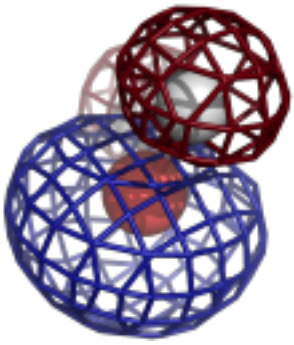
VOLUME TWO



COMPUTATIONAL CRYSTALLOGRAPHY NEWSLETTER

JANUARY MMXI

LABELIT SYMMETRY KING ROSETTA TLS TWINNING

Table of Contents

Editor

Nigel W. Moriarty, NWMoriarty@LBL.Gov

Contributors

P. D. Adams, P. V. Afonine, D. Baker, L. J. Bourhis, G. Bunkóczi, V. B. Chen, F. DiMaio, N. Echols, J. J. Headd, R. W. Grosse-Kunstleve, V. Y. Lunin, N. W. Moriarty, R. J. Read, D. C. Richardson, J. S. Richardson, N. K. Sauter, T. C. Terwilliger, A. Urzhumtsev, C. Williams

PHENIX News

New releases

A new tool for automated partitioning a model into TLS groups, `phenix.find_tls_groups`, is now available. This tool is available in the GUI and command-line interfaces and can take advantage of additional available CPU to generate the atom selection for a refinement run. The automatically defined TLS groups can be readily visualised and edited in the GUI. This tool and all others mentioned here are available in *PHENIX* version 1.7.

Visualisation of mutil-criteria kinemage graphics is now available in *PHENIX* and is discussed in the short communications on page 6.

Generation of ensembles for Molecular Replacement (MR) is the goal of new release called `phenix.ensembler`. Another new release integrates MR and Rosetta in *PHENIX*. For more details, see the short communications for `phenix.ensembler` on page 8 and `phenix.mr_rosetta` on page 10.

New features

Reference model restraints

Reference model restraints are used to steer refinement in cases where the working data

The Computational Crystallography Newsletter (CCN) is a regularly distributed electronically via email and the PHENIX website, www.phenix-online.org/newsletter. Feature articles, meeting announcements and reports, information on research or other items of interest to computational crystallographers or crystallographic software users can be submitted to the editor at any time for consideration. Submission of text by email or word-processing files using the CCN templates is requested.

set is low resolution, but there is a known related structure solved at higher resolution. The higher resolution reference model is used to generate a set of dihedral restraints that are applied to each matching dihedral in the working model. Sequence alignment is handled automatically in cases of sequence dissimilarity, including handling for deletions and insertions. Alternatively, selections may be used to hand-specify the desired reference group in a parameter file. To use (also available in GUI) add the following to the input.

```
main.reference_model_restraints=True
reference_model.file=my_reference.pdb
```

To specify reference group(s),

```
refinement.reference_model.reference_group
{
reference = chain A and resseq 130:134
selection = chain B and resseq 120:124
}
```

For a full list of reference model options, please see the `phenix.refine` documentation.

Augmented base-pairing restraints for RNA

RNA base-pairing restraints have been augmented to include many non-Watson-Crick pairings, including all 28 Saenger types. Saenger nomenclature is now used by default to specify applicable base-pairs. Base-pairs are determined automatically from the input model, but may also be specified by hand in a parameter file. To specify a base-pair by type in a parameter file:

```
nucleic_acids {
  base_pair {
    base1 = "chain \"A\" and resseq 54"
    base2 = "chain \"A\" and resseq 72"
    saenger_class = "XIX"
  }
}
```

For a full list of secondary structure options, please see the on-line documentation for `phenix.refine`.

Crystallographic meetings and workshops

20th West Coast Protein Crystallography Workshop, 20-23 March, 2011

The bi-annual WCPCW is being held at a new location, Monterey Plaza Hotel in Monterey, CA on the 20th to the 23rd of March. *PHENIX* developers will be in attendance.

RapiData, 3-8 April, 2011

RapiData, the annual data collection and structure-solving course is being held from the 3rd to the 8th of April. *PHENIX* developers will be in attendance.

International Conference on Structural Genomics May 10-14, 2011

The International Structural Genomics Organisation (ISGO) is holding a conference in Toronto, Canada from the 10th to the 14th of May. The "*PHENIX* Crystallography Software Workshop" is an all day event on the first day.

American Crystallographic Association, 28 May – 2 June, 2011

A workshop, entitled "Introduction to *PHENIX* for beginning to advanced crystallographers" is planned for the 2011 Meeting of the American Crystallographic Association in New Orleans, Louisiana. The workshop is being held on the 28th May with further information available at www.amercrystalassn.org.

XXII Congress and General Assembly of the International Union of Crystallography (IUCr), 22-30 August 2011

The 22nd Congress and General Assembly of the IUCr will be held in Madrid, Spain on the 22nd to the 30th of August.

PHENIX User's Workshop, 17 March, 2011

A *PHENIX* user's workshop is being planned in Berkeley, California on the 17th of March for local area students, postdocs and other interested parties. Please contact Nat Echols at NEchols@lbl.gov for further information.

Expert advice

Fitting Tips

Vincent Chen, Christopher Williams and Jane Richardson, Duke University

Now that model building is highly automated, for example in *PHENIX*, crystallographers get much less experience with the fun of fitting good maps and perhaps could use tips about what to look for in the difficult cases where the automated methods don't suffice requiring that they do it themselves. In this series, we will try to pass along our own hard-earned rules of thumb. Most of these are what we call "systematic errors", because they happen repeatedly in a similar pattern caused by a misleading appearance in the electron density or a misleading assumption about what should be true in the model. That means they occur fairly often in certain circumstances, but fortunately there is usually a recipe for what needs to be done to fix them.

Our initial fitting tip is one especially relevant at low resolution (around 3Å and worse), first noticed recently when helping to rebuild ribosome structures (in collaboration with Jamie Cate). We encountered places where an extended β strand, instead of the usual alternation of peptide direction, showed three carbonyls in a row that all pointed the same general direction. Figure 1 shows an example of a three-stranded antiparallel β sheet with two such problems (left panel, with the CO triples labeled in red, the all-atom clashes as pink spikes and Ramachandran outliers as green lines). This prevents good H-bonding, produces clashes, often places a sidechain on the wrong side of the sheet and, usually, leads to Ramachandran outliers.

The misleading feature that causes this mistake is that in the 2.5 to 3Å resolution range the carbonyls start to lose definition in the map and a strand becomes a rather smooth, round tube. Therefore the electron density no longer provides the clues to peptide orientation that both people and software rely on at better resolution.

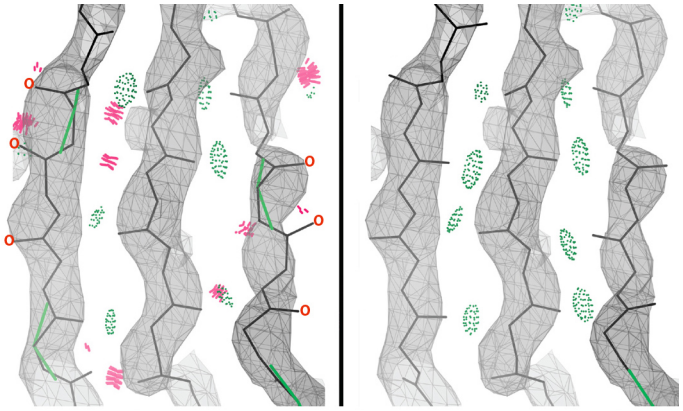


Figure 1: Model on left has a misfit β strand containing three consecutive parallel carbonyls and the corrected model on the right.

The fixing procedure begins with a near-180° flip of the peptide containing the central of the three COs, followed by geometry regularization, refitting of the affected sidechain and optimization of backbone β-sheet H-bonding. The result, as in the right panel of figure 1, should correct the clashes, Ramachandran outliers, improve H-bonding (pillows of green dots) and fit the density at least a bit better.

An alpha-test version of an automated diagnostic for these "three CO" misfit cases identified hundreds of examples in low-resolution PDB structures. See and try out the illustrated case in PDB file `3I1N` (or `3i1n` in lower case for clarity) chain G (L6), residues 95-113. However, note that in future structures for the E coli 70S ribosome many such issues will have been corrected.

FAQ

There are two methods to add links between atoms in *PHENIX*. Which one should I use?

Both methods have their place. Either can be used in most situations, however, one is designed for adding one or two links with little setup while the other is more extensible to larger numbers of links and also reusable in other protein models.

```
refinement.geometry_restraints.edits {
  zn_selection = chain X and resname ZN and resid 200 and name ZN
  his117_selection = chain X and resname HIS and resid 117 and name NE2
  asp130_selection = chain X and resname ASP and resid 130 and name OD1
  bond {
     action = *add
     atom_selection_1 = $zn_selection
     atom_selection_2 = $his117_selection
     distance_ideal = 2.1
     sigma = 0.02
     slack = None
  }
  bond {
     action = *add
     atom_selection_1 = $zn_selection
     atom_selection_2 = $asp130_selection
     distance_ideal = 2.1
     sigma = 0.02
     slack = None
  }
  angle {
     action = *add
     atom_selection_1 = $his117_selection
     atom_selection_2 = $zn_selection
     atom_selection_3 = $asp130_selection
     angle_ideal = 109.47
     sigma = 5
  }
}
```

Figure 2: Example of the "edits" syntax for linking atoms with bonds and angles in phenix.refine.

The simplest method is the "edits" scheme specific to *PHENIX*. The format uses the phil syntax (cctbx.sf.net/libtbx_phil.html) that drives the command interface for *PHENIX*. The basic concept involved is selecting atoms and performing an action. The selection syntax is the same as the selection syntax in other portions of *PHENIX* and is shown in figure 2.

The first lines create selections of three atoms that can be used in the subsequent actions. Note that the selections must parse to a single atom but can be either all the alternative locations of an atom or a specific alternative location. The use of variables such $zn_selection reduces clutter and errors in the bond and angle scopes but is not required. The spelt-out selection can be used inside the bond or angle entries but care should be taken with multiple entries containing the same atoms.

The most common action is adding a geometric entity and is performed using the *add syntax. The ideal value of the geometry entities and the sigma value are needed by the refinement. Additional values include a

```
data_link_NGA-THR
#
loop_
_chem_link_bond.link_id
_chem_link_bond.atom_1_comp_id
_chem_link_bond.atom_id_1
_chem_link_bond.atom_2_comp_id
_chem_link_bond.atom_id_2
_chem_link_bond.type
_chem_link_bond.value_dist
_chem_link_bond.value_dist_esd
 NGA-THR  1 C1  2 OG1   single   1.439  0.020
loop_
_chem_link_angle.link_id
_chem_link_angle.atom_1_comp_id
_chem_link_angle.atom_id_1
_chem_link_angle.atom_2_comp_id
_chem_link_angle.atom_id_2
_chem_link_angle.atom_3_comp_id
_chem_link_angle.atom_id_3
_chem_link_angle.value_angle
_chem_link_angle.value_angle_esd
 NGA-THR  1 C1  2 OG1 2 CB    108.700  3.000
 NGA-THR  1 O5  1 C1  2 OG1   112.300  3.000
#
```

Figure 3: Example of the cif_link syntax for linking atoms with bonds and angles in phenix.refine.

slack value that enables a flat-bottomed potential well and a symmetry operator for linking symmetry-related atoms. The list of geometric restraints that can be added in this fashion currently includes bonds, angles, dihedrals and planes.

The second technique is the standard link mechanism from the Monomer Library (www.ccp4.ac.uk/html/mon_lib.html). Two files are required to perform the same function as edits but this approach is more efficient for defining many links.

The data_link file defines the geometric restraints involved in the linking between two "residues". In the example in figure 3, the saccharide specified by the code NGA is specified as bonding to the oxygen of a threonine (THR). The first line of the file specifies the id of the link. Subsequent lines specify the bond between the C1 of the NGA and the OG1 of THR along with an ideal value and estimated standard deviation (ESD) or

sigma value. Two angles are also specified below the bonds.

The second file is used to specify the residues that are to be linked. Figure 4 shows the syntax required by `phenix.refine` to apply the link, `NGA-THR`, to the two selected residues. Note that the ordering of the selection is important.

```
refinement.pdb_interpretation.apply_cif_link {
  data_link = NGA-THR
  residue_selection_1 = chain X and resname NGA and resid 900
  residue_selection_2 = chain X and resname THR and resid 42
}
```

Figure 4: Example of the `cif_link` syntax for linking atoms with bonds and angles in `phenix.refine`.

Using the latter method a library of the links could be developed and then simply applied to the desired residues in a refinement. This is a powerful feature and the Monomer Library contains a number of useful pre-defined links such as saccharide linking to protein and glycosidic bonds that can be accessed using the "`apply_cif_link`" demonstrated in figure 4.

Currently, different modules in *PHENIX* use one or the other technique to communicate information. Metal coordination uses the edits to addition restraints to `phenix.refine`. *eLBOW* uses the `cif_link` method to link ligands to a protein residue.

Multi-criterion kinemage graphics in PHENIX

Jeffrey J. Headd[a], Vincent B. Chen[c], Nathaniel Echols[a], Nigel W. Moriarty[a], David C. Richardson[c], Jane S. Richardson[c] and Paul D. Adams[a,b]

[a]*Lawrence Berkeley National Laboratory, Berkeley, CA 94720*
[b]*Department of Bioengineering, University of California at Berkeley, Berkeley, CA 94720*
[c]*Department of Biochemistry, Duke University Medical Center, Durham, NC 27710*

Correspondence email: JJHeadd@LBL.Gov

An important component of crystallographic structure determination is validation of model quality. The MolProbity (Chen *et al.*, 2010) web server provides a variety of metrics to evaluate model quality and presents the analysis in both tabular and graphical forms. The KiNG structure viewer (Chen *et al.*, 2009) used by MolProbity displays structural analysis in the visual form of a multi-criterion kinemage, allowing the user to quickly identify both interesting *structural* features and model-building errors with criterion-specific graphical cues. *PHENIX* has previously incorporated MolProbity validation methods, such as rotamer outliers and steric clashes, in tabular form (Adams *et al.*, 2010), but has not provided the same level of graphical validation available from MolProbity. To complement tabular analysis, *PHENIX* now features multi-criterion kinemage graphics. Multi-criterion kinemages generated in *PHENIX* contain the same visual validation graphics as presented by MolProbity, which includes graphics for rotamer and Ramachandran outliers, steric clashes, angle and bond-length deviations, Cβ deviations and ribose pucker outliers for nucleic acids. Figure 1 depicts an example multi-criterion kinemage displayed in KiNG for ubiquitin (pdbID: 1UBQ). See figure 2 for a complete depiction of all available graphical validation metrics.

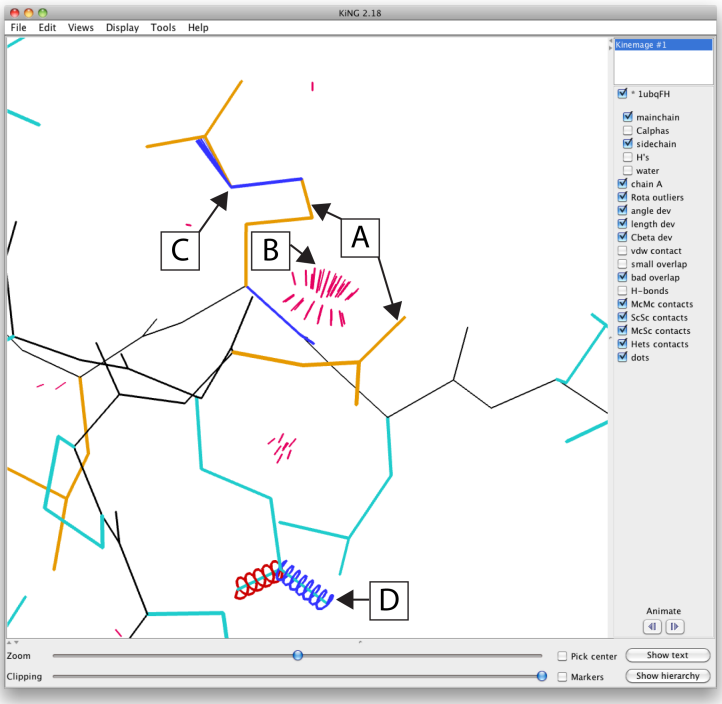


Figure 1: Close-up within a sample multi-criterion kinemage for ubiquitin (pdbID: 1UBQ). (A) Sidechain rotamer outliers are shown in gold as seen here for Arg A 72 and Asp A 39. (B) Significant steric overlaps (> 0.4Å) are shown by hot pink spikes. (C) Angle deviations greater that 4σ are flagged, blue for angles that are too small and red for angles that are too large. (D) Bond-length deviations greater than 4σ are flagged, blue for bonds that are too short and red for bonds too long.

As discussed in the July 2009 issue of the Computational Crystallography Newsletter, *PHENIX* now includes KiNG as a core component, so displaying kinemage graphics is natively available. When using the `phenix.refine` GUI, a multi-criterion kinemage is generated by default. At the completion of a refinement run, the *MolProbity* → *Summary* tab will have a button labeled "show validation in KiNG", as seen in figure 3. Clicking this button will launch KiNG and load the multi-criterion kinemage for interactive use. The .kin file is also available in the refinement project directory.

Multi-criterion kinemages are also available via the command line tools. To generate a kinemage on the command line, run:

```
phenix.kinemage 1ubq.pdb
```

This command will automatically add hydrogens if not present (crucial for correct contact analysis) and generate a multi-criterion kinemage in a file named `1ubq.kin`. To specify the output file name and/or location, run:

```
phenix.kinemage 1ubq.pdb out_file=/path/to/1ubq_multi.kin
```

Once generated, the multi-criterion kinemage may be viewed in KiNG by running:

```
phenix.king 1ubq_multi.kin
```

Serendipitously, restraints generated for novel ligands, modified amino acids and nucleic acids and other non-standard molecules with *eLBOW* (Moriarty *et al.*, 2009) can be used to augment structure validation. The kinemage generation methods in *PHENIX* use these definitions for both graphical rendering and model validation, providing more facile functionality over the current state-of-the-art in MolProbity for novel molecular handling. Kinemage generation in the `phenix.refine` GUI handles custom restraints incorporation automatically. For command line generation, a restraints file (`.cif`) may be included by running:

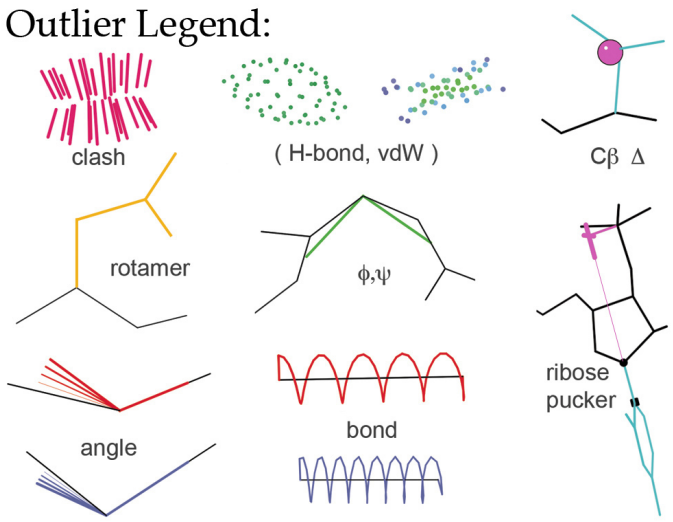


Figure 2: Legend of all outlier symbols used in multi-criterion kinemages generated by PHENIX. Taken from Chen *et al.*, 2010.

```
phenix.kinemage 1ubq.pdb cif=ligands.cif
```

The flexibility of the kinemage format and many display features of KiNG provide great opportunity for development of new crystallographic-specific validation graphics in *PHENIX*. KiNG is capable of displaying parallel coordinates, for example (Chen *et al.*, 2009), which will be useful for displaying trends across all validation criteria for tracking model improvement throughout the building and refinement process. Development is ongoing and future releases of *PHENIX* will feature expanded validation techniques to aid in structure solution.

