol.*, 291, 651-660]
- 340 amino acids
- 2 molecules in asymmetric unit (O and Q)
- each molecule has a NAD-binding and a catalytic domain
- Data to 2.05Å



GAPDH before and after TLS

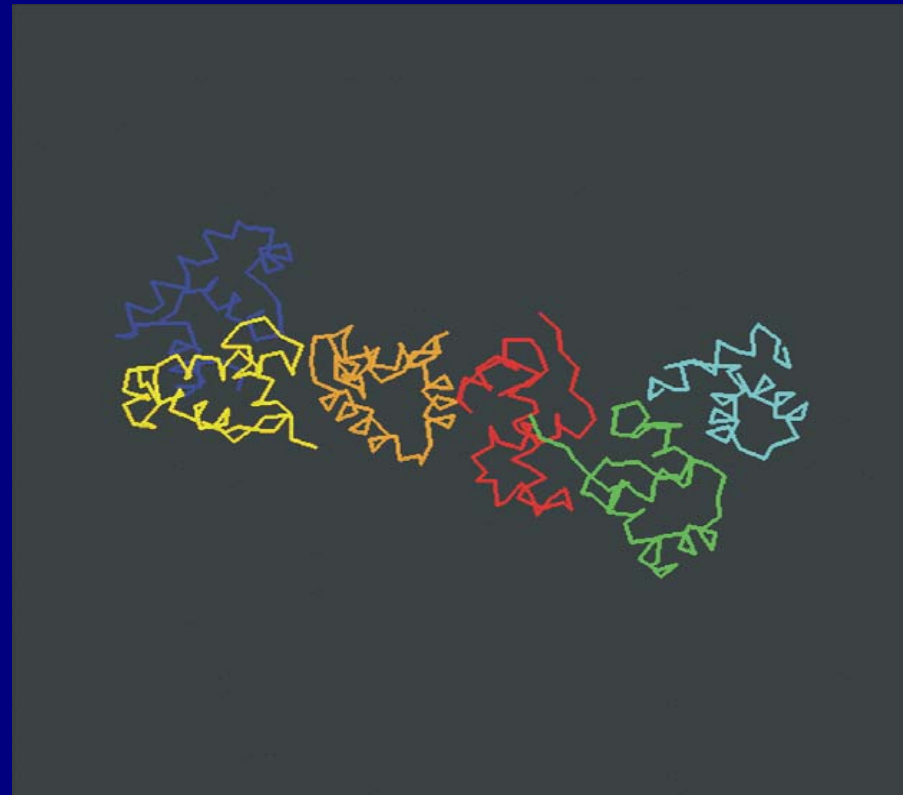
TLS	R	R_{free}
0	22.9	29.5
1	21.4	25.9
4	21.1	25.8