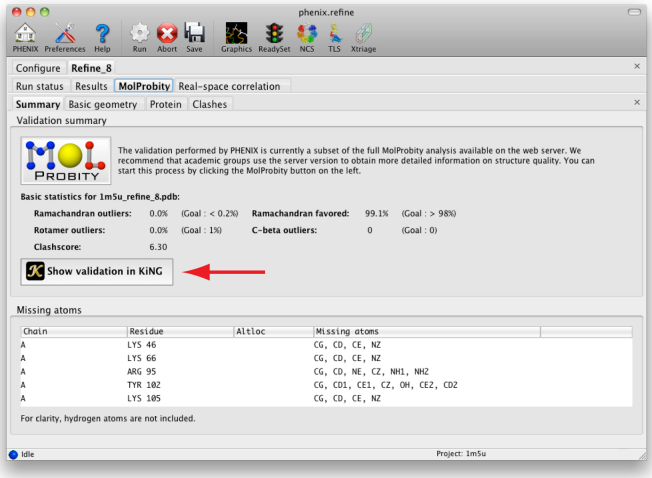


Figure 3: A multi-criterion kinemage is available in the `phenix.refine` GUI by clicking on the "Show validation in KiNG" button in the *MolProbity* tab. A multi-criterion kinemage is also generated in the comprehensive validation GUI in the same manner (not shown).

References

Adams P.D., Afonine P.V., Bunkóczi G., Chen V.B., Davis I.W., Echols N., Headd J.J., Hung L.-W., Kapral G.J., Grosse-Kunstleve R.W., McCoy A.J., Moriarty N.W., Oeffner R., Read R.J., Richardson D.C., Richardson J.S., Terwilliger T.C. and Zwart P.H.. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Cryst. D66:213-221.

Chen V.B., Davis I.W. and Richardson D.C. (2009) KiNG (Kinemage, Next Generation): A versatile interactive molecular and scientific visualization program. Protein Science 18:2403-2409.

Chen V.B., Arendall III, W.B., Headd J.J., Keedy D.A., Immormino R.M., Kapral G.J., Murray L.W., Richardson J.S. and Richardson D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Cryst. D66:12-21.

Moriarty N.W., Grosse-Kunstleve R.W. and Adams P.D. (2009) electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. Acta Cryst. D65:1074-1080.

phenix.ensembler: a tool for multiple superposition

Gábor Bunkóczi and Randy J. Read

Department of Haematology, University of Cambridge, CIMR, Wellcome Trust/MRC Building, Hills Road, CB2 0XY, Cambridge, UK

Correspondence email: gb360@cam.ac.uk

In molecular replacement, a collection of superposed simple models ("ensemble" models) can give superior signal over any of the individual components. For example, hen egg white lysozyme can be solved with either mouse digestive lysozyme (2FBD, 40% identical) or apo bovine α-lactalbumin (1F6R, 40% identical) with both of them yield similar quality solutions (Table 1). However, when the two models are combined, solution quality (measured both in terms of translation function Z-score or final log-likelihood gain) increases (Table 1).

Table 1. Comparison of model quality between a two-member ensemble and its components.

Model	TFZ	Refined LLG
mouse digestive lysozyme	9.7	91.66
apo bovine α-lactalbumin	11.4	95.88
ensemble	12.6	146.49

In an ideal ensemble, individual components capture possible conformations for the flexible parts of the structure, while emphasizing the rigid core. Therefore, to create an ensemble, models have to be selected based on their conformational variability, while their rigid core has to be identified and superposed.

`phenix.ensembler` was written to automate the creation of models as much as possible as well as offer some help in deciding what models to include. It takes a series of PDB files as arguments (potentially unedited models from the PDB as identified by a homology search) and optional alignment files.

1. The program reads all PDB files and analyses all chains. All non-protein chains are discarded; multiple copies of the same chain are retained.
2. Equivalent residue positions in protein chains are aligned using one of the following methods (residue alignment):
   a. `ssm`: uses the secondary-structure matching (Krissinel & Henrick, 2004) algorithm (default).
   b. `muscle`: automatically create a multiple alignment using `phenix.muscle` (Edgar, 2004). Results are possibly less accurate than those from `ssm`, but are applicable for any protein chains, even those without any secondary structure.
   c. `alignments`: reads alignments provided on the command line, thereby giving full control over residue alignment.
   d. `resid`: aligns residues with identical residue number and insertion code.
3. Equivalent atoms in residues are aligned based on atom name (atom alignment) and atoms participating in the superposition are selected (default: C_α only).
4. Equivalent positions are then superposed using a multiple superposition algorithm. This gives better results than a series of pair-wise superpositions if there are significant differences between the structures. There are two algorithms implemented:
   a. `gapless` (Diamond, 1992). This is a fast algorithm, but it can only use sites that are present in all structures superposed and therefore may discard a significant number of sites when superposing several protein chains with distant homology (default).

b. `gapped` (Wang & Snoeyink, 2008). This algorithm can take alignment gaps into account in superposition and therefore can make use of more sites, giving more precise results.

5. Superposition is then alternated with weighting until convergence. Currently, there are two weighting schemes implemented:
   a. `unit`. This gives an unweighted superposition.
   b. `robust_resistant`. This assigns a weight to each superposed position based on the root mean-square deviation between all structures according to a robust-resistant weight function (default). Tolerance of the weighting can be controlled by the weighting.robust_resistant.critical parameter, with lower values down-weighting deviating positions more progressively. This approach proved very efficient in correcting for incorrect site alignment and identifying identical regions (Figure 1).
6. Superposed structures are analysed and clusters are determined based on structural similarity (controlled by the `configuration.clustering` parameter). Resulting clusters can be used to classify protein chains according to conformational variability. Depending on the number of protein chains, some experimentation may be necessary with the clustering distance parameter, so that an optimum is found between the two extremes (each structure forms a separate cluster *vs.* all structure belong to the same cluster). A representative from each cluster may then be chosen for inclusion in the ensemble.
7. After optionally sorting according to sequence identity, chain length, weighted and unweighted r.m.s.d., protein chains are transformed and written out.

The resulting ensemble model can be used directly in molecular replacement.

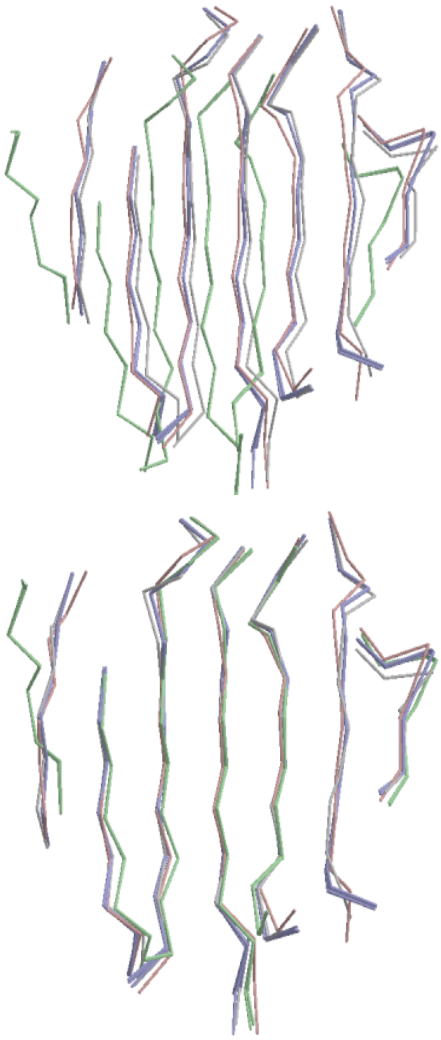


Figure 1. Correcting residue misalignment with weighting in superposition. Upper is the unweighted superposition using an alignment from *CLUSTALW* (Larkin et. al., 2007). The lower is the weighted superposition with the same alignment, resulting in much better structural agreement.

References

Diamond, R. (1992). On the multiple simultaneous superposition of molecular structures by rigid body transformations. *Protein Sci.* **1**, 1279-1287.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792-1797.

Krissinel, E. & Henrick, K. (2004). Secondary-structure matching, a new tool for fast protein structure alignment in three dimensions. *Acta Cryst.* **D60**, 2256-2268.

Larkin, M. A, Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948.

Wang, X. & Snoeyink J. (2008). Defining and computing optimum RMSD for gapped and weighted multiple-structure alignment. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **5**, 525-533.

phenix.mr_rosetta: A new tool for difficult molecular replacement problems

Tom Terwilliger[a], Randy Read[b], Frank DiMaio[c] and David Baker[c]
[a]Los Alamos National Laboratory, Los Alamos NM 87545
[b]University of Cambridge, Department of Haematology , Cambridge, CB2 0XY, UK
[c]University of Washington, Department of Biochemistry, Seattle, WA, 98195, USA

Correspondence email: terwilliger@lanl.gov

What is phenix.mr_rosetta?

The *PHENIX* development team is working with the Baker laboratory at the University of Washington to combine the power of Rosetta structure modeling with *PHENIX* automated molecular replacement (MR), model-building, density modification and refinement. The basic idea is to find MR solutions with `phenix.automr`, rebuild them with Rosetta, including electron density map information, then rebuild those models with `phenix.autobuild`. The combination of Rosetta rebuilding and phenix rebuilding is the key part of this method. MR solutions are found with `phenix.automr` (Phaser), scored with LLG (optionally following Rosetta relaxation), the best solutions are picked and rebuilt with Rosetta including map information, the resulting models are scored with Rosetta, rescored with LLG, and the top models are rebuilt with `phenix.autobuild`.

What is phenix.mr_rosetta good for?

`phenix.mr_rosetta` can be very useful for cases where the search model used in molecular replacement is slightly too distant to rebuild successfully with `phenix.autobuild`. It can also be useful in cases where the model is too distant to even find a molecular replacement solution and pre-refinement with Rosetta can yield an improved search model.

How do I run mr_rosetta?

You can run `phenix.mr_rosetta` in a very automated way, or as a tool to find molecular replacement solutions and to systematically improve them. To run `phenix.mr_rosetta` you need to have both *PHENIX* (any recent version) and Rosetta (development version as of this writing, or version 3.2 or later once available), installed.

The basic inputs for `phenix.mr_rosetta` are pretty simple: (1) a data file with F, SIGF and freeR flags, (2a) a search model and an alignment file, or (2b) an hhpred file with a list of alignments and PDB file names and (3) a pair of fragments files that you create and download from the Robetta server.

You can get the `hhpred` file with alignments by pasting your sequence into the server at toolkit.tuebingen.mpg.de/hhpred. This takes about 10 minutes. You can get the fragments files by pasting your sequence into the Robetta server at robetta.bakerlab.org/fragmentsubmit.jsp. This takes a few hours to run, depending on the length of your sequence.

Once you have these files, you simply edit a simple script file for `phenix.mr_rosetta` that specifies these files (and other parameters if you wish). A typical command-line run of `phenix.mr_rosetta` is shown at the right.

```
phenix.mr_rosetta \
   seq_file=seq.dat \
   data=coords1.mtz \
   search_models=coords1.pdb \
   already_placed=True \
   fragment_files = test3.gz \
   fragment_files = test9.gz \
   rosetta_models=20 \
   ncs_copies=2 \
   space_group=p212121  \
   use_all_plausible_sg=False \
   nproc=200 \
   group_run_command=qsub
```

Does phenix.mr_rosetta require a cluster to run?

`phenix.mr_rosetta` does require building a number of Rosetta models with each model taking from 10-60 minutes to build with a single processor. In many cases, `phenix.mr_rosetta` can succeed with

as few as 20 models in each cycle. This means that a computer with 4 processors can be quite sufficient to run `phenix.mr_rosetta` and can finish in a day or so. In very challenging cases, as many as 2000 models may need to be built (the best models are picked and the density from the top 20% of models is averaged) making a cluster the best option. `phenix.mr_rosetta` can run on a Sun Grid Engine (SGE) cluster and on a Condor cluster. It may also run on other clusters. To run on a cluster you simply specify the command that you use to submit jobs ("`qsub`" for a SGE cluster for example).

Advanced uses of `phenix.mr_rosetta`

Once you have used `phenix.mr_rosetta` a few times, you will find that you can control where it starts and what it does in quite some detail. You can choose a particular solution that it is working on and have it build Rosetta models for that solution, then write out a table of results that you can examine. This way you can combine your intuition with the scoring that `phenix.mr_rosetta` uses to optimize your search.

You can also pre-refine your search model. This just means running Rosetta modeling on your search model without including any information from the crystallographic experiment. This can be very useful because Rosetta modeling can improve your search model and allow molecular replacement to succeed in cases where it might otherwise fail completely.

Where can I read more?

You can see all about running `phenix.mr_rosetta` in the *PHENIX* documentation.

Fuzzy space group symbols: H3 and H32

Ralf W. Grosse-Kunstleve[a], Nathaniel Echols[a] and Paul D. Adams[a,b]
[a]*Lawrence Berkeley National Laboratory, Berkeley, CA 94720*
[b]*Department of Bioengineering, University of California at Berkeley, Berkeley, CA 94720*

Correspondence email: RWGrosse-Kunstleve@LBL.Gov

Introduction

Most applications in *PHENIX* (Adams et al., 2010) have to process space group symbols and most use the *sgtbx* space group toolbox for this purpose (Grosse-Kunstleve et al., 2002). The *sgtbx* space-group symbol interpreter accepts a large variety of inputs, for example "19", "P212121", or "P2(1)2(1)2(1)". Unfortunately it is not straightforward to accommodate the "H3" and "H32" symbols commonly used in macromolecular crystallography for the rhombohedral space groups *R*3 (No. 146) and *R*32 (No. 155), respectively. Currently these symbols appear in about 2.5% (1764 out of 69655) files in the PDB (www.pdb.org). This article explains the difficulties and how *PHENIX* handles rhombohedral space group symbols.

Conflicting definitions for H3 and H32

The main reference work for information about space groups and space-group symbols is *International Tables for Crystallography Volume A* (ITA) (Hahn, 1983-2006). Table 1.2 of ITA defines "symbols for the conventional centring types", which are *P*, *C*, *A*, *B*, *I*, *F*, *R* and *H*. The definition of the *H* symbol is not commonly used or known; the fractional coordinates of the lattice points within the unit cell are defined as 0,0,0; $\frac{2}{3},\frac{1}{3},0$; $\frac{1}{3},\frac{2}{3},0$. For comparison, the lattice points associated with the widely used *R* symbol are 0,0,0; $\frac{2}{3},\frac{1}{3},\frac{2}{3}$; $\frac{1}{3},\frac{2}{3},\frac{1}{3}$. The definition of the *H* symbol dates back to the precursor of ITA, *Internationale Tabellen zur Bestimmung von Kristallstrukturen* (Hermann, 1935) in which it is used for the description of 18 trigonal and 27 hexagonal space groups in "triple cell" settings (cci.lbl.gov/cctbx/multiple_cell.html). These settings were replaced by primitive or *R*-centered settings in *International Tables for Crystallography Volume I* (Henry & Lonsdale, 1952). In modern editions of the International Tables, the triple cell settings only appear in a column of ITA Table 4.3.1. According to this table, the symbol "H3" designates the triple-cell setting of space group 143, conventionally known as *P*3; the symbol "H32" (formatted $H3_2$) designates the triple-cell setting of space group 145, conventionally known as $P3_2$.

In general the macromolecular field uses the standard Hermann-Mauguin space-group symbols defined by ITA, but the PDB introduced a conflicting de-facto standard for "H3" and "H32". The root of the conflict is probably to be sought in the well-known ambiguities of Hermann-Mauguin symbols. A Hermann-Mauguin symbol uniquely identifies the space group type (one of the 230 crystallographic space group types), but not the setting. In the case of the seven rhombohedral space groups, the same Hermann-Mauguin symbol is used for a setting with a *R*-centered hexagonal basis system (e.g. "*R*3 (hexagonal axes)" in ITA) and a primitive setting with a rhombohedral basis system (e.g. "*R*3 (rhombohedral axes)"). Figure 1 shows how the basis systems are related. The ITA notation for the information about the setting is long compared to the Hermann-Mauguin symbol. This has lead many authors of crystallographic software and reference tables to introduce a more

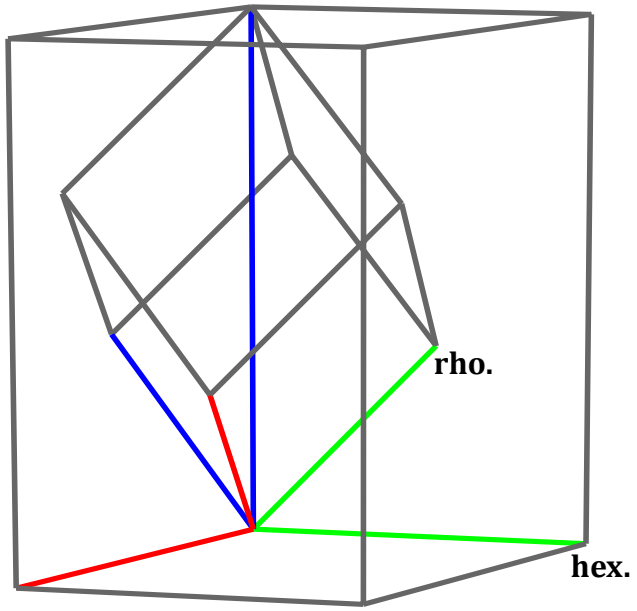


Figure 1: Basis systems of *R*3 settings with "**hex**agonal axes" and "**rho**mbohedral axes" (ITA nomenclature). The basis vectors are colored **a**=red, **b**=green, **c**=blue. For more information refer to ITA chapter 5.

compact notation. For example, the symbols "R3:h" and "R3:r" appear in the International Tables for Crystallography Volume B (Shmueli, 2001) Table A1.4.2.7 and the IUCr symCIF dictionary (Brown, 2005). The PDB has gone one step further by re-interpreting the first character of the "R3" and "R32" Hermann-Mauguin symbols as information about the setting. This confusion of centering type (ITA Table 1.2) and setting information is compounded by the conflict with the *H* centering type symbol.

The widely used *SCALEPACK* software (www.hkl-xray.com) has adopted a similar approach as the PDB, but in exactly the opposite way. According to the *SCALEPACK* manual, the symbols "R3" and "R32" correspond to the PDB symbols "H3" and "H32", respectively, and vice versa. Table 1 summarizes the symbols used for space groups 146 and 155 in different contexts.

Table 1: Summary of symbols used for space groups 146 and 155 in different contexts.

Space-group No.	Int. Tab. Vol. A	Int. Tab. Vol. B	PDB	SCALEPACK
146	R3 (hexagonal axes)	R3:h	H3	R3
	R3 (rhombohedral axes)	R3:r	R3	H3
155	R32 (hexagonal axes)	R32:h	H32	R32
	R32 (rhombohedral axes)	R32:r	R32	H32

Handling of H3 and H32 in PHENIX

An obvious way to disambiguate the rhombohedral space group symbols in Table 1 is to take the unit cell parameters into account. *PHENIX* makes use of this approach in some contexts, but it is not always practical. In many situations space group symbols and unit cell parameters are processed independently and combined only in advanced stages of involved procedures. A simple example is the validation of space group symbols in the *PHENIX* Graphical User Interface (GUI). The symbols are validated and standardized the moment the user presses the enter key or moves the input focus. It would be far more complicated to defer the validation until unit cell parameters are available. Potentially the GUI manages multiple unit cells and space group symbols and it may therefore not even be straightforward to formalize the relationships.

Until recently, the *PHENIX* GUI did not recognize the "H3" and "H32" symbols. Starting with *PHENIX* installer *dev-603*, the *sgtbx* library used by the GUI supports these symbols, following the de-facto PDB standard. We had been hesitant to take this step because it is the only non-conformance of the *sgtbx* library with the ITA standards. Eventually a considerable stream of negative feedback convinced us to value practicality higher than the principle of full ITA compliance. Since the new support for the PDB symbols is implemented in the *sgtbx* library, all applications using this library (`phenix.refine`, `phenix.xtriage`, `phenix.phaser`, to name just a few) also support these symbols now.

In the contexts of reading and writing PDB and CCP4 MTZ files, *PHENIX* included support for the PDB symbols for many years already. The interpretation of PDB `CRYST1` records is implemented in the `iotbx.pdb` module (Grosse-Kunstleve & Adams, 2010). In this context the rhombohedral space group symbols are used essentially only to infer the space group type. The choice of basis system, hexagonal vs. rhombohedral, is determined by inspection of the unit cell parameters. Internally *PHENIX* uses the symbols "R3:H", "R3:R", "R32:H" and "R32:R" (they appear, for example, in the GUI), but when formatting `CRYST1` records for output the PDB symbols are used instead.

Space group information from MTZ files is processed similarly, but in most cases the space group symbol included in the MTZ file is not actually used since the symmetry is usually defined unambiguously via lists of symmetry operations. When generating output MTZ files, *PHENIX* uses a copy of a CCP4 symmetry library file (`lib/data/symop.lib` in CCP4) to obtain the CCP4 space group symbol by matching the symmetry operations in the *sgtbx* space group object with the operations listed in the CCP4 library file. For the rhombohedral space groups, this mechanism produces the PDB symbols.

Starting with *PHENIX* installer *dev-603*, symmetry information read from merged *SCALEPACK* files is also subject to disambiguation via unit cell parameters.

Conclusion

The significant amount of programming effort and user frustration caused by the "H3" and "H32" space-group symbols is a good example of how time can be lost by not adopting long established standards. This is not meant to suggest revising the PDB as this would certainly only increase the confusion. Rather, it is a salient reminder to the entire crystallographic methods developer community to avoid ad-hoc approaches when possible. Ever more automated systems require highly reliable components and unambiguous semantics. The effort spent early on ensuring reliability and clarity is usually rewarded many times over as time passes.

References

Adams, P. D. et al. (2010). Acta Cryst. **D66**, 213-221.

Brown, I. D. (2005). ftp://ftp.iucr.org/cifdics/cif_sym_1.0.1.dic

Grosse-Kunstleve, R. W., Adams, P. D. (2010). Computational Crystallography Newsletter **1**, 4-11.

Grosse-Kunstleve, R. W., Sauter, N. K., Moriarty, N. W., Adams, P. D. (2002). J. Appl. Cryst. **35**, 126-136.

Hahn, T. (1983-2006). International Tables for Crystallography, Vol. A. Dordrecht: Kluwer.

Hermann, C. (1935). Internationale Tabellen zur Bestimmung von Kristallstrukturen. Berlin: Gebrüder Bornträger.

Henry, N. F. M., Lonsdale, K. (1952). International Tables for X-ray Crystallography, Vol. I. Birmingham: Kynoch Press.

Shmueli, U. (2001). International Tables for Crystallography, Vol. B. Dordrecht: Kluwer.

Visualizing the raw diffraction pattern with *LABELIT*

Nicholas K. Sauter[a]

[a]*Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720*

Correspondence email: NKSauter@LBL.Gov

Introduction

This article focuses on creating publication-quality pictures that illustrate diffraction data. While numerous tools are available for the routine conversion of raw data files into common image formats, other plots require specialized markup and can only be produced by custom-written software. A familiar example is the need to label each Bragg spot in the diffraction pattern with its proper Miller index. Also, synchrotron beamlines with very fast detectors such as the Pilatus-6M have emphasized the advantage of collecting data with thin rotation slices, which improve the signal-to-noise ratio but also leave the image sparsely populated with Bragg spots. This raises the need for a visualization tool where several consecutive images are summed together to give a more recognizable lattice. Another mechanism to conveniently examine the lattice is to plot the signal that corresponds to plane sections through reciprocal space. Old-style precession cameras would generate this type of photograph experimentally, but for modern rotation geometry it is necessary to synthesize such images with pixels taken from different shots over the whole data set. Code for all of these applications is available within the *LABELIT* package, which is distributed with *PHENIX* and available for download at www.phenix-online.org.

All examples discussed here are executed through the command line, as *LABELIT* has not yet been incorporated into the *PHENIX* graphical interface. Figures are documented in a special subdirectory within the source code: `labelit/publications/ccn_visualization` and can be reproduced by following the instructions contained therein. General documentation for *LABELIT* is at cci.lbl.gov/labelit.

Prerequisite indexing with *labelit.index*

Since the desired illustrations depend on knowledge of the Miller indices, we first perform an autoindexing step to identify the principle axes and unit cell dimensions of the crystal. It is assumed that the *PHENIX* package (any release subsequent to 1 Jan 2011) is installed and added to the path by sourcing the appropriate setup file (`phenix_env` or `phenix_env.sh`). As currently implemented, the program *MOSFLM* must also be placed on path under the alias `ipmosflm`; this is used during the autoindexing step to obtain a firm estimate of the resolution limits.

Indexing is done in a new current working directory (`cwd`), with the raw data frames placed either in `cwd` or any other directory:

```
cwd> labelit.index <data_path>/file_template_1_###.img 1 90
```

In this example the diffraction pattern is indexed from two 1° rotation images (#1 and #90) spaced widely apart to achieve the highest accuracy. Details are found in the *LABELIT* documentation. Of interest below, any lattice with higher than triclinic symmetry can be described in multiple Bravais settings:

```
LABELIT Indexing results:
Solution  Metric fit  rmsd  #spots  crystal_system  unit_cell
:)    5     0.197 dg  0.100    494  orthorhombic oP  84.6 123.3 174.2 90.0 90.0 90.0
:)    4     0.197 dg  0.105    496    monoclinic mP  84.6 123.4 174.3 90.0 90.0 90.0
:)    3     0.197 dg  0.099    493    monoclinic mP  84.6 174.2 123.3 90.0 90.0 90.0
:)    2     0.052 dg  0.081    496    monoclinic mP 123.3  84.6 174.0 90.0 90.2 90.0
:)    1     0.000 dg  0.082    498     triclinic aP  84.6 123.3 174.0 90.2 90.0 90.0
MOSFLM Integration results:
Solution  SpaceGroup Beam x   y  distance  Resolution Mosaicity RMS
:)    5         P222  94.08  94.11  179.99       2.20    0.050       0.040
      1           P1  94.08  94.09  180.02       2.14    0.050       0.027
```

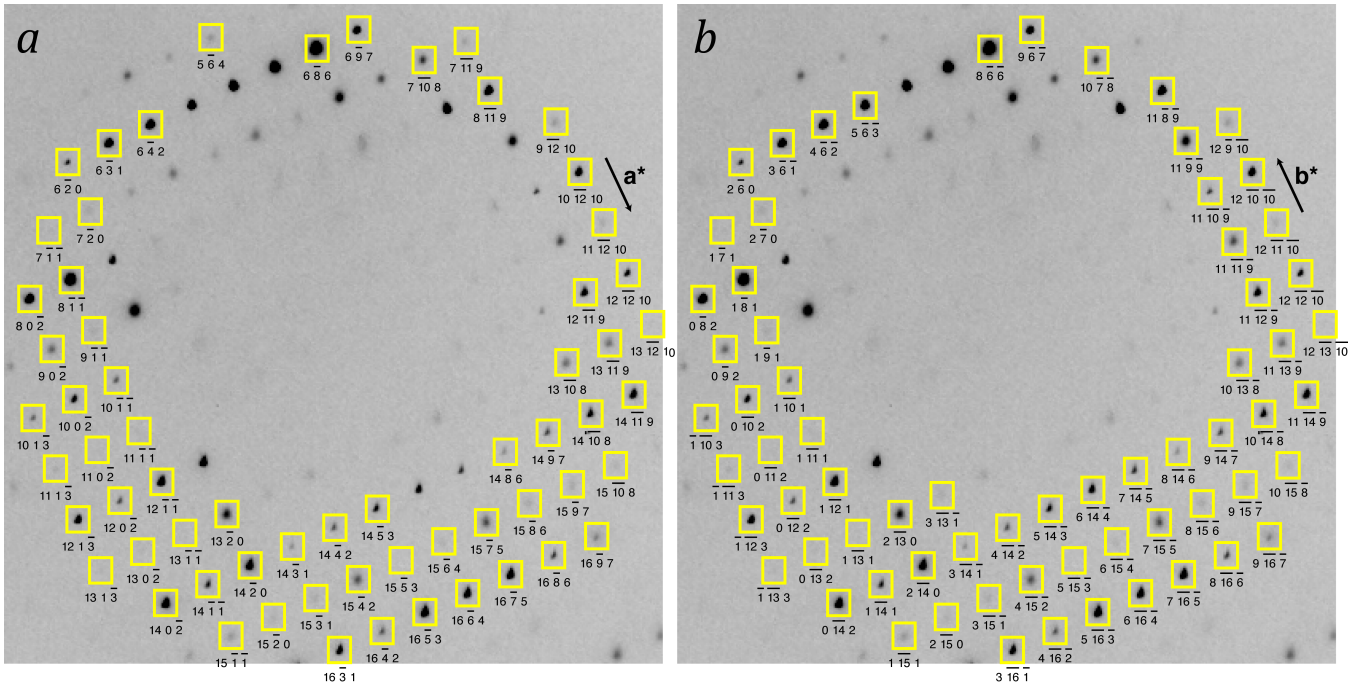


Figure 1. Detail of a rotation image from the 2qyv dataset, illustrated with `labelit.image`. The Bravais lattice is orthorhombic in model (*a*) and monoclinic in model (*b*), corresponding to the above-listed "*LABELIT* solution" numbers 5 and 2, respectively. While describing the same physical data, the two models differ in the orientations of the reciprocal cell axes (**a***, **b***, **c***) and the resultant Miller index labels attached to each Bragg spot. As the two lattice solutions are refined separately, slightly different subsets of Bragg spots are predicted by the two models (yellow boxes), but this distinction is normally erased in the subsequent steps of postrefinement and data integration.

As the printout shows, this orthorhombic lattice may be viewed either in the orthorhombic setting, or alternately in the triclinic or three different monoclinic settings. These are the indexing results from a dataset used by the Joint Center for Structural Genomics (JCSG) to solve the structure of Protein Data Bank entry 2qyv. The JCSG data repository, available for download from www.jcsg.org (Elsliger *et al.*, 2010), is used for the examples in figures 1, 3 & 4.

To create the subsequent illustrations, the indexing results must be kept in their cached location in the files `cwd/DISTL_pickle`, `cwd/LABELIT_pickle` and `cwd/LABELIT_possible`. Therefore, if multiple datasets are to be indexed, separate working directories should be created. Cached information in `cwd` can be deleted with:

```
cwd> labelit.reset
```

PDF-format rendering of diffraction images: *labelit.image*

The first product of interest is a simple picture of a rotation photograph, with Bragg spots labeled as shown in figure 1.

As with other programs in the *PHENIX* family, keyword input for `labelit.image` may be provided either at the command line or from an "effective parameter file" listing the desired keywords in a structured format. Relevant keywords can be listed out with:

```
cwd> labelit.image
cwd> labelit.image help # detailed descriptions for each keyword
```

Output from the undecorated `labelit.image` command may be used as a template to create the

effective parameter file `param.eff`:

```
bravais_choice = 5
image_number = 1
window_fraction = 0.4
window_center_x = 0.5
window_center_y = 0.5
image_brightness = 1.0
pdf_output{
  file = output.pdf
  box_size = 500
}
markup{
  bragg_spot{
    box = True
    linewidth = 0.04
    profile_shrink = 0
    color = yellow
  }
  miller_index{
    legend = False
    font_size = 10
    color = black
    vertical_offset = 10
  }
  inliers = False
}
```

The final image is computed with the command

```
cwd> labelit.image param.eff
```

or by conveniently supplying scoped command-line keywords for all non-default values:

```
cwd> labelit.image bravais_choice=5 \
                   image_number=1 \
                   pdf_output.file=output.pdf \
                   markup.miller_index.legend=True
```

Keywords and default values are as follows:

`bravais_choice=None`

This required keyword identifies the integer Bravais setting number to use for labeling the Bragg spots, as enumerated in the "*LABELIT* solution" column of the *LABELIT* output. The list of possible settings can be viewed again with the command `labelit.stats_index`. Fig. 1 illustrates how the output varies with different Bravais choices; the same physical data are displayed but differently numbered Miller indices are attached to each spot. The correct Bravais choice is not necessarily known at the time of indexing. In the case shown (PBD entry 2qyv) the published symmetry happens to be $P2_12_12$, corresponding to `bravais_choice=5` (figure 1a).

`image_number=None`

The integer image sequence number to use in the illustration (required). It is not necessary to supply the full file name, as the directory and file template are already cached by `labelit.index`.

```
window_fraction=1.0
```

Fractional length of the full image x and y dimensions to be used for illustration. A window_fraction of 0.5 would render a 1500 × 1500 square section of a 3000 × 3000 pixel raw image. To zoom in on an image detail, pick progressively smaller values.

```
window_center_x=0.5
window_center_y=0.5
```

Fractional offset on the full image to be used as the center of the section to be illustrated. The center of the raw image is at the coordinates (`window_center_x = 0.5`, `window_center_y = 0.5`), with x and y being the slow and fast directions on the image, respectively. On the printed page, slow is vertical and fast is horizontal, with the origin in the upper left corner.

```
image_brightness=1.0
```

Factor used to multiply the pixel values to produce a customized brightness. By default, a brightness scale is automatically calculated for each image, such that the 90th-percentile pixel is shown as saturated (black). An image_brightness > 1.0 makes pixels more saturated (darker).

```
pdf_output.file=None
```

Required file name for the output illustration. The top of the printed page will show the image file name and relevant information taken from the file header, while the `labelit.index` results and `labelit.image` command will be summarized at the bottom.

```
pdf_output.box_size=500
```

Number of points (unit of length, 1/72 inch) for the square edge of the illustration on the printed page.

```
markup.bragg_spot.box=True
```

Boolean value to toggle the boxes that locate the predicted position of each Bragg spot.

```
markup.bragg_spot.linewidth=0.04
```

Width of the printed lines used to outline each Bragg spot (in mm). Adjust this value to improve the clarity of the illustration if the Bragg spots are too congested.

```
markup.bragg_spot.profile_shrink=0
```

Number of pixels to shrink the box edge for outlining Bragg spots. By default, the rectangular box is sized to contain the average profile of the bright spots used by `labelit.index` for indexing, plus a two-pixel margin on each side. Use this keyword to improve clarity if necessary.

```
markup.bragg_spot.color=yellow
```

Color used for the Bragg spot boxes, as defined in the PDF-generating package *Reportlab*.

```
markup.miller_index.legend=True
```

Boolean value to toggle the inlining of Miller index HKL values underneath each Bragg spot.

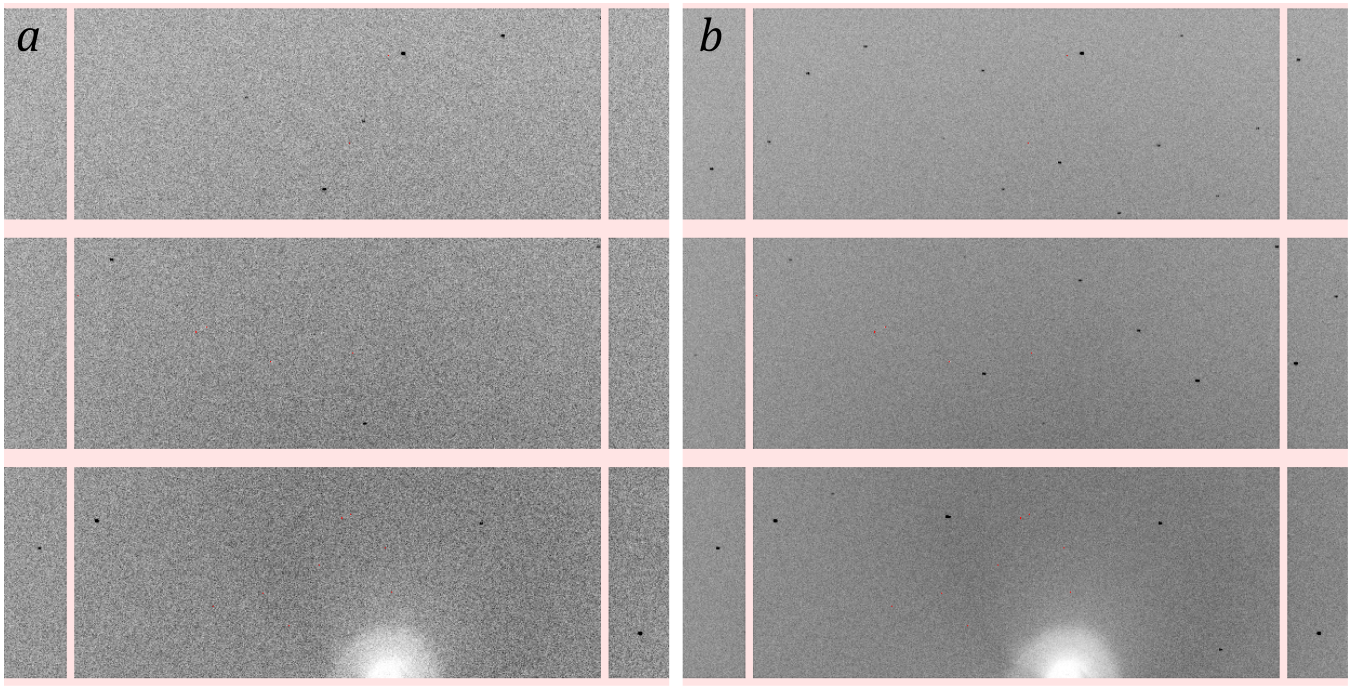


Figure 2. Detail from a cubic-lattice diffraction pattern taken from a single image (*a*, Δφ=0.2°) and a stack of five images (*b*, Δφ=1.0°). The rectangular areas are 195 × 487-pixel modules on a Pilatus-6M detector.

```
markup.miller_index.font_size=10
markup.miller_index.color=black
markup.miller_vertical_offset=10
```

For the inlined HKL values, font size in points, ink color as defined in *Reportlab*, and vertical offset in the downward direction in points, so that the HKL value does not overlap the spot.

```
markup.inliers=False
```

Boolean value to toggle dots locating the bright spots used by `labelit.index` for indexing. Red dots indicate the spot center of mass, pink dots show the maximum pixel.

Summation of consecutive thin-sliced images

Diffraction photographs of still crystals, or those with an extremely small rotation angle, will record a correspondingly thin slice through reciprocal space. Such photographs may exhibit few Bragg spots, especially if the unit cell is small and those that are captured will represent partial slices through the rocking curve, not full intensities. A sparse diffraction pattern from a 0.2° rotation photograph is shown in figure 2a.

For the purpose of illustrating the diffraction, it can be advantageous to stack consecutive rotation shots on top of each other, thus summing the partial intensities and filling out the layer slices so that the lattice pattern is more readily apparent. Such a construction is shown in figure 2b. The picture was created by supplying an `image_range` keyword:

```
cwd> labelit.image image_range=1,5
```

Keywords `image_range` and `image_number` are mutually exclusive and cannot be supplied together.

It is hoped that the availability of this method for stacking images will encourage crystallographers to

make it a common practice to acquire very finely sliced rotation images, which is now practical with the introduction of fast pixel array detectors such as the Pilatus-6M. Thin-sliced data (Pflugrath, 1999) have several advantages including improved signal-to-noise and the ability to model the Bragg spots with three-dimensional profiles. If it is easy to stack images for routine viewing, this removes the objection that thinly-sliced images are difficult to examine visually.

Note: it is possible to go immediately from raw images to PDF-format pictures *without* indexing first, if the object is to simply render the image without any markup. A separate command is provided for this purpose:

```
cwd> labelit.pdf <image template (/home/data/lysozyme_###.img)>
```

Keyword options are:

```
image_number = 1
image_range = 1,5
window_fraction = 0.4
window_center_x = 0.5
window_center_y = 0.5
image_brightness = 1.0
pdf_output{
  file = output.pdf
  box_size = 500
}
```

Synthesis of pseudo-precession photographs: *labelit.precession_photo*

While molecular structure is ideally explored with perfect crystals that give sharp Bragg peaks, it has been the imperfections that have posed large challenges over the years. The visual examination of images certainly plays an important role in diagnosing specific types of disorder (Nave, 1999). Aside from mosaicity, or isotropic disorder that gives rise to wide rocking curves (the diffraction of a Bragg spot over a large rotation angle), recent papers have highlighted specific types of long-range disorder that produce recognizable signatures at the "non-Bragg" positions of the diffraction pattern. For example, incommensurate modulation (Borgstahl *et al.*, 2009), a periodic distortion of the crystal lattice, generates discrete satellite Bragg spots; while lattice translocation disorder (Tsai *et al.*, 2009), the slight displacement of successive crystal layers, creates a pattern of streaks on specific spots.

Historically, the availability of Buerger precession cameras (*e.g.*, Blundell & Johnson, 1976) made it easy to examine specific planar layers of the reciprocal lattice, after painstaking alignment of the principal crystallographic axes (**a***, **b***, **c***) relative to the camera reference frame. In certain cases (Bragg & Howells, 1954), differences in the Bragg spot shape could be described as a function of Miller index. Achieving this type of convenient plot with modern rotation data requires a software calculation for two reasons: first, each image represents a curved surface of reciprocal space, not a plane; and secondly, the crystal axes are now rarely prealigned with the camera.