Contributions to equivalent isotropic B_s



Example GerE

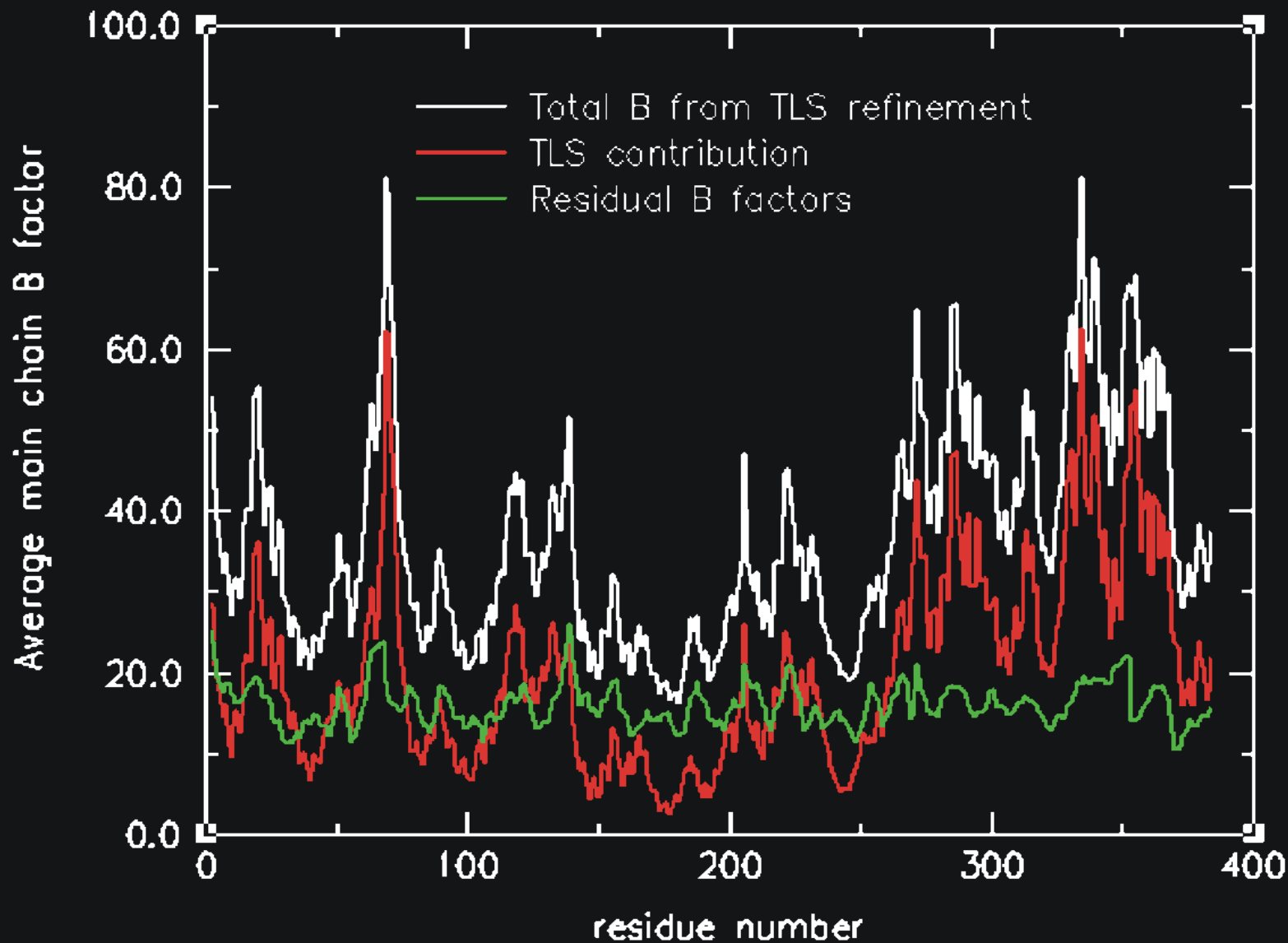
- Transcription regulator from *B. subtilis* [Ducros, V.M. et al., (2001) Crystal structure of GerE, the ultimate transcriptional regulator of spore formation in Bacillus subtilis, *J. Mol. Biol.*, 306, 759-771]
- 74 amino acids
- 6 chains A-F in asu
- Data to 2.05Å