Assuming that images are available from a wide enough rotational range, the requisite planar section can be synthesized. However, it is best to keep in mind that there are implicit limitations. We assume, for example, that the crystal is rigidly fixed to the goniometer rotor, so its orientation is exactly known for each source image from the dataset. Deviations from this ideal will degrade the synthesized image, particularly at higher scattering angles. Also, our implementation does not apply scaling corrections. Thus, factors such as accumulated radiation dose that change the sample over time will cause symmetry-related reflections to appear unequal in intensity. Geometric approximations are unavoidable: in order to create a mapping between the raw image and reciprocal space coordinates, it is assumed that each raw image represents the center of its rotation range (for example, a 1° rotation image covering φ=[0°,1°] is uniformly assigned the value φ=0.5°). Moreover, image pixels far from the

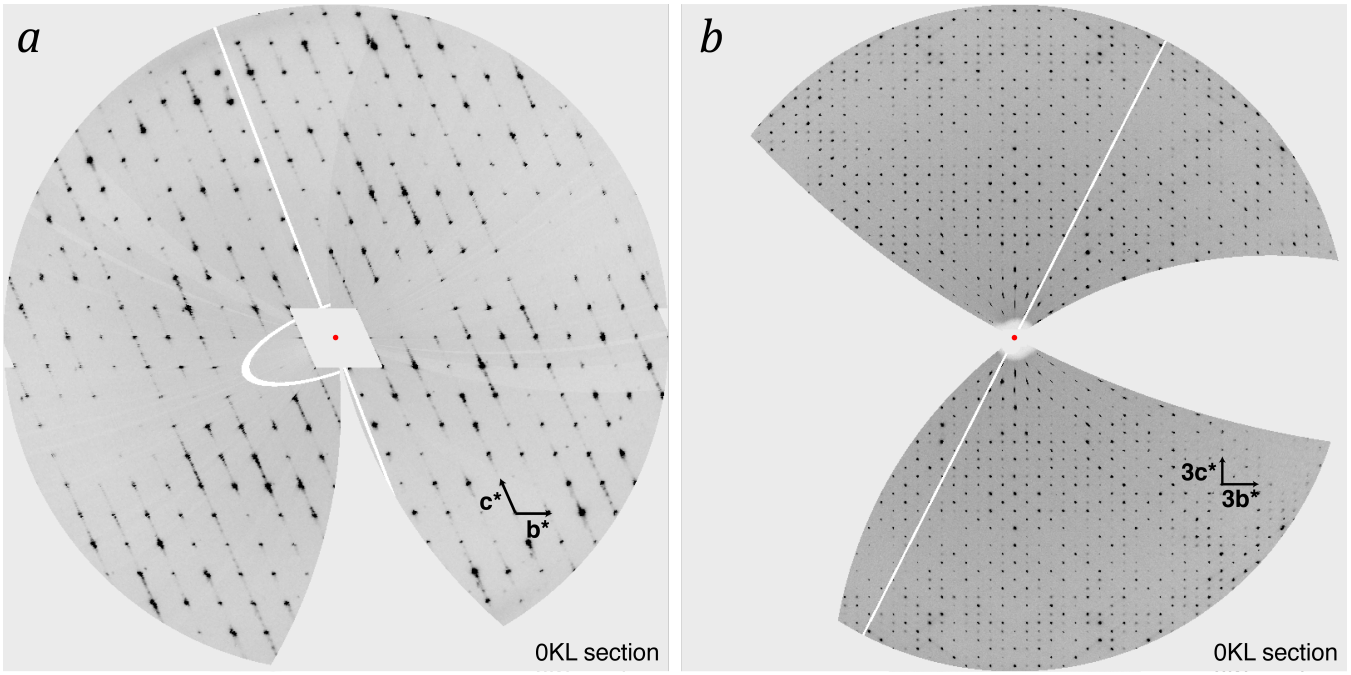


Figure 3. Reciprocal space sections illustrated with `labelit.precession_photo` for structures 1vk8 (*a*) and 2qyv (*b*). The origin of the streaks extending along the **c*** axis in (*a*) was not examined in the original publication (Dermoun *et al.*, 2010); but is presumably associated with lattice disorder.

rotation axis map to a larger rotational path through reciprocal space and appear as large quadrilaterals on the synthesized image. Clearly, the best sampling is obtained with fine rotational slicing (Pflugrath, 1999) and small image pixels.

Despite these caveats, `labelit.precession_photo` can be used (figure 3) to clearly illustrate phenomena that we recently cited: streaky Bragg spots of unknown origin associated with PDB code 1vk8 (Sauter & Poon, 2010) and a pattern of alternating weak and strong Bragg spots due to pseudotranslational symmetry in PDB structure 2qyv (Sauter & Zwart, 2009). The command line keywords for `labelit.precession_photo` are handled exactly as described above for `labelit.image`:

```
bravais_choice=None
image_range=None
pdf_output.file=None
pdf_output.box_size=500
```

Identical to the parameters described for `labelit.image`. A "None" value indicates required input. Here it is advantageous to specify an `image_range` covering the entire dataset, so that the coordinate grid of the synthesized image is filled in to the largest extent.

```
pixel_width=600
```

The width of the synthesized coordinate grid in pixels, which is then fit into the `pdf_output.box_size` dimension expressed in points.

```
resolution_outer=3.0
```

The high resolution limit of the requested plot, expressed in Ångstroms.

`intensity_full_scale=256`

Intensity value on the raw image that is treated as fully saturated (black).

`plot_section="H,K,0"`

Determine which principle axes are in the plane of the printed page; either **a***, **b*** (`H,K,0`); **b***, **c*** (`0,K,L`); or **a***, **c*** (`H,0,L`). Also, upper- and lower-layers can be sectioned, *i.e.*, "`H,K,1`"; "`H,K,-1`"; etc.

`layer_width=0`

The width of the reciprocal space section to be illustrated, given in fractional Miller index units. For example, if `plot_section` is "`H,K,0`" and `layer_width` is 0.05, then all image pixels mapping to reciprocal space coordinates between `H,K,-0.025` and `H,K,0.025` are plotted, with overlapping pixels being averaged. By default `layer_width=0`, corresponding to a section thickness of one pixel.

`apply_symmetry=None`

Point-group symmetry used to average the data. This option is not recommended by default, but is provided to respond to user comments that `labelit.precession_photo` printouts do not always look like true precession photographs. The full reciprocal space layer is not always covered. This is because the experimental rotation range is often less than the full 180° required for full coverage. Can point-group symmetry be applied to get an illustration that *looks* like a full precession photograph? Caution must be exercised, as such an operation could erase distinctions that might be important! As noted above the synthesized image normally reflects differences due to radiation damage, as well as variations in other factors such as the incident beam flux and the length of the absorption path. Moreover, the true symmetry of the diffraction may not be as high as implied by the Bravais choice, as with a monoclinic crystal with β angle close to 90°, which can be plotted within an orthorhombic cell. With these warnings in mind, the user can choose to impose point-group symmetry on the diffraction pattern if desired, with the `apply_symmetry` keyword.

Two choices must be made when selecting the point-group symmetry. First, should Friedel symmetry, $HKL = \overline{H}\overline{K}\overline{L}$, be imposed or not? Second, for certain crystal systems (such as tetragonal) there are alternate Laue groups to select. The Laue group cannot be selected automatically based on indexing alone, as it is necessary to compare symmetry-equivalent intensities after integration and scaling. To make the full matrix of choices clear, the user should type the undecorated command:

```
cwd> labelit.precession_photo
```

which outputs a table enumerating the point group choices for each possible Bravais setting, for example:

Bravais_choice	Lattice	Laue-group	Reflection-symmetry	Friedel-only
9	tP	4/mmm	422	-1
9	tP	4/m	4	-1
8	oC	mmm	222	-1
7	mC	2/m	2	-1
6	mC	2/m	2	-1
5	oP	mmm	222	-1
4	mP	2/m	2	-1
3	mP	2/m	2	-1
2	mP	2/m	2	-1
1	aP	-1	1	-1

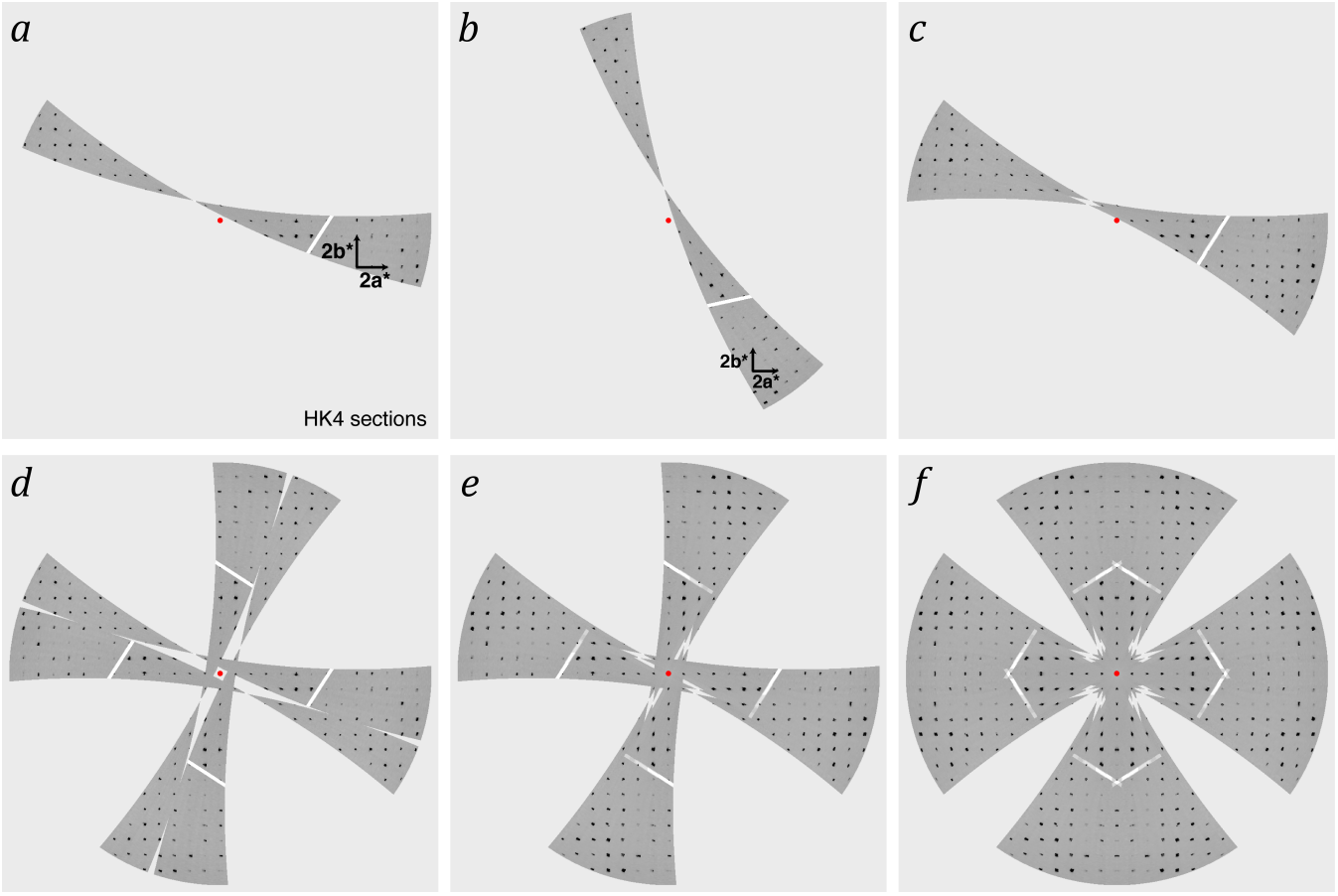


Figure 4. L=4 sections from the 1o3u diffraction pattern (Eriandsen *et al.*, 2004). The space group of the structure is $P4_12_12$, and the data are plotted either in the primitive tetragonal setting (*a*, *c*-*f*) or the *C*-centered orthorhombic setting (*b*). Point group symmetries imposed on the data are $\bar{1}$ (*c*), 4 (*d*), 4/m (*e*), and 4/mmm (*f*).

Various combinations of `bravais_choice` and `apply_symmetry` produce drastically different pictures, as illustrated in figure 4. Each panel plots the same 20° of rotation data from a tetragonal diffraction pattern. The `bravais_choice` keyword changes the axes on which the data are plotted without altering the intensities that are displayed, as seen by comparing panels (*a*, tP) and (*b*, oC). In contrast, the `apply_symmetry` option has the effect of increasing the reciprocal space coverage and/or averaging symmetry-redundant measurements of the displayed reciprocal space coordinates. For example, the application of Friedel symmetry (*c*) brings data from the L=–4 section on to the L=4 layer. Application of 4-fold symmetry (*d*) produces a clover-leaf pattern around the L-axis, while the combination of both Friedel and 4-fold symmetry (*e*) combines both effects. Finally, 4/mmm symmetry (*f*) yields two mirror planes at H=0 and K=0.

PNG- and GIF-format output

For completeness, we mention that *LABELIT* can also generate PNG-format images of the diffraction pattern:

```
cwd> labelit.png <image file> <output file.png> [-large]
cwd> labelit.overlay_distl <image file> <output file.png> [-large]
cwd> labelit.overlay_index <image file> <output file.png> [-large]
cwd> labelit.overlay_mosflm <image file> <output file.png> [-large]
```

The command `labelit.png` generates an undecorated image, while `labelit.overlay_distl` colors the subsets of bright spots either used (green) or not used (blue) for indexing. Commands

`labelit.overlay_index` and `labelit.overlay_mosflm` add markup of the predicted lattice for the highest Bravais choice, as refined by either *LABELIT* or *MOSFLM*, respectively. The only allowed keyword is `–large`, which imposes a one-to-one mapping between raw data pixels and pixels on the generated picture, otherwise the raw data pixels are binned in 2 × 2 squares.

An animated GIF-format movie can be generated of the entire dataset (this does not require indexing):

```
cwd> labelit.dataset_animation <template> <first image> <last image> <out>
cwd> labelit.dataset_animation /home/user/mydata/lyso_###.img 1 90 out.gif
```

Extensibility of Python code

Developers should be aware that the features discussed here could easily be extended by simple scripting in Python language. The applications discussed above are built on standard `cctbx` components for handling of detector formats (`iotbx.detectors`) and command-line keywords (`libtbx.phil`). Third party extensions are used for standard image formats (*Python Image Library*) and generation of PDF output (*Reportlab*).

Acknowledgments

Comments from software users were instrumental in developing the finished product. In particular, input from Tillman Heinisch (Universität Basel) and Jason Porta (University of Nebraska Medical Center) contributed significantly to `labelit.precession_photo`. The financial support of the National Institutes of Health / National Institute of General Medical Sciences under grant number R01-GM077071 is gratefully acknowledged. Operation of LBNL is partly supported by the US Department of Energy under Contract No. DE-AC02-05CH11231.

References

Blundell TL, Johnson, LN (1976). *Protein Crystallography*. London, Academic Press, Ltd.

Bragg WL, Howells ER (1954). X-ray diffraction by imidazole methaemoglobin. *Acta Crystallogr.* **7**, 409.

Dermoun Z, Foulon A, Miller MD, Harrington DJ, Deacon AM, Sebban-Kreuzer C, Roche P, Lafitte D, Bornet O, Wilson IA, Dolla A (2010). TM0486 from the hyperthermophilic anaerobe *Thermotoga maritima* is a thiamin-binding protein involved in response of the cell to oxidative conditions. *J. Mol. Biol.* **400**, 463-476.

Elsliger MA, Deacon AM, Godzik A, Lesley SA, Wooley J, Wüthrich K, Wilson IA (2010). The JCSG high-throughput structural biology pipeline. *Acta Crystallogr.* F**66**, 1137-1142.

Eriandsen H, Canaves JM, Elsliger MA von Delft F, Brinen LS, Dai X, Deacon AM, Floyd R, Godzik A, Grittini C, Grzechnik SK, Jaroszewski L, Klock HE, Koesema E, Kovarik JS, Kreusch A, Kuhn P, Lesley SA, McMullan D, McPhillips TM, Miller MD, Morse A, Moy K, Ouyang J, Page R, Robb A, Quijano K, Schwarzenbacher R, Spraggon G, Stevens RC, van den Bedem H, Velasquez J, Vincent J, Wang X, West B, Wolf G, Hodgson KO, Wooley J, Wilson IA (2004). Crystal structure of an HEPN domain (TM0613) from *Thermotoga maritima* at 1.75 Å resolution. *Proteins* **54**, 806-809.

Plugrath JW (1999). The finer things in X-ray diffraction data collection. *Acta Crystallogr.* D**55**, 1718-1725.

Sauter NK, Poon BK (2010). Autoindexing with outlier rejection and identification of superimposed lattices. *J. Appl. Crystallogr.* **43**, 611-616.

Sauter NK, Zwart PH (2009). Autoindexing the diffraction patterns from crystals with a pseudotranslation. *Acta Crystallogr.* D**65**, 553-559.

Electron density illustrations

Ralf W. Grosse-Kunstleve[a] and Luc J. Bourhis[b]
[a]*Lawrence Berkeley National Laboratory, Berkeley, CA 94720, U.S.A.*
[b]*Bruker AXS SAS, Champs-sur Marme, 77447 Marne-la-Vallée Cedex 2, France*

Correspondence email: RWGrosse-Kunstleve@LBL.Gov

Introduction

Crystallographers across all specializations invariably inspect electron density maps. This small article illustrates the effects of selected fundamental factors that shape a map, primarily to explain why macromolecular maps are often not scaled to absolute units (such as $e/Å^3$). Instead, "sigma scaling" is used, which is to compute the signal-to-noise ratio. Technically, this is to divide the density values in the map by their standard deviation (commonly called "sigma"), with the result that the standard deviation of the scaled values is one. Useful contour levels for graphical display are easily predictable for sigma-scaled maps, typically in the range 0.5-2. By contrast, the range of absolute electron density values is far more dependent on these factors:

- Isotropic Displacement Parameters (known as U_{iso} or B_{iso})
- High Resolution Limit
- Grid Resolution Factor
- Omission of the Fourier coefficient F_{000}

Here we show using a series of plots that these factors lead to a wide spread of absolute electron density values. We also take a brief look at how these parameters determine whether neighboring atoms are resolved in a density map.

B_{iso} dependence

For brevity, in the following we refer to isotropic displacement parameters as B_{iso} (see for example Grosse-Kunstleve & Adams (2002) and references therein). The B_{iso} dependence of electron density values is best illustrated via the Analytical Fourier Transform (AFT) since this eliminates the influences of the other factors in the list above. The formula for the AFT of a scatterer with displacement parameter B_{iso} and an N-Gaussian approximation to the X-ray scattering factor was given by Agarwal (1978) (see also Afonine & Urzhumtsev (2004) and references therein):

$$\rho(r) = \sum_{i=1}^{N} a_i \left(\frac{4\pi}{b_i + B_{iso}}\right)^{3/2} e^{-\frac{4\pi^2 r^2}{b_i + B_{iso}}} \qquad (1)$$

Here $\rho(r)$ is the electron density at a distance r (in Å) from the atomic center. a_i and b_i are the coefficients of the N-Gaussian approximation to the scattering factor $f(s)$ at the diffraction angle measurement $s = \sin\theta\,/\lambda$:

$$f(s) = \sum_{i=1}^{N} a_i e^{-b_i s^2} \qquad (2)$$

In the examples that follow, we work with a 5-Gaussian approximation (Grosse-Kunstleve et al., 2004) to the X-ray scattering factor of carbon. The corresponding Gaussian approximations of the International Tables for Crystallography Volume C (1992) and Waasmaier & Kirfel (1995) could not be used for the present purpose because they both involve a constant term, which is equivalent to a Gaussian term with $b_i = 0$. If the isotropic displacement parameter approaches zero, this leads to numerical instabilities when computing the AFT, as is apparent from equation (1). However, it should be noted that the International Tables, Wassmaier & Kirfel and 5-Gaussian approximations yield nearly identical results for B_{iso} values that are significantly different from zero.

Figure 1 shows AFT results with selected B_{iso} values ranging from 0.0 to 1.0 Å^2 in figure 1 (a) and powers

of 2

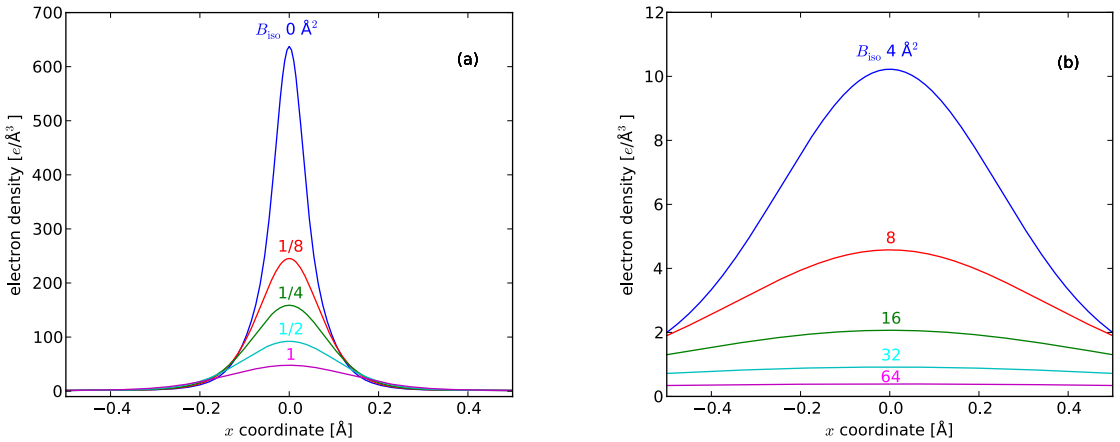


Figure 1: AFT results with selected B_{iso} values. The x coordinate is equivalent to the radial distance r of equation (1).

ranging from 4 to 64 Å^2 in figure 1 (b). The plots show that the absolute values of the electron densities are highly dependent on the displacement parameter. The electron density integrated over the unit cell is equal to the six carbon atom electrons for all values of B_{iso}, but the plots show how this charge is spread over the unit cell as B_{iso} increases, or conversely how it is concentrated around the carbon site as B_{iso} decreases. As a result, there cannot be any universally meaningful electron density contour level when the electron density is on an absolute scale.

High resolution limit

Figure 2 compares AFT and Discrete Fourier Transform (DFT) results with B_{iso}=0 Å^2. The unit cell used in the calculation is a cube with edge length 5 Å, but this value is not critical. The grid resolution factor for the DFTs is 1/4, which is used to determine the approximate map grid spacing by multiplication with the high resolution limit, for example $0.5\ \text{Å} \cdot 1/4 = 0.125\ \text{Å}$. The grid is constrained to uneven values to accommodate a symmetric range of Miller indices, for exampling ranging from -20 through 20 in each dimension, which leads to a slightly smaller grid spacing (5 Å / 41 = 0.122 Å).

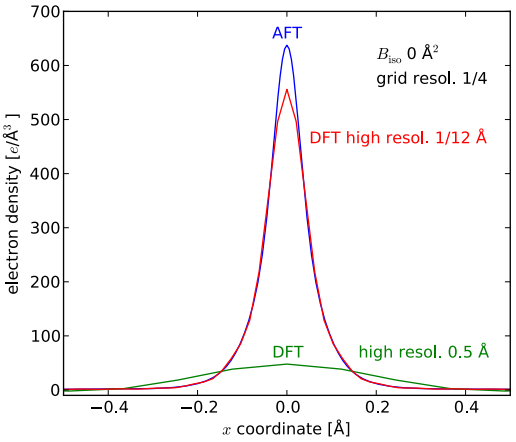


Figure 2: AFT and DFT results with B_{iso}=0 Å^2.

The blue plot is the AFT as before in figure 1. The red plot is the DFT of F_{calc} Fourier coefficients computed up to a high resolution of 1/12 Å (0.083 Å), which is the nominal limit of the 5-Gaussian approximation. The red plot is shown here mainly to demonstrate that the AFT is approximated well at an extremely high DFT resolution. The green plot is the DFT with Fourier coefficients up to a high resolution of 0.5 Å, which is still a very high resolution for macromolecular structures. The green plot illustrates that the omission of Fourier coefficients beyond the 0.5 Å resolution limit has a large impact on the electron density values if B_{iso}=0.

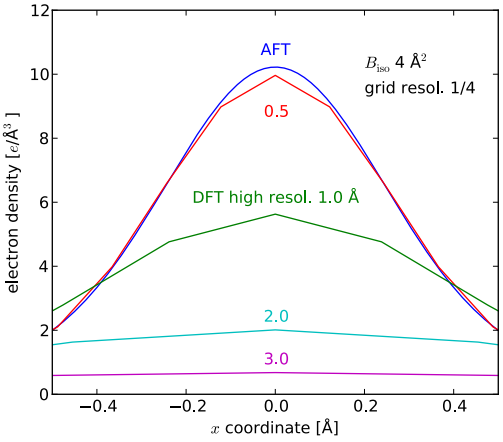


Figure 3: AFT and DFT results with B_{iso}=4 Å^2.

Figure 3 is similar to figure 2, but B_{iso}=4 Å^2 is used, a value in the typical range for small molecule crystal structures. In

this case the AFT and DFT electron density values are in close agreement if the high resolution limit is 0.5 Å. The plots with lower resolution limits (1.0, 2.0 and 3.0 A) demonstrate the still very strong influence of the resolution limit on the absolute electron density values given a non-zero B_{iso}.

Figure 4 is similar to the previous figure but B_{iso}=32 Å^2 is used, a value more typical for macromolecular crystal structures. Comparison of figures 3 and 4 illustrates that the discrepancy of the AFT and DFTs depends less strongly on the resolution limit if the value of the displacement parameter increases.

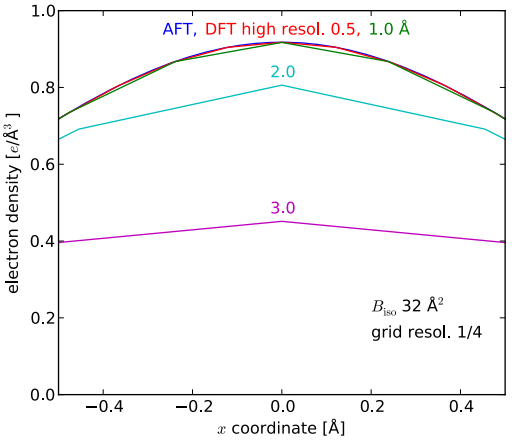


Figure 4: AFT and DFT results with B_{iso}=32 Å^2.

Grid resolution factor

Figure 5 illustrates the influence of the grid resolution factor on the appearance of the electron density. B_{iso}=4 Å^2 and a high resolution limit of 1.0 Å is used in the plots. The resolution factors are 1/2, 1/3, 1/4 and 1/8. It can be seen that the absolute electron density values are hardly affected by the resolution factor. Smaller resolution factors lead to smoother densities, but imply increased memory and runtime requirements. In most situations the factor 1/3 or 1/4 is a useful compromise.

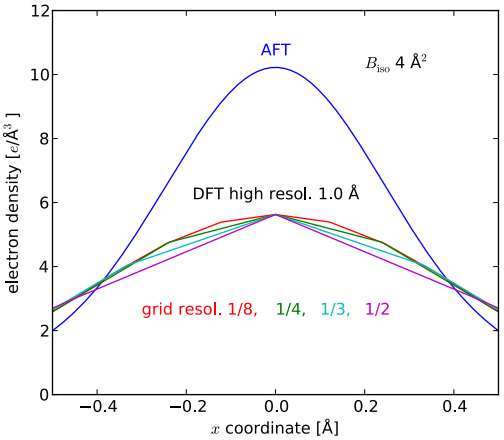


Figure 5: DFT results with a selection of grid resolution factors.

Omission of the Fourier coefficient F_{000}

Figure 6 illustrates the effect of omitting the F_{000} structure factor in the DFT, as is common practice. The contribution to the electron density is the constant $\langle\rho\rangle = F_{000}/V$, where V is the volume of the unit cell. The plots show the fractions contributed by $\langle\rho\rangle$ to electron density maxima at carbon positions of the model with PDB code 1ab1 (www.pdb.org); the values plotted are average fractions over all carbon positions. The DFTs were computed using F_{calc} with high resolution limits 0.5, 1.0, 2.0, 3.0 & 4.0 Å and carbon B_{iso} values 0, 4, 32, 64 and 128 Å^2. Similar calculations with other smaller and larger protein models showed that the plots in Figure 6 are typical. The $\langle\rho\rangle$ fraction of the electron density is a function of both the resolution and the displacement parameters. For very high resolution and small B_{iso} the $\langle\rho\rangle$ fraction is negligible but can grow to be significant at low resolutions and large B_{iso}. The $\langle\rho\rangle$ fraction in typical protein structures ranges from about 10% to 30%.

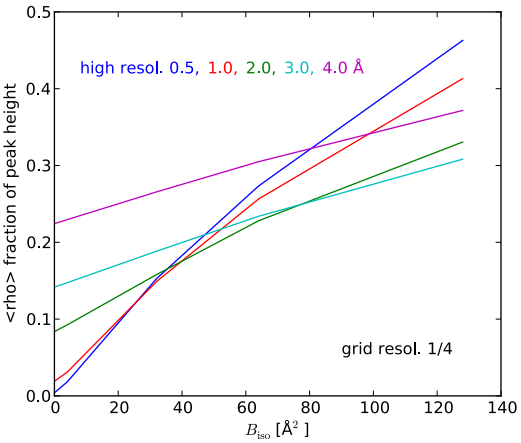


Figure 6: $\langle\boldsymbol{\rho}\rangle = \boldsymbol{F_{000}/V}$ contributions to maxima at carbon positions.

Two-atom examples

Theoretically, features separated by more than half the high-resolution limit of the Fourier coefficients can be distinguished in a DFT map. For example, two point scatterers separated by 1.5 Å should appear as separate maxima in a map if the Fourier coefficients extend to 3 Å or higher. Figure 7 illustrates the practical limits for resolving two neighboring carbon atoms with selected B_{iso} values and data resolution limits. The Gaussian shapes of both the X-ray scattering factor and the isotropic displacements lead to a

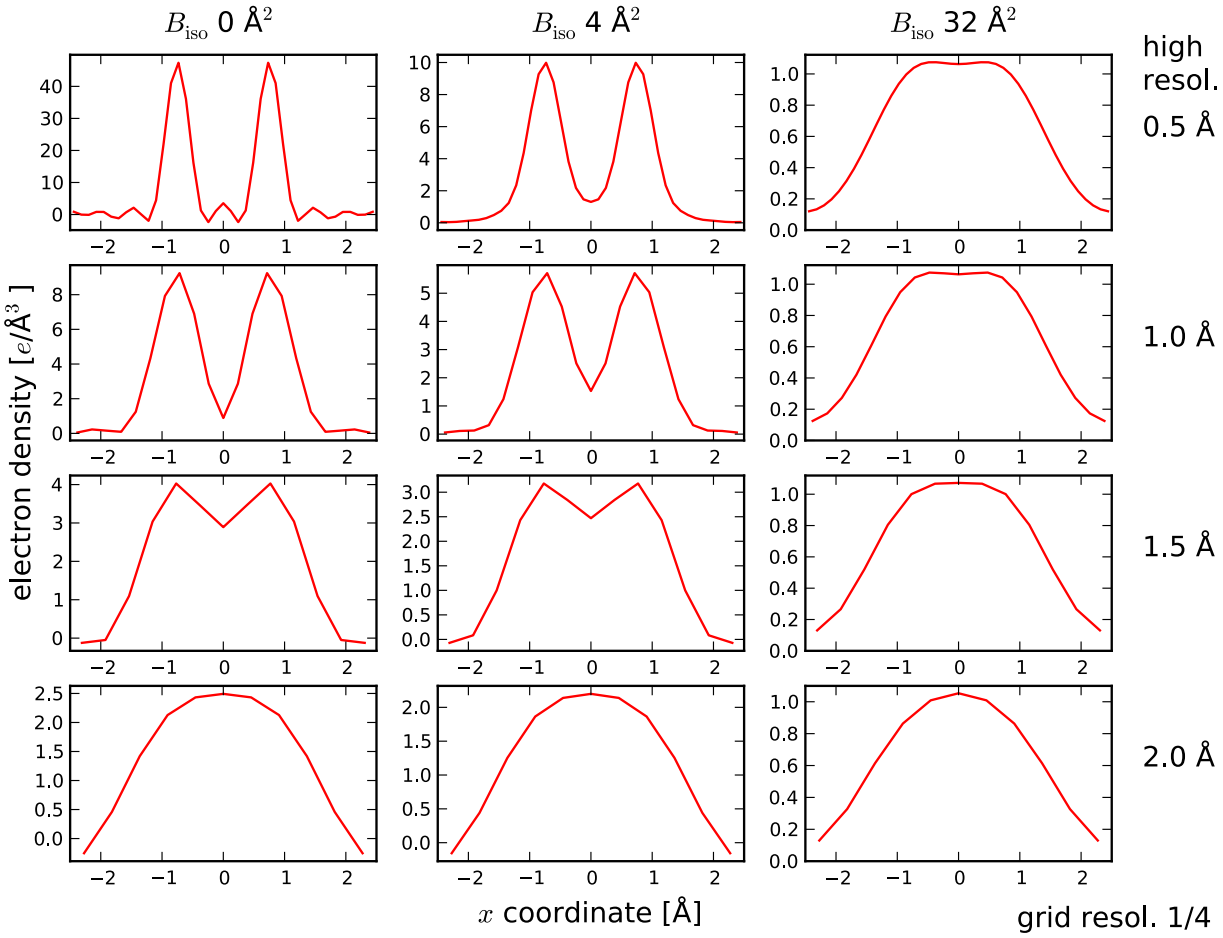


Figure 7: Examples of electron densities around two carbon atoms separated by 1.5 Å.

blurring of the electron density. Two carbon atoms separated by 1.5 Å are difficult to distinguish at 1.5 Å resolution even with B_{iso}=0 Å^2. Increasing B_{iso} values quickly reduce the effective resolution further.

Figure 7 also provides further examples of the spread of absolute electron density values; note the different scales of the y-axes in the matrix of plots.

References

Afonine, P.V. & Urzhumtsev, A.G. (2004). Acta Cryst. A**60**, 19-32.

Agarwal, R.C. (1978). Acta Cryst. A**34**, 791-809.

Grosse-Kunstleve, R.W., Adams, P.D. (2002). J. Appl. Cryst., **35**, 477-480.

Grosse-Kunstleve, R.W., Sauter, N.K., Adams, P.D. (2004). Newsletter of the IUCr Commission on Crystallographic Computing, **3**, 22-31.

International Tables for Crystallography Volume C (1992). Edited by A.J.C. Wilson, Kluwer Academic Publishers, Dordrecht/Boston/London.

Waasmaier, D., Kirfel, A. (1995). Acta Cryst., A**51**, 416-431.

Maximum likelihood refinement for twinned structures

Vladimir Y. Lunin
Institute of Mathematical Problems of Biology, Russian Academy of Sciences, Pushchino, Moscow region, 142290 Russia
Correspondence email: lunin@impb.psn.ru

Synopsis

A Rice-type probability distribution with properly adjusted parameters is a reasonable approximation for observed structure factor magnitudes when merohedral (or pseudo-merohedral) twinning is present. This allows to extend easy the usual maximum likelihood refinement technique for twinned structures.

Introduction

One of the most attractive ideas in crystallographic refinement is to enhance refinement power by maximisation of a likelihood function instead of the conventional minimisation of the least squares (LSQ) criterion. A special type of this likelihood function, which was used primarily for the evaluation of model quality (Lunin, 1982; Lunin & Urzhumtsev, 1984; Read, 1986; Lunin & Skovoroda, 1995; Urzhumtsev *et al*., 1996), was shown to be a good new goal function for the refinement of atomic models (Pannu & Read, 1996; Bricogne & Irwin, 1996; Murshudov et al., 1997; Adams et al., 1997). Now Maximum Likelihood (ML) refinement is an essential feature of such mainstream program complexes as phenix.refine (Afonine *et al*., 2005), REFMAC (Murshudov *et al*., 1997), Phaser (Read, 2001) and others. In this paper a simple way to extend this idea for refinement of twinned structures is discussed. In recent decades, twinning has been shown to be an important feature of macromolecular crystals (Yeates, 1997; Yeates & Fam, 1999; Dauter, 2003; Parsons, 2003; Lebedev et al., 2006). Twinned crystals are composed of separate differently orientated crystal specimens. If certain conditions for unit-cell parameters and orientation of the specimens are met (merohedral twinning), the reciprocal-space lattices of different domains coincide and the measured intensity of a diffracted beam becomes the sum of two (or more) different intensities that come from different specimens. The basic idea of ML refinement, namely, to maximise the probability to reproduce the observed values after allowed random corrections of the model have been done (Lunin, Afonine & Urzhumtsev, 2002), can be easily extended to take twinning into account. The problem is how to calculate the likelihood function practically. In this paper, a simple approximation for the likelihood function in the case of twinned intensities is suggested and tested. The found approximation allows conserving the usual "shape" of ML criterion and requires small corrections only in computational procedures.

1. Glossary

In this paper we distinguish four kinds of values:

- "true" or "theoretical" values that correspond to real structure and that are the goal of our study;
- "observed" values (intensities, or structure factor magnitudes) that were obtained in an experiment; they differ from the true values by experimental errors;
- "calculated" values that were obtained with the use of some model; these value differ from the true ones due to errors present in the model and approximations used in calculations;
- random values (structure factors, or errors) appear when we consider a value calcuated with the use of randomly generated additives.

2. Maximum likelihood refinement for twinned structures

2.1. Merohedral twinning

The phenomena of merohedral twinning may appear when the exact or approximate (pseudo-merohedral twinning) symmetry of crystal lattice exceeds the symmetry of crystal content. If $\mathbf{R}$ is such "extra" rotation symmetry operation and two specimens of the crystal linked by this rotation are

present in the X-ray beam simultaneously, then the diffraction patterns from two specimens overlap. If $I^{true}(\mathbf{s})$ is the intensity of $\mathbf{s}=(h,k,l)$ the indexed reflection for the first specimen, then theoretically, the measured intensity should be equal to

$$J^{theor}(\mathbf{s}) = (1-\kappa) I^{true}(\mathbf{s}) + \kappa I^{true}(\mathbf{R}^T\mathbf{s}) \tag{1}$$

Here $1-\kappa$ and κ are relative volumes of two specimens. The twinning fraction κ may be estimated by different methods (Yeates, 1988; Dauter, 2003; Lunin *et al*., 2007), which are outside of the scope of this paper.

2.2. Conventional least squares refinement

If some preliminary model of the studied structure exists a conventional least square refinement of the model parameters $\mathbf{q}$ could be performed by minimisation of discrepancy

$$Q_{LSQ}(\mathbf{q}) = \sum_{\mathbf{s}} \left(H^{calc}(\mathbf{s};\mathbf{q}) - H^{obs}(\mathbf{s}) \right)^2 \Rightarrow \min \tag{2}$$

with

$$H^{obs}(\mathbf{s}) = \sqrt{J^{obs}(\mathbf{s})}, \tag{3}$$

$$H^{calc}(\mathbf{s}) = \sqrt{(1-\kappa)\left|\mathbf{F}^{calc}(\mathbf{s};\mathbf{q})\right|^2 + \kappa\left|\mathbf{F}^{calc}(\mathbf{R}^T\mathbf{s};\mathbf{q})\right|^2}, \tag{4}$$

where $J^{obs}(\mathbf{s})$ are experimentally measured intensities from the twinned crystal.

A weakness of criterion (2) is that the model can contain irremovable errors so that no combination of flexible model parameters makes $\mathbf{F}^{calc}$ equal to $\mathbf{F}^{true}$ (and correspondingly H^{calc} equal to H^{true}). As a simple example, a portion of the atoms may be absent in the current model and their absence can not be compensated by moving of the atoms present in the model. Furthermore the target values H^{obs} in (2) differ from H^{true} by experimental errors. Furthermore, these errors cannot be corrected by changing of model parameters either. To some extent, these shortcomings may be overcome in the framework of maximum likelihood approach.

2.3. Maximum likelihood refinement

Maximum likelihood approach draws into refinement additional information present in the form of some statistical pattern for irremovable errors. For example one can consider measurement errors

$$\delta(\mathbf{s}) = J^{obs}(\mathbf{s}) - J^{true}(\mathbf{s}) \tag{5}$$

as independent random variables normally distributed with zero mean and variance $\sigma^2(\mathbf{s})$ (estimated in the experiment for every reflection separately). Similarly one can assume the any absences of atoms in the current model be uniformly distributed in the unit cell. After statistical properties of irremovable errors has been modeled the question may be asked "How large is the probability to get the calculated values equal to the observed values after random corrections have been introduce following the defined statistical pattern of irremovable errors?" This probability is called as “statistical likelihood” and it may be calculated for the independent observations as the product

$$L = \prod_{\mathbf{s}} Probability\left\{H^{cor}(\mathbf{s};\mathbf{q}) = H^{obs}(\mathbf{s})\right\}, \tag{6}$$

where $H^{cor}(\mathbf{s})$ stands for the random value that is the result of improvement of calculated values by random corrections. The likelihood may be adopted as a measure of goodness of the current structure model: the larger the likelihood value (6) the more reasonable the model. The choice of the model parameters resulting in the largest likelihood lies in the basis of maximum likelihood approach in mathematical statistics.