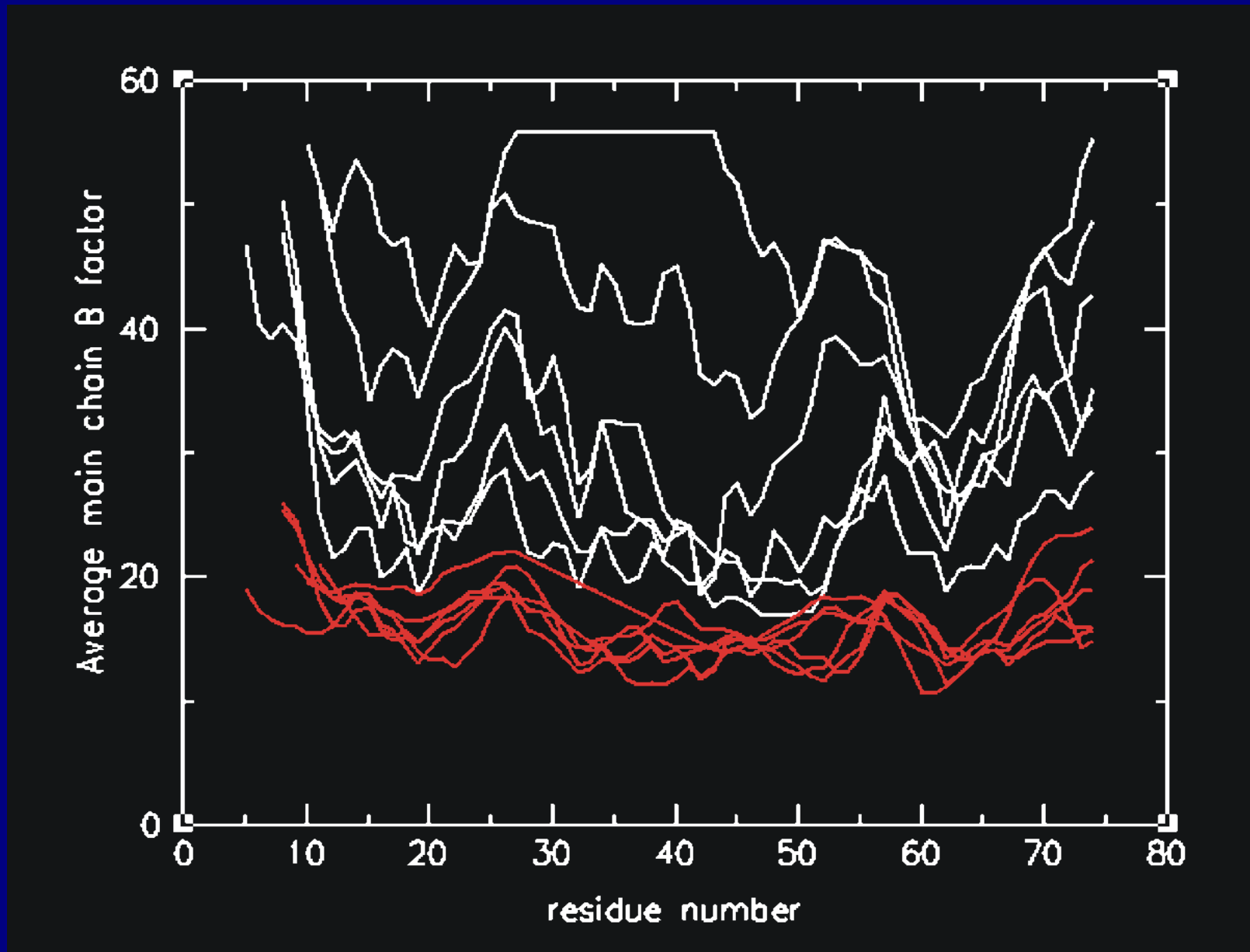
Refinement GerE

Model	TLS	NCS	R	R_{free}	CC_B
1	0	No	21.9	29.3	0.519
2	0	Yes	22.5	30.0	0.553
3	6	Yes	21.4	27.2	0.816

Contribution to equivalent isotropic B_s



Bs from NCS related chains



Summary TLS

TLS parameterization allows to partly take into account anisotropic motions at modest resolution ($> 3.5 \text{ \AA}$)

TLS refinement might improve refinement statistics of several percent

TLS refinement in *REFMAC5* is fast and therefore can be used routinely

TLS parameters can be analyzed to extract physical significance

[Howlin, B. & al. (1993) TLSANL: TLS parameter-analysis program for segmented anisotropic refinement of macromolecular structures, J. Appl. Cryst. 26, 622-624]

[Painter, J. & Merritt, E.A. (2005) A molecular viewer for the analysis of TLS rigid-body motion in macromolecules, Acta Cryst. D61, 465-471]

Refmac People

Garib Murshudov

Alexei Vagin

Andey Lebedev

Fei Long

Dan Zhou

James Foadi

Roberto Steiner

Financial support