2.4. Calculation of the likelihood

One of the key problems in applications of the ML approach is the calculation of probability distributions for "randomly corrected" values, *i.e.* for $H^{cor}(\mathbf{s})$ values in our case. If the irremovable errors are restricted to the model incompleteness and measurement errors one can define random value $H^{cor}(\mathbf{s})$ as

$$H^{cor}(\mathbf{s};\mathbf{q}) = \sqrt{(1-\kappa)\left|\mathbf{F}^{cor}(\mathbf{s};\mathbf{q})\right|^2 + \kappa\left|\mathbf{F}^{cor}(\mathbf{R}^T\mathbf{s};\mathbf{q})\right|^2 + \delta(\mathbf{s})}, \tag{7}$$

$$\mathbf{F}^{cor}(\mathbf{s};\mathbf{q}) = \mathbf{F}^{part}(\mathbf{s};\mathbf{q}) + \mathbf{U}^{lost}(\mathbf{s}). \tag{8}$$

Here $\mathbf{q}$ is a set of current model parameters; $\mathbf{F}^{part}(\mathbf{s};\mathbf{q})$ are usual (deterministic) structure factors corresponding to the current partial model; $\delta(\mathbf{s})$ is a random error distributed with the normal distribution with zero mean and variance $\sigma^2(\mathbf{s})$; $\mathbf{U}^{lost}(\mathbf{s})$ are random structure factors calculated from randomly generated atomic positions.

A reasonable approximation to probability distribution of random variable $F^{cor}(\mathbf{s})$ is known (see e.g., Srinivasan & Parthasarathy, 1976) and for a general type non-centrosymmetric reflection is

$$P_F(F;\mathbf{s}) = \frac{2F}{\Sigma_Q(s)}\exp\left[-\frac{F^2 + \left(F^{part}(\mathbf{s})\right)^2}{\Sigma_Q(s)}\right] I_0\left(\frac{2FF^{part}(\mathbf{s})}{\Sigma_Q(s)}\right) \tag{9}$$

with

$$\Sigma_Q(s) = \sum_{j=1}^{N_{lost}} f_j^2(s), \tag{10}$$

where $f_j(s)$ are atomic scattering factors for the missing atoms and I_0 is the modified Bessel function. The distribution of the form (9) is often known as the Rice distribution.

The distribution for $F^{cor}(\mathbf{s})$ may have a more general form with two parameters $\alpha(\mathbf{s})$and $\beta(\mathbf{s})$ (Lunin 1982; Lunin & Urzhumtsev, 1984; Uzhumtsev *et al.*, 1996)

$$P_F(F;\mathbf{s}) = \frac{2F}{\beta(\mathbf{s})}\exp\left[-\frac{F^2 + \alpha^2(\mathbf{s})\left(F^{part}(\mathbf{s})\right)^2}{\beta(\mathbf{s})}\right] I_0\left(\frac{2\alpha(\mathbf{s})FF^{part}(\mathbf{s})}{\beta(\mathbf{s})}\right) \tag{11}$$

If more sources of errors are taken into account, *e.g.* atomic scattering factors are not known exactly, common scaling factor should be applied to $H^{cor}(\mathbf{s})$ before comparing with $H^{obs}(\mathbf{s})$, some errors are present in positions of atoms included in a rigid body block. Non-uniform prior coordinate distributions for the missing atoms may be involved in this scheme as well (Afonine *et al.*, unpublished).

2.5. Rice-type approximation for twinned intensity distribution

To get the final distribution that could be used to calculate the likelihood (6) one should calculated the triple convolution of probability distributions for $\left|\mathbf{F}^{cor}(\mathbf{s})\right|^2$, $\left|\mathbf{F}^{cor}(\mathbf{R}^T\mathbf{s})\right|^2$ and $\delta(\mathbf{s})$. This is not a simple task and the result cannot be presented in a simple analytical form. To avoid this problem one can skip the problem of convolution and suppose that the random values $H^{cor}(\mathbf{s})$ obey directly the distribution (11) with adequately designed $\alpha(\mathbf{s})$and $\beta(\mathbf{s})$ parameters. A reason for this hypothesis is that distribution (11) originates from Gaussian two-dimensional distribution. It is simply the marginal distribution of the distance-coordinate when the two-dimensional Gaussian distribution is written in polar coordinates. Due to the Central Limit Theorem (of theory of probabilities) Gaussian distribution appears in many different circumstances so that one can hope that its derivative (11) is suitable for $H^{cor}(\mathbf{s})$ values as well.

To check this hypothesis a series of test was performed (See section 3 below for the details). In these tests simulated sets of $H^{obs}(\mathbf{s})$ values were checked against theoretical distributions (11) with properly adjusted parameters. It was found that correspondence is very good for relatively strong reflections and is reasonable for weak ones. More definitely, the similarity of empirical and theoretical curves depended mostly on relative value of intensity of reflection in comparison with the part of the true intensity corresponding to the missing atoms. To characterise a reflection power we used the ratio $\lambda = H^{calc}(\mathbf{s})\big/\sqrt{\Sigma_Q(s)}$, where $\Sigma_Q(s)$ is mean intensity corresponding to the lost atoms defined by (10). Figure 1 shows empirical and theoretical probability distribution for different values of this ratio. A more accurate comparison was made with the use of χ^2 criterion. For large enough values of the ratio ($\lambda > 1$) these tests give no reason to reject the null hypothesis that the distribution of $H^{cor}(\mathbf{s})$ values is consistent with probability distribution (11). At the same time, for weak reflection the confidence level was small enough (see test 9 in table 1) and the approximation becomes less

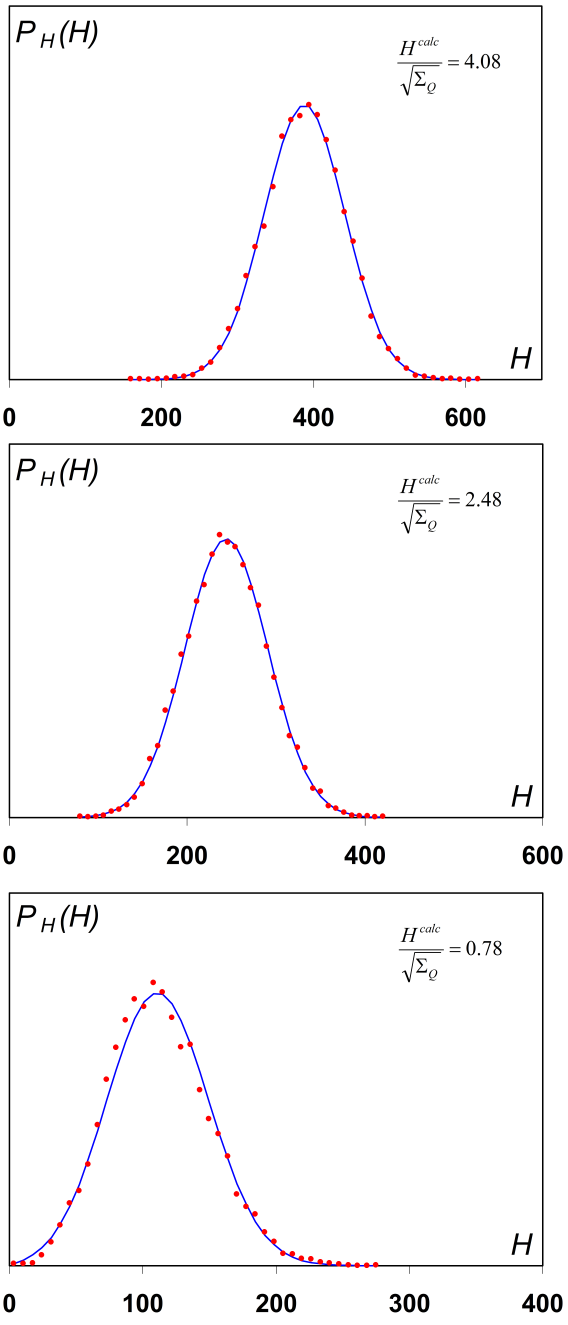


Figure 1. Theoretical distributions (11) with optimally defined parameters α/β (solid line) and distributions for simulated H^{obs} values (dots) for reflections with different $H^{calc}\big/\sqrt{\Sigma_Q}$ ratio. (See section 3 for details).

accurate (figure 1c). It is noteworthy that the tests did not reject the hypothesis when the twinning fraction was set as $\kappa = 0$. This means that in the usual ML refinement (without twinning) experimental measurement errors may be "absorbed" by α / β values.

A similar approach may be used if one has more then one twinning operation present. The only difference is that the calculated values $H^{calc}(\mathbf{s})$ are a mixture of more then two intensities.

2.6. Maximum likelihood refinement for twinned crystals

If the parameters α and β are defined and $\mathbf{q}$ is the set of flexible parameters of the partial model (*i.e.* parameters that are allowed to be changed in the refinement) then the likelihood function (6) takes the form

$$L(\mathbf{q}) = \prod_{\mathbf{s}} \frac{2H^{obs}}{\beta(\mathbf{s})} \exp\left[-\frac{\left(H^{obs}(\mathbf{s})\right)^2 + \alpha^2(\mathbf{s})\left(H^{calc}(\mathbf{s};\mathbf{q})\right)^2}{\beta(\mathbf{s})} \right] I_0\left(\frac{2\alpha(\mathbf{s}) H^{obs}(\mathbf{s}) H^{calc}(\mathbf{s};\mathbf{q})}{\beta(\mathbf{s})} \right) \quad (12)$$

(This formula supposes the non-centosymmetric reflection of a general type only be used in the calculation; see section 1.7 below for a more general case).

The calculation of logarithm of (12), changing the sign and omitting the terms that do not depend on the flexible model parameters, reduces the problem of maximisation of (12) to the problem of minimisation of the standard ML criterion

$$Q_{ML}(\mathbf{q}) = \sum_{\mathbf{s}} \left\{ \frac{\alpha^2(\mathbf{s})\left(H^{calc}(\mathbf{s};\mathbf{q})\right)^2}{\beta(\mathbf{s})} - \ln I_0\left(\frac{2\alpha(\mathbf{s}) H^{obs}(\mathbf{s}) H^{calc}(\mathbf{s};\mathbf{q})}{\beta(\mathbf{s})} \right) \right\} \Rightarrow \min \quad (13)$$

with $H^{obs}(\mathbf{s})$ and $H^{calc}(\mathbf{s})$ values defined as (3-4). The only difference with the conventional ML-criterion is that $F^{calc}(\mathbf{s})$ values as replaced by $H^{calc}(\mathbf{s})$.

2.7. Maxumum likelihood estimates of α and β parameters

The minimisation (13) supposes that the parameters $\alpha(\mathbf{s})$ and $\beta(\mathbf{s})$ are known for all reflections. To define these parameters providing model parameters are known, the maximum likelihood principle may be applied again (Lunin & Urzhumtsev, 1984; Read, 1986; Lunin&Skovoroda, 1995). Let us suppose that for a thin spherical shell in the reciprocal space parameters $\alpha(\mathbf{s})$ and $\beta(\mathbf{s})$ are the same for all reflection in the shell, i.e. $\alpha(\mathbf{s}) = \alpha$, $\beta(\mathbf{s}) = \beta$ and the product in (12) is calculated over reflection from this zone only. Consider now that the model parameters $\mathbf{q}$ in (12) be known and fixed, then the likelihood function depends on two parameters α and β only and these parameters may be defined by maximisation of the likelihood again. As before, this maximisation can be reduced to minimisation of

$$Q_{\alpha\beta}(\alpha,\beta) = K \ln\beta + \frac{1}{\beta}\sum_{\mathbf{s}}\left(H^{obs}(\mathbf{s})\right)^2 + \frac{\alpha^2}{\beta}\sum_{\mathbf{s}}\left(H^{calc}(\mathbf{s})\right)^2 - \sum_{\mathbf{s}} \ln I_0\left(\frac{2\alpha H^{obs}(\mathbf{s}) H^{calc}(\mathbf{s})}{\beta} \right) \quad (14)$$

where K is the number of reflections included in the sums. Now only the terms depending on α and β are included in minimisation criterion. A technique of minimisation of this function is described in

(Lunin & Skovoroda, 1995).

As it was shown before for usual refinement (Lunin, 1982; Lunin & Skovoroda, 1995) parameters α and β and model parameters $\mathbf{q}$ are highly biased if defined from the same set of reflections. So that α and β parameters should be defined on the base of test reflections only (*i.e.* ones excluded from refinement). The same observation is valid in the presence of twinning as well. Furthermore, some special attention should be paid when splitting the full set of reflections into work and test parts. The pair of reflections linked by the twinning law (*i.e.* $\mathbf{s}$ and $\mathbf{R}^T\mathbf{s}$) should have similar type ("work" or "test").

2.8. Remark on centrosymmetric and special reflections

The distribution (11) may be used as is for a general type non-centrosymmetric reflection. To go to a more common case we should correct distribution (11) for non-centrosymmetric reflection as

$$P_H\left(H;\mathbf{s}\right)=\frac{2H}{\varepsilon\left(\mathbf{s}\right)\beta\left(\mathbf{s}\right)}\exp\left[-\frac{H^2+\alpha^2\left(\mathbf{s}\right)\left(H^{calc}\left(\mathbf{s}\right)\right)^2}{\varepsilon\left(\mathbf{s}\right)\beta\left(\mathbf{s}\right)}\right]I_0\left(\frac{2\alpha\left(\mathbf{s}\right)HH^{calc}\left(\mathbf{s}\right)}{\varepsilon\left(\mathbf{s}\right)\beta\left(\mathbf{s}\right)}\right) \quad (15)$$

Here the parameter $\varepsilon\left(\mathbf{s}\right)$ depends only on the reflection indices and on the particular space group $\Gamma=\left\{\left(\mathbf{G}_\nu,\mathbf{t}_\nu\right)\right\}_{\nu=1}^{n}$ and may be calculated as the number of reciprocal-space symmetries $\mathbf{G}_\nu^T$ that when applied to the vector $\mathbf{s}$ leave it invariable, i.e. $\mathbf{G}_\nu^T\mathbf{s}=\mathbf{s}$. For centrosymmetric reflection the corresponding probability distribution has the form

$$P_H\left(H;\mathbf{s}\right)=\sqrt{\frac{2}{\pi\varepsilon\left(\mathbf{s}\right)\beta\left(\mathbf{s}\right)}}\exp\left[-\frac{H^2+\alpha^2\left(\mathbf{s}\right)\left(H^{calc}\left(\mathbf{s}\right)\right)^2}{2\varepsilon\left(\mathbf{s}\right)\beta\left(\mathbf{s}\right)}\right]\cosh\left(\frac{\alpha\left(\mathbf{s}\right)HH^{calc}\left(\mathbf{s}\right)}{\varepsilon\left(\mathbf{s}\right)\beta\left(\mathbf{s}\right)}\right) \quad (16)$$

The changes in (12-14) are straightforward. We do not include the statistical weight $\varepsilon\left(\mathbf{s}\right)$ into $\beta\left(\mathbf{s}\right)$ parameter as it would destroy the hypothesis that all $\beta\left(\mathbf{s}\right)$ are the same in a thin spherical shell in the reciprocal space, which is used to define α and β parameters provided the current model is fixed (see section 2.7).

3. Testing the null hypothesis

3.1. Tests and results

A series of tests were performed to check to what extent the Rice-type distribution (11) might approximate the distribution of measured twinned intensity for a particular reflection. In every such test a sample of "experimental" magnitudes $\left\{H_j^{obs}\right\}_{j=1}^{N}$ was generated by a Monte Carlo type procedure for some fixed reflection $\mathbf{s}$ taking into account three sources of uncertainties:

- "missing" atoms, i.e. ones present in the real structure, but absent in the current partial model;
- an experimental error when measuring intensities;
- an arbitrary scale for the measured intensities.

The parameters α^{opt} and β^{opt} in the theoretical distribution

$$P^{theor}\left(H;\alpha,\beta\right)=\frac{2H}{\beta}\exp\left[-\frac{H^2+\alpha^2\left(H^{calc}\right)^2}{\beta}\right]I_0\left(\frac{2\alpha HH^{calc}}{\beta}\right) \qquad (17)$$

that matches best the "observed" magnitudes were defined by maximisation of the likelihood

$$L\left(\alpha,\beta,\right)=\prod_{j=1}^{N}P^{theor}\left(H_j^{obs};\alpha,\beta\right)\Rightarrow \max \qquad (18)$$

(see details below). Figure 1 allows visual comparison of optimally fitted theoretical distributions with the empirical ones calculated from generated H_j^{obs} sets. These graphs reveal quite reasonable correspondence and back the idea that distribution (11) is suitable to describe deviations in experimental magnitudes when a twinning and experimental errors are present. To estimate the correspondence quantitatively the "null hypotheses" that $\left\{H_j^{obs}\right\}_{j=1}^{N}$ are random numbers generated following to $P^{theor}\left(H;\alpha^{opt},\beta^{opt}\right)$ probability distribution was checked under χ^2 criterion. Table 1 provides the obtained χ^2 value as well as the significance value defined as the probability to get the obtained or even worse (larger) value of χ^2 in the case when $\left\{H_j^{obs}\right\}_{j=1}^{N}$ are really obtained with the tested distribution (i.e. the null hypothesis is true). The null hypotheses should be rejected if the corresponding confidence level is small. Table 1 shows that for strong reflections one has no reasons to reject the null hypothesis. This is not the case for weak reflections. Nevertheless, the differences in empirical and theoretical distributions concern mostly "tails" of distribution. In ML refinement one tries to adjust the H^{calc} values to maxima of corresponding distributions, so that inaccuracies in "tails" of distributions for weak reflections should not affect the results of refinement too greatly.

Table 1. Results from a number of tests explained in the text. H^{calc} and Σ_Q are defined by equations (4) and (10) correspondingly.

test number	κ	$\frac{\sigma_I}{\left(H^{calc}\right)^2}$	$\frac{H^{calc}}{\sqrt{\Sigma_Q}}$	theoretical		defined		# bins	χ^2	confidence p	Fig.
				α	β	α^{opt}	β^{opt}				
1	0.	0.	5.04	1.	8701.	1.0	8690.	29	26.3	0.61	
2	0.	0.	2.37	1.	8675.	0.998	8721.	29	34.2	0.23	
3	0.	0.	0.50	1.	8601.	0.678	9828.	29	20.6	0.87	
4	0.	0.05	5.04			1.001	8897.	29	36.0	0.17	
5	0.	0.05	2.37			0.998	8832.	29	27.6	0.54	
6	0.	0.05	0.50			0.682	9816.	29	21.6	0.83	
7	0.36	0.05	4.08			1.011	5647.	29	24.1	0.73	1a
8	0.36	0.05	2.48			1.039	4435.	29	24.2	0.72	1b
9	0.36	0.05	0.78			1.435	3087	29	114.	$4.2\text{x}10^{-12}$	1c

3.2. Test parameters

Every particular test was defined by a set of parameters:

- the indexes of two twinned reflections $\mathbf{s}=(h_1,k_1,l_1)$ and $\mathbf{R}^T\mathbf{s}=(h_2,k_2,l_2)$ and magnitudes and phases of the corresponding structure factors calculated for a partial model; in our test the partial model was obtained by exclusion of 10% of atoms randomly from 1C5E (Yang *et al.*, 2000) model;
- the twin fraction value κ; (this value was defined as 0.36 for 1C5E crystal);
- the space group ($P2_1$ for 1C5E);
- the number $N_{missing}$ of atoms that are present in the whole structure, but are absent in the partial model; the positions of these atoms were considered as independent random variables uniformly distributed in the unit cell; $N_{missing}$ was set to 209 in our tests (10% of the whole number of atoms in 1C5E model);
- the accuracy of measuring of intensities σ_I; the measuring error δ was considered as normally distributed random variable with zero mean and variance σ_I^2; σ_I was set to zero or to $0.05*\left(H^{calc}\right)^2$ in our test;
- the scale factor $Scale$ converting magnitudes into some relative scale; It was found that this coefficient influence on optimal α and β values, bit does not influence on the quality of approximation; the results in table 1 correspond to value $Scale=1$;
- start value for random number generator, number of bins for χ^2 statistics, *etc.*

As a result simulated magnitudes H^{obs} were calculated as

$$H^{obs}=Scale*\sqrt{I^{twin}+\delta}\,, \tag{19}$$

$$I^{twin}=(1-\kappa)\left|\mathbf{F}^{part}(h_1,k_1,l_1)+\mathbf{U}(h_1,k_1,l_1)\right|^2+\kappa\left|\mathbf{F}^{part}(h_2,k_2,l_2)+\mathbf{U}(h_2,k_2,l_2)\right|^2 \tag{20}$$

where $\mathbf{U}(h,k,l)$ is a structure factor calculated from randomly generated coordinates of $N_{missing}$ atoms and δ is generated randomly with Gaussian distribution $N(0,\sigma_I^2)$. A characteristic value

$$\lambda=\frac{H^{calc}}{\sqrt{\Sigma_Q}} \tag{21}$$

with Σ_Q define by (10) was calculated for every test.

3.3. Special case: no twinning, no measurement errors

If the twinning and measurement errors are absent close formulas (9-10) exist for the probability distribution and parameters α and β. This may be used to check the developed numerical procedure (see section 4 below) for estimation of α and β parameters as well as to check the extent the generated $\left\{H_j^{obs}\right\}_{j=1}^N$ values are consistent with the theoretical distribution. Results of three such tests carried out for different λ-ratio occupy the first three lines in table 1. It is worthy of noting that though for weak reflection (test 3) the α^{opt} and β^{opt} parameters are different from the theoretical ones but the consistency of sample data with the found distribution is not worse then when using the distribution with theoretical parameter values. The number of bins, χ^2 value and confidence p are 31, 26.0 and 0.72 correspondingly in the last case.

3.4. Conventional maximum likelihood: no twinning, measurement errors present

Tests 4-6 were performed for the case when twinning is absent, but experimental errors are present. In this case $H^{calc}(\mathbf{s})$ coincides with $F^{part}(\mathbf{s})$. The goal of these tests was to estimate to what extent distributions (11) may be used in conventional ML-refinement without explicit corrections for measurement errors. The results support an idea that in refinement (13) these errors are taken into account implicitly be means of changing of the α and β parameters.

3.5. Common case

Tests 7-9 included both a twinning and measurement errors. For strong and medium reflections the distributions (11) describe well sample data $H^{obs}(\mathbf{s})$. For weak reflections ($\lambda < 1$) results are worse (test 9). Nevertheless the visual comparison of graphs in figure 1c shows that at "qualitative" level the difference is not too dramatic and these distributions could be used for weak reflections as well.

4. Maximum likelihood estimates of the distribution parameters

A problem we meet in these tests is how to define the parameters of distribution (17) that makes it the most consistent with sample data $\left\{H_j^{obs}\right\}_{j=1}^{N}$. Formally, this problem is a particular case of a more general problem studied in (Lunin & Skovoroda, 1995) where sample data were supposed to have different probability distributions (with different H_j^{calc} values) linked through the common α and β parameters. Nevertheless, this particular case is a "singular point" of the more general approach as some key parameter

$$\Omega = \left\langle \left(H^{obs}\right)^2 \left(H^{calc}\right)^2 \right\rangle - \left\langle \left(H^{obs}\right)^2 \right\rangle \left\langle \left(H^{calc}\right)^2 \right\rangle \qquad (22)$$

is equal to zero if all H_j^{calc} are the same. This requires a separate study of the problem.

4.1. Normalisation of variables

First we rewrite the problem (17-18) using a more convenient for analysis notation. Let the raw moments of the sample data $\left\{H_j^{obs}\right\}_{j=1}^{N}$ be

$$B_1 = \left\langle H^{obs} \right\rangle = \frac{1}{N}\sum_{j=1}^{N} H_j^{obs}, \quad B_2 = \left\langle \left(H^{obs}\right)^2 \right\rangle, \quad B_4 = \left\langle \left(H^{obs}\right)^4 \right\rangle, \qquad (23)$$

where $\langle . \rangle$ means the arithmetic mean, *i.e.* for any function $\psi(H)$ we define

$$\left\langle \psi\left(H^{obs}\right) \right\rangle = \frac{1}{N}\sum_{j=1}^{N} \psi\left(H_j^{obs}\right) \quad . \qquad (24)$$

If at least two observation in the set $\left\{H_j^{obs}\right\}_{j=1}^{N}$ are different, then

$$B_2 - B_1^2 = \left\langle \left(H^{obs} - \left\langle H^{obs} \right\rangle\right)^2 \right\rangle > 0 \,, \qquad (25)$$

so that

$$\frac{B_1^2}{B_2} < 1 \quad , \quad \frac{B_1}{\sqrt{B_2}} < 1 \quad . \tag{26}$$

Let normalised variables be

$$u = \sqrt{\frac{B_2}{\beta}} \, , \quad t = \frac{\alpha H^{calc}}{\sqrt{\beta}} \, , \quad z^{obs} = \frac{H^{obs}}{\sqrt{B_2}} \quad , \tag{27}$$

then the maximisation of the likelihood (17-18) is equivalent to minimisation of the function

$$Q(u,t) = -2\ln u + u^2 + t^2 - \left\langle \ln I_0\left(2utz^{obs}\right)\right\rangle , \qquad u \geq 0, t \geq 0 \, . \tag{28}$$

(The statistical weight $\varepsilon(\mathbf{s})$ is included into β parameter in this case).

4.2. The minimum of Q(u,t) at the border of the allowed region

Let study first the behavior of function $Q(u,t)$ at the border of the allowed region $u \geq 0, t \geq 0$. This function tends to infinity if $u \to 0$. Using an asymptotic expansion

$$\ln I_0(x) = x - \frac{1}{2}\ln x + \ldots \quad , x \to +\infty \tag{29}$$

we get for large values of the product ut

$$Q(u,t) = \left(t - \frac{B_1}{\sqrt{B_2}}u\right)^2 + \left(1 - \frac{B_1^2}{B_2}\right)u^2 + \ldots \quad , \quad ut \to +\infty \, , \tag{30}$$

so that $Q(u,t)$ function grows to infinity if u or t parameter grows. If $t = 0$, then the function

$$Q(u,0) = -2\ln u + u^2 \tag{31}$$

has the unique minimum at $u = 1$. As a result, the minimum at the border of the allowed region (including infinity) is attained for $(u,t) = (1,0)$.

4.3. Stationary point equation

If a minimum is attained at some inner point, then at this point one has

$$\begin{cases} \dfrac{\partial Q}{\partial u} = -\dfrac{2}{u} + 2u - t\Lambda(ut) = 0 \\ \dfrac{\partial Q}{\partial t} = 2t - u\Lambda(ut) = 0 \end{cases} \tag{32}$$

where the function $\Lambda(x)$ is defined as

$$\Lambda(x) = \left\langle 2z^{obs}\frac{I_1\left(2z^{obs}x\right)}{I_0\left(2z^{obs}x\right)}\right\rangle \quad . \tag{33}$$

Multiplying the first equation in (32) by u, the second by t and taking the difference we get

$$u^2 - t^2 = 1 \quad , \quad u = \sqrt{1+t^2} \tag{34}$$

The second equation in (32) may be used now to get the necessary condition for the point of the minimum

$$\Psi(t) = 2t - \sqrt{1+t^2}\,\Lambda\left(t\sqrt{1+t^2}\right) = 0 \quad . \tag{35}$$

4.4. Uniqueness of solution of Ψ(t)=0

We have $\Psi(0) = 0$ and using the asymptotic

$$\frac{I_1(x)}{I_0(x)} = 1 - \frac{1}{2x} + \ldots \quad , \quad x \to +\infty \tag{36}$$

we get

$$\Psi(t) = 2\left(1 - \frac{B_1}{\sqrt{B_2}}\right)t + \ldots \quad , \quad t \to +\infty \quad . \tag{37}$$

so that (due to (26)) $\Psi(t)$ increases for large t. As a result we have two alternatives possible

- $\Psi(t)$ increases starting from $t = 0$, so that $t = 0$ is the only solution of the equation (35); this means that the global minimum of $Q(u,t)$ is attained at $(1,0)$;
- $\Psi(t)$ decreases first, then grows to infinity, so that it exists a second solution of equation (35) corresponding to the global minimum.

To resolve this alternative it is necessary to study the vicinity of the point $t = 0$. Using the asymptotic

$$\frac{I_1(x)}{I_0(x)} = \frac{1}{2}x - \frac{1}{16}x^3 + \ldots \quad , \quad x \to 0 \tag{38}$$

we get

$$\Psi(t) = \left(\frac{B_4}{B_2^2} - 2\right)t^3 + \ldots \quad , \quad t \to 0 \quad . \tag{39}$$

It follows from the last formula that

- if $B_4 > 2B_2^2$, then $t = 0$ is the only solution for equation $\Psi(t) = 0$ and the global minimum of $Q(u,t)$ is attained at $u^{opt} = 1, t^{opt} = 0$; correspondingly $\alpha^{opt} = 0, \beta^{opt} = B_2 = \left\langle \left(H^{obs}\right)^2 \right\rangle$;
- if $B_4 < 2B_2^2$, then the equation $\Psi(t) = 0$ has non-zero solution t^{opt} and the global minimum of $Q(u,t)$ is attained at $\left(\sqrt{1+\left(t^{opt}\right)^2}, t^{opt}\right)$; parameters $\alpha^{opt}, \beta^{opt}$ are defined in this case as

$$\alpha^{opt} = \frac{\sqrt{B_2}}{H^{calc}} \frac{t^{opt}}{\sqrt{1+\left(t^{opt}\right)^2}} \quad , \quad \beta^{opt} = \frac{B_2}{1+\left(t^{opt}\right)^2} \quad . \tag{40}$$

4.5. Solution of the equation Ψ(t)=0

Due to non-monotonic behavior of function $\Psi(t)$ we used a zero-order method (without of the use of derivatives) to solve it. Nevertheless any other method can be used as well with necessary precautions.

5. Switching in the likelihood function

Let denote $J_j^{obs} = \left(H_j^{obs}\right)^2$, then the switching condition $B_4 < 2B_2^2$ may be written as

$$\left\langle \left(J^{obs} - \left\langle J^{obs} \right\rangle \right)^2 \right\rangle < \left\langle J^{obs} \right\rangle^2 \tag{41}$$

or

$$\sqrt{\text{variance}\left(J^{obs}\right)} < \text{mean}\left(J^{obs}\right) \quad . \tag{42}$$

The obtained condition means that the likelihood approach suggest to use a simple (Wilson) distribution

$$P^{theor}(H) = \frac{2H}{\left\langle \left(H^{obs}\right)^2 \right\rangle} \exp\left[-\frac{H^2}{\left\langle \left(H^{obs}\right)^2 \right\rangle} \right] \tag{43}$$

to describe the observed values instead of a more complicated distribution (17), if deviations of intensities J^{obs} from the mean are too large (more then the mean intensity value). To some extent we can interpret this as the likelihood function indicates that the calculated H^{calc} from a partial model is too unreliable and should not be used. Such switching is a rather typical feature of likelihood functions used for refinement or evaluation of atomic models in crystallography (Lunin, 1982; Lunin & Skovoroda, 1995; Lunin *et al.*, 2002).

Acknowledgments

This work was supported in part by grant 10-04-00254-a of Russian Foundation for Basic Research.

References

Adams, P. D., Pannu, N. S., Read, R. J. & Brünger, A. T. (1997). Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. *Proc Natl Acad. Sci. USA*, **94**, 5018-5023.

Afonine, P.V., Grosse-Kunstleve, R.W. & Adams, P.D. (2005). The Phenix refinement framework. *CCP4 Newsl.* **42**, contribution 8.

Bricogne, G. & Irwin, J. (1996). Maximum-likelihood structure refinement: theory and implementation within BUSTER + TNT. *Proceedings of the CCP4 Study Weekend*, Daresbury Laboratory, Warrington, England, 85-92.

Dauter, Z. (2003). Twinned crystals and anomalous phasing. *Acta Cryst.* D**59**, 2004-2016.

Lebedev, A.A., Vagin, A.A. & Murshudov, G.N. (2006). Intensity statistics in twinned crystals with

examples from the PDB. *Acta Cryst*. D**62**, 83-95.

Lunin, V.Y. (1982). The use of maximum likelihood approach to estimate phase errors in protein crystallography. *Preprint* (Russian), Pushchino, Russia. http://www.impb.ru/pdf/LCM_1982_3r.pdf

Lunin, V.Y. & Urzhumtsev, A.G. (1984). Improvement of Protein Phases by Coarse Model Modification. *Acta Cryst*., A**40**, 269-277.

Lunin, V.Y. & Skovoroda, T.P. (1995). R-free Likelihood-Based Estimates of Errors for Phases Calculated from Atomic Models. *Acta Cryst*. A**51**, 880-887.

Lunin, V.Y., Afonine P.V. & Urzhumtsev, A.G. (2002). Likelihood-based refinement. I. Irremovable model errors. *Acta Cryst*. A**58**, 270-282.

Lunin, V.Y., Lunina, N.L. & and Baumstark, M.W. (2007). Estimates of the twinning fraction for macromolecular crystals using statistical models accounting for experimental errors. *Acta Cryst*. , D**63**, 1129–1138

Murshudov, G.N., Vagin, A.A. & Dodson,E.J. (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Cryst*. D**53**, 240-255.

Pannu, N. S. & Read, R. J. (1996). Improved structure refinement through maximum likelihood. *Acta Cryst*. A**52**, 659-668.

Parsons, S. (2003). Introduction to twinning. *Acta Cryst*. D**59**, 1995-2003.

Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst*. A**42**, 140-149.

Read, R.J. (2001). Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Cryst*. D**57**, 1373-1382.

Srinivasan, R. & Parthasarathy, S. (1976). *Some statistical applications in X-ray crystallography*. Pergamon Press.

Urzhumtsev, A.G., Skovoroda, T.P. & Lunin, V.Y. (1996). A procedure compatible with X-PLOR for the calculation of electron-density maps weighted using an R-free-likelihood-based approach. *J.Appl.Cryst*., **29**, 741-744.

Yang, F., Dauter, Z. & Wlodawer, A. (2000). Effects of crystal twinning on the ability to solve a macromolecular structure using multiwavelength anomalous diffraction. *Acta Cryst*. D**56**, 959-964.

Yeates, T. O. (1988). Simple statistics for intensity data from twinned specimens. *Acta Cryst*. A**44**, 142-144.

Yeates, T. O. (1997). Detecting and overcoming crystal twinning. *Methods Enzymol*. **276**, 344-358.

Yeates, T.O. & Fam, B.C. (1999). Protein crystals and their evil twins. *Structure*, **7**, R25-R29.

TLS for dummies

Alexandre Urzhumtsev[a,b], Pavel V. Afonine[c] and Paul D. Adams[c,d]

[a]*IGBMC, CNRS-INSERM-UdS, 1 rue Laurent Fries, B.P.10142, 67404 Illkirch, France*
[b]*Physics Department, University of Nancy, B.P. 239, Faculté des Sciences et des Technologies, 54506 Vandoeuvre-lès-Nancy, France*
[c]*Lawrence Berkeley National Laboratory, One Cyclotron Road, BLDG 64R0121, Berkeley, CA 94720 USA.*
[d]*Department of Bioengineering, University of California Berkeley, Berkeley, CA 94720 USA.*

Correspondence email: sacha@igbmc.fr

Abstract

TLS model of a rigid-body harmonic displacement introduced in crystallography by Schomaker & Trueblood (1968) became a regular tool in macromolecular studies and is a part of most of modern refinement packages. There are a very large number of publications relevant to *TLS* and explaining its different aspects. However, these publications typically lack the details essential for understanding how the *TLS* model actually works, or contain too much of the formal mathematical details that are difficult to comprehend for readers without advanced mathematical background. In these notes we do not present any new development of the *TLS* model. Instead, we consider many simple examples that illustrate important features of the model. Using these examples, a general case is studied resulting in the widely known formulae. Simplified formulae are given for several special cases that may occur in macromolecular modeling and refinement. We believe that these notes may be useful for individuals who want to understand the basics of *TLS* modeling and not just use it as a "black box", as well as for crystallographic software developers wanting to implement some specific features described here.

Table of contents

1. Introduction

Crystallographic modeling of uncertainties in atomic positions uses different kinds of parameters. Some of them describe the same phenomenon with different accuracy; an example is isotropic or anisotropic atomic displacement parameters (*ADP*). Other parameters describe these uncertainties at different levels, such as individual atomic vibrations inside an atomic group or a thermal motion of this group as a whole (see Afonine et al., 2010 and references therein). In this sense, one should not consider an atomic group motion only as a way to reduce the number of parameters to work at low resolutions but as a way to better describe uncertainties in atomic positions.

Modeling a rigid group motion is based on the fact that any displacement of a rigid body may be considered as a superposition of a rotation around a given axis and a translation (see, for example, Goldstein, 1950). Eventually, these two motions may be correlated. When a rigid group oscillates, that is moves around its mean position, the term 'libration' is used instead of 'rotation' as it has been introduced in crystallography by Cruickshank (1956b). In what follows we stay within a harmonic model approximation that assumes small atomic displacements.

While some procedures to model harmonic rigid-group displacement and refine corresponding parameters have been suggested previously (see for example Pawley, 1963, 1964) nowadays the *TLS* model of a rigid-body harmonic displacement (Schomaker & Trueblood, 1968) is the mostly used one. Here *T*, *L* and *S* stand for 3 matrices describing translation, libration and their correlation (screw-rotation), respectively. It has been demonstrated that *TLS* modeling may provide reasonable results even for larger-scale vibrations (see for example Painter & Merritt, 2005). This ability to cover a broader vibrational range is considered a powerful feature of this modeling, although the results obtained at such inappropriate conditions should be interpreted with a caution.

There is a lot of literature about *TLS* modeling such as Johnson (1970, 1980), Scheringer (1973), Dunitz (1979), Stuart & Phillips (1985), Howlin et al. (1989), Tickle & Moss (1999), Winn et al. (2001), Painter & Merritt (2005, 2006a), Coppens (2006), Zucker et al. (2010) and references therein. The goal of this article is to give some technical details and practical computation schemes that are not available in the referenced above articles. Also, some specific cases are discussed (such as a *TLS* modeling with a fixed axis) that can be directly used in crystallographic structure refinement as an alternative to a traditional group ADP refinement. Differently from many articles on the subject, we try to keep all derivations and formulae at the basic level of the mathematics permitting most of readers to understand and reproduce them easily. We progress by short sections from easier specific cases to more complex and general ones.

2. Description of motion

2.1. Atomic displacement parameter (*ADP*)

A crystallographic atomic model represents not only time- and space-averaged positions $\mathbf{r}_n$ of atoms but also the uncertainties in these positions. These uncertainties result in blurring of atomic peaks in the Fourier maps and are characterised by atomic displacement parameters, *ADP*, also known as *B*-factors, isotropic or anisotropic. To simplify the analysis we suppose that all unit cells of the crystal have exactly the same structure and all uncertainties in atomic positions come from harmonic atomic motion only (as opposed to anharmonic large-scale motion resulting in distinct alternative conformations typically modeled using occupancies).

More formally, let's suppose that there is a Cartesian coordinate system with the origin **O** and the three orthonormal basis vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$ (the vectors are orthogonal to each other and are of a unit length). A position of an atom *n* at a moment *t* is defined by the coordinates (x_n, y_n, z_n) of $\mathbf{r}_n$ in this coordinate system (*e.g.* the PDB coordinates) and by the coordinates $(q_{nx}(t), q_{ny}(t), q_{nz}(t))$ of an instant deviation

$\mathbf{q}_n(t)$ from $\mathbf{r}_n$. The electron density in each point $\mathbf{r}_n + \mathbf{q}_n$ is proportional to the frequency (probability) $p_n(\mathbf{q}_n)$ characterizing the occurrence of the atom in this point. When (q_{nx}, q_{ny}, q_{nz}) are small, the logarithm of this probability, considered as a continuous smooth function, can be expanded into the Taylor series on these coordinates. When this expansion is done around a peak of the distribution, *i.e.* the most frequent position, the linear terms of this expansion are equal to zero as they correspond to the gradient of this function. (*Remark*: For a development around the mean position the linear term of the Taylor series does not necessarily vanish. For an atom in a single conformation it vanishes because this mean position can be expected to be close to or coincide with the most frequent value. For an atom in several alternative conformations such an expansion is done around each peak independently resulting in individual *ADP*s for each conformer). In a harmonic approximation, we have this expansion up to quadratic terms

$$p_n(\mathbf{q}) \approx \tilde{p}_n(\mathbf{q}) = \tag{2.1}$$

$$= \kappa_n \cdot \exp\left(\alpha_{nxx} q_x^2 + \alpha_{nyy} q_y^2 + \alpha_{nzz} q_z^2 + 2\alpha_{nxy} q_x q_y + 2\alpha_{nxz} q_x q_z + 2\alpha_{nyz} q_y q_z\right)$$

Non-harmonic approximations are discussed for example in Trueblood *et al*. (1996) Coppens (2006) and references therein. Expression (2.1) can be put into a standard form where the coefficients of the quadratic function of the atomic coordinates are presented by a symmetric matrix U_n^{-1}

$$\tilde{p}_n(\mathbf{q}) = (2\pi)^{-3/2} (\det U_n)^{-1/2} \exp\left(-\frac{1}{2}\mathbf{q}^\tau U_n^{-1} \mathbf{q}\right) \tag{2.2}$$

(see for example Trueblood *et al*., 1996). Here and in what follows the vector of instant deviation $\mathbf{q}$ is presented by a column vector of its coordinates $\begin{pmatrix} q_x \\ q_y \\ q_z \end{pmatrix}$ and τ stands for transposition, $(q_x, q_y, q_z) = \begin{pmatrix} q_x \\ q_y \\ q_z \end{pmatrix}^\tau$. The matrix U_n is defined as (see for example Cruickshank, 1956a and references therein)

$$U_n = \left\langle \mathbf{q}_n \mathbf{q}_n^\tau \right\rangle = \left\langle \begin{pmatrix} q_{nx} \\ q_{ny} \\ q_{nz} \end{pmatrix} \begin{pmatrix} q_{nx} & q_{ny} & q_{nz} \end{pmatrix} \right\rangle = \begin{pmatrix} \langle q_{nx}^2 \rangle & \langle q_{nx} q_{ny} \rangle & \langle q_{nx} q_{nz} \rangle \\ \langle q_{nx} q_{ny} \rangle & \langle q_{ny}^2 \rangle & \langle q_{ny} q_{nz} \rangle \\ \langle q_{nx} q_{nz} \rangle & \langle q_{ny} q_{nz} \rangle & \langle q_{nz}^2 \rangle \end{pmatrix} \tag{2.3}$$

In (2.3) the symbol $\langle\ \rangle$ means the time average and index *n* in U_n means that each atom is related to its matrix of displacements. This matrix is non-negative definite, *i.e.* $\left(\mathbf{q}^\tau U_n \mathbf{q}\right) \geq 0$ for any vector $\mathbf{q}$.

If a group of atoms moves together oscillating as a rigid body then atomic displacements $\mathbf{q}_n$ for different atoms are not independent but are expressed through some common parameters and the corresponding U_n also can be expressed as a combination of several common matrices. Here we are

going to show corresponding relations and their derivation. Obviously, such a description of an atomic group as a rigid body is only an approximation; this problem and ways to take next-level details into account were discussed for example by Johnson (1970b), Scheringer (1978b,c), Schomaker & Trueblood (1984), Dunitz *et al*. (1988), Bürgi (1989), Moore (2009) and others.

Obviously, a physical phenomenon – the probability of atomic displacement $\mathbf{q}_n$ in the crystal – is invariant with respect to the mathematical description used to describe it. This means that if we change the coordinate system, the coordinates of $\mathbf{q}_n$ change, so the matrix U_n should change accordingly, following some rules. Appendices A1 and A2 remind these basic rules, without going into advanced definitions of quadratic forms, tensors and other mathematical entities.

2.2. Rotation axes and their parameterisation

To define a displacement of a rigid body with respect to its original position, one needs to know a translation vector, a rotation axis and a rotation angle. An axis may be characterised by a unit vector ***l*** = (l_x, l_y, l_z) along it and by a point **w** = (w_x, w_y, w_z) that belongs to the axis.

In order to define a position of the rotation axis parallel, for example, to the coordinate axis **k** (figure 2.1a), two parameters are sufficient. For example these may be Cartesian coordinates of the point **w** = $(w_x, w_y, 0)$ in which this axis crosses the plane **Oij**. Alternatively, these may be polar coordinates R and α of this point. Note that with such a choice **Ow** by construction is normal to **k** and therefore **w** is the closest to the origin **O** among all points of the axis.

For an arbitrary axis ***l*** that passes through the origin **O** (figure 2.1b) two parameters are sufficient to define its direction. For example these may be two polar angles: the angle ψ between ***l*** and the **k** axis and the angle φ between **i** and the projection of ***l*** into the plane **Oij**.

To fully define an axis in an arbitrary position and not crossing the origin **O** (figure 2.1c), two parameters are required for its orientation, for example ψ and φ as above and two more for its position (***l*** defines unambiguously the plane normal to it and containing the origin **O**; in this plane

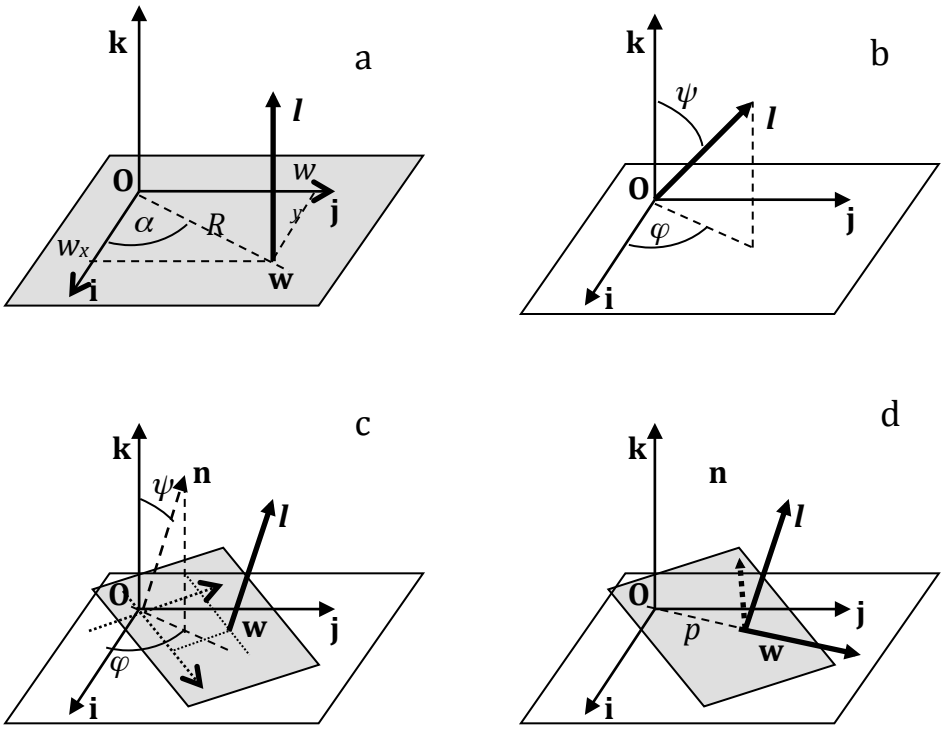


Figure 2.1. Schematic illustration of the definition of a rotation axis ***l***. (a) Axis parallel to **k**; its position is defined by 2 parameters, either by w_x and w_y, or by R and α. (b) Axis in an arbitrary orientation crossing the origin; its orientation is defined by 2 angles. (c) Axis in a general position; **n** is the vector parallel to ***l*** and crossing the origin; the plane normal to **n** and ***l*** is in grey. Similar to (b), two parameters are sufficient to define orientation of **n** and ***l***; to define the position of ***l*** two more parameters are required as in (a); they are coordinates of the intersection **w** of ***l*** with the 'grey' plane. (d) Normalized **Ow** and ***l*** are considered as the rotated **i** and **k**; p is the distance |**Ow**|.

the two coordinates, Cartesian or polar, of the point **w** in which ***l*** crosses the plane define its position). Obviously other points **w** may be chosen as discussed later.

Alternatively, the sufficiency of 4 parameters can be understood as following. Let **w** be a point on the rotation axis ***l*** such that the vectors ***l*** and **Ow** are orthogonal; in other words **w** is in the plane normal to ***l*** and crossing the origin (figure 2.1d). Then three Euler angles are sufficient to describe the orientation of **Ow** and ***l*** (for example a rotation of the base vectors so that rotated **i** coincides with **w** and rotated **k** coincides with ***l***) and the forth parameter is the shift p of the ***l*** axis along the direction **Ow**.

2.3. Rotation: linear approximation

When considering a libration of a body around a given axis, a displacement of each point may be expanded in series on the rotation angle δ. Eventually, high-order series may be considered; see for example, Johnson & Levy (1974), Johnson (1980), Coppens (2006) and references therein. Tickle & Moss (1999) mentioned that "*The harmonic model is applicable only if the motion is purely translational, but provided the libration amplitudes are not too large it is a good approximation*". Cruickshank (1956c) gave a value of 8° (0.14 radians) for the oscillation amplitude as a limit of this approximation.

In particular, working with small libration amplitudes means that a displacement towards the rotation axis is of a next order of magnitude than the displacement normal to the axis. For example, when the body is rotated around the coordinate axis **k** by an angle δ, the point **r** with the coordinates (1,0,0) gets the coordinates ($\cos\delta$, $\sin\delta$, 0). Replacing the exact displacement ($\cos\delta$-1, $\sin\delta$, 0) by its *linear approximation* **v** we neglect all terms starting from δ^2 in the Taylor series for cosine and sine, and the approximate coordinates of the shift are

$$(\cos\delta\,, \sin\delta, 0) - (1,0,0) \approx (0, \delta, 0) = \mathbf{v} \qquad (2.4)$$

More generally, for small angles δ any point positioned at the distance $R = 1$ from the rotation axis is displaced by a distance $d \approx R\delta = \delta$ (in radians). In these notes we use this parameter d instead of δ to define the rotation amplitude.

Also within a linear approximation, the point $(x_n,0,0)$ as well as $(x_n,0,z_n)$ are shifted by the vector $(0, x_n d, 0)$. Similarly, the point $(0,y_n,z_n)$ is shifted by $(-y_n d, 0, 0)$ and a general-position point $\mathbf{r}_n = (x_n,y_n,z_n)$ is shifted by

$$\mathbf{q}_n \approx \mathbf{v}_n = (-y_n d, x_n d, 0) = d\left[\mathbf{k} \times \mathbf{r}_n\right] \qquad (2.5)$$

where × is a vector cross product.

For a rotation around an arbitrary axis ***l*** crossing the origin, a linear approximation $\mathbf{v}_n$ to the displacement $\mathbf{q}_n$ of a point $\mathbf{r}_n$ may be expressed as

$$\mathbf{v}_n = d\left[\boldsymbol{l} \times \mathbf{r}_n\right] \qquad (2.6)$$

The latter is often presented in an equivalent form

$$\mathbf{v}_n = dA_n\boldsymbol{l} \qquad (2.7)$$

with the matrix

$$A_n = \begin{pmatrix} 0 & z_n & -y_n \\ -z_n & 0 & x_n \\ y_n & -x_n & 0 \end{pmatrix} \qquad (2.8)$$

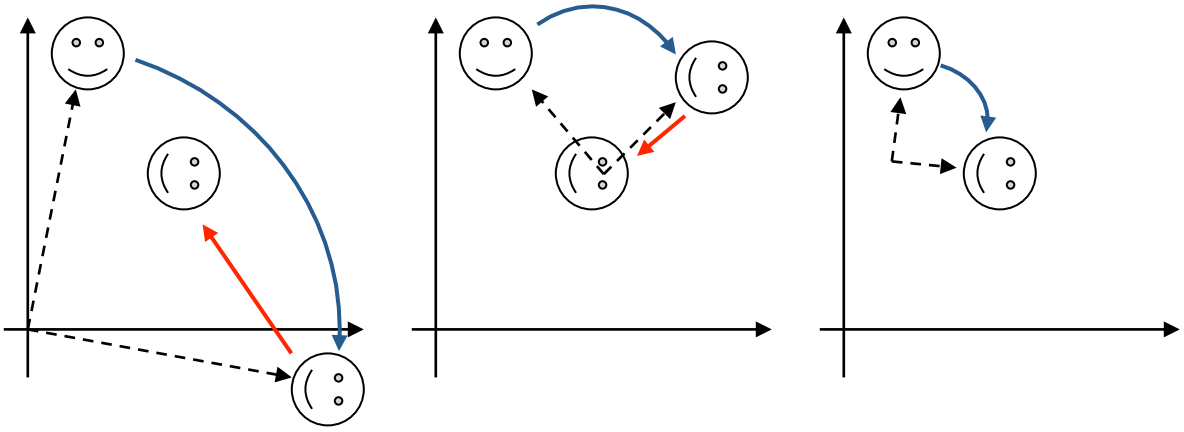


Figure 2.2. Three examples of possible combinations of a rotation (blue arrow) followed by a translation (red arrow) corresponding to the same transformation. The dashed lines show the rotation radius and indicate the rotation axis, always perpendicular to the plane of the page. Left image: the axis crosses the origin. Central image: the axis crosses the final object; corresponding translation is smaller than at the left image. Right image: the same transformation is presented as a pure rotation.

expressed through the Cartesian coordinates (x_n, y_n, z_n) of the point $\mathbf{r}_n$ in the coordinate system defined above (section 2.1). If $\boldsymbol{l}$ passes through a point $\mathbf{w} = (w_x, w_y, w_z)$ and not through the origin $\mathbf{O}$ (see for example section 2.2), vector $\mathbf{v}_n$ becomes

$$\mathbf{v}_n = d[\boldsymbol{l} \times (\mathbf{r}_n - \mathbf{w})] = d[\boldsymbol{l} \times \mathbf{r}_n] - d[\boldsymbol{l} \times \mathbf{w}] = dA_n \boldsymbol{l} - d\boldsymbol{l} \times \mathbf{w} \qquad (2.9)$$

Here $\boldsymbol{l} \times \mathbf{w}$ is a given vector, the same for all points of the oscillating rigid body; on the contrary $dA_n \boldsymbol{l}$ is point-dependent in the same fashion as A_n.

One can note that (2.9) presents a rotation around an axis in a general position as a rotation around an axis at the origin followed by a translation $-d\boldsymbol{l} \times \mathbf{w}$ common for all points. Inversely, a rotation around an axis at the origin (or at any other point) followed by a translation normal to this axis can be always presented as a pure rotation by the same angle. The position of this rotation axis is unique. Figure 2.2 inspired by figure 2 of the *TLSView Manual* (Merritt, pymmlib.sourceforge.net/tlsview/tlsview.html) illustrates this fact.

2.4. Choice of the point at the axis

Obviously, the shift (2.9) is independent of the choice of the point at the rotation axis $\boldsymbol{l}$. If we substitute a point $\mathbf{w}$ by another point $\mathbf{w}' = (w'_x, w'_y, w'_z)$ at the same axis, this gives:

$$\mathbf{v}'_n = d[\boldsymbol{l} \times (\mathbf{r}_n - \mathbf{w}')] = d[\boldsymbol{l} \times (\mathbf{r}_n - \mathbf{w} + \mathbf{w} - \mathbf{w}')] =$$

$$= d[\boldsymbol{l} \times (\mathbf{r}_n - \mathbf{w})] + d[\boldsymbol{l} \times (\mathbf{w} - \mathbf{w}')] = d[\boldsymbol{l} \times (\mathbf{r}_n - \mathbf{w})] = \mathbf{v}_n \qquad (2.10)$$

since $\mathbf{w} - \mathbf{w}'$ is collinear to $\boldsymbol{l}$.

2.5. Non-linear effects

The (omitted) second-order term in the development of the libration-based shift (2.5) into the Taylor series on the rotation angle corresponds to the displacement toward the rotation axis. This term is

responsible for two effects. The first one is the curvature of electron density for individual atoms that appears as "*banana-shaped contours*" (Howlin *et al.*, 1989); some authors use the term "*boomerang shape*". To the best of our knowledge such effects were not reported in practical macromolecular studies.

The second effect is a shrinking of an apparent bond length, the distance between the centers of two covalently linked atoms as we see them in the electron density maps (obviously, of a high enough resolution), in comparison with the actual bond length. This problem was discussed by Cruickshank (1956c, 1961), Busing & Levy (1964), Schomaker & Trueblood (1968), Scheringer (1972a,b), Stuart & Phillips (1985), Dunitz *et al*. (1988). Similarly, Scheringer (1978a) and Haestier *et al*. (2008) discusses a modification of apparent bond angles. Cruickshank (1956c) and Haneef *et al*. (1985) estimated the bond-length correcting value as 0.010-0.015 Å. Howlin *et al*. (1989) showed some larger values, up to 0.06 Å. Interestingly, Burns *et al.* in 1967 wrote that "*... it has become fairly common practice at the end of a molecular crystal structure determination to analyze the anisotropic temperature parameters on the assumption that the molecule is rigid. Often the purpose is no more than the correction of bond lengths...*". Such examples can be found in Becka & Cruickshank (1961), Birnbaum (1972), Downs *et al*. (1992), Steiner & Saenger (1993), Dunitz (1999).

In these notes we stay within a linear approximation (2.5-2.9).

3. Special case 1: rigid body translation

The simplest case is a pure translational motion, *i.e.* an oscillation of a rigid body without rotation compound. For such a displacement all points of the body are shifted by the same vector $\mathbf{q}_n = \mathbf{u}$. For this particular motion we introduce a special notation for the matrix U_n, the same for all points of the group:

$$U_n = T = \begin{pmatrix} \langle u_x^2 \rangle & \langle u_x u_y \rangle & \langle u_x u_z \rangle \\ \langle u_x u_y \rangle & \langle u_y^2 \rangle & \langle u_y u_z \rangle \\ \langle u_x u_z \rangle & \langle u_y u_z \rangle & \langle u_z^2 \rangle \end{pmatrix} \qquad (3.1)$$

By definition, T is symmetric and therefore is defined by 6 elements, 3 at the diagonal and 3 off-diagonal.

Also by definition T is non-negative definite, therefore it has three non-negative eigenvalues corresponding to three mutually orthogonal eigenvectors. In the basis composed of these normalised eigenvectors $\mathbf{i}_t$, $\mathbf{j}_t$, $\mathbf{k}_t$, matrix T becomes

$$T_t = \begin{pmatrix} \langle t_x^2 \rangle & 0 & 0 \\ 0 & \langle t_y^2 \rangle & 0 \\ 0 & 0 & \langle t_z^2 \rangle \end{pmatrix} \qquad (3.2)$$

with the eigenvalues at the diagonal. They are variances of the displacement along these three new axes. Zero off-diagonal elements mean that these displacements are non-correlated with each other and that they can be used as three new parameters of the problem. Note that for an isotropic translational displacement with $\langle t_x^2 \rangle = \langle t_y^2 \rangle = \langle t_z^2 \rangle$ the matrix T is diagonal in any basis.

Since both the basis $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ and the basis $(\mathbf{i}_t, \mathbf{j}_t, \mathbf{k}_t)$ are orthonormal, the transformation between them can be only a rotation. The corresponding transformation matrix R_t (see Appendix A1) can be defined by three rotation angles (see for example Urzhumtsev & Urzhumtseva (1997) for various

parameterisations used in crystallography for rotation matrices). Together with $\langle t_x^2 \rangle, \langle t_y^2 \rangle, \langle t_z^2 \rangle$ this makes the total number of parameters to be equal to 6 as above.

It follows from Appendix A2, coordinates (t_x, t_y, t_z) of a shift **u** in the basis $(\mathbf{i}_t, \mathbf{j}_t, \mathbf{k}_t)$ are linked to its coordinates (u_x, u_y, u_z) in the basis $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ by relation

$$\begin{pmatrix} u_x \\ u_y \\ u_z \end{pmatrix} = R_t \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} \tag{3.3}$$

and the matrix T in the initial base is

$$T = R_t \, T_t \, R_t^{\tau} \tag{3.4}$$

4. Special case 2: rotation axis parallel to k

4.1. Rotation around k axis

Now we consider a pure rotation of a rigid body with no translation component. For simplicity, we start from a rotation by a small angle δ around the axis **k**. As previously, if d_z is a random value for a shift of a point at a distance equal to 1 from **k** (section 2.3), its probability distribution defines a corresponding shift $\mathbf{q}_n \approx \mathbf{v}_n = (-y_n d_z, x_n d_z, 0)$ of a point $\mathbf{r}_n = (x_n, y_n, z_n)$ giving its matrix U_n (2.3) as

$$U_n = \begin{pmatrix} \langle q_{nx}^2 \rangle & \langle q_{nx} q_{ny} \rangle & \langle q_{nx} q_{nz} \rangle \\ \langle q_{nx} q_{ny} \rangle & \langle q_{ny}^2 \rangle & \langle q_{ny} q_{nz} \rangle \\ \langle q_{nx} q_{nz} \rangle & \langle q_{ny} q_{nz} \rangle & \langle q_{nz}^2 \rangle \end{pmatrix} = \begin{pmatrix} y_n^2 \langle d_z^2 \rangle & -x_n y_n \langle d_z^2 \rangle & 0 \\ -x_n y_n \langle d_z^2 \rangle & x_n^2 \langle d_z^2 \rangle & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$= \langle d_z^2 \rangle \begin{pmatrix} y_n^2 & -x_n y_n & 0 \\ -x_n y_n & x_n^2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{4.1}$$

The same results can be also obtained as

$$U_n = \langle \mathbf{q}_n \mathbf{q}_n^{\tau} \rangle = \langle (d_z A_n \mathbf{k})(d_z A_n \mathbf{k})^{\tau} \rangle = \langle d_z^2 A_n \mathbf{k} \mathbf{k}^{\tau} A_n^{\tau} \rangle = \langle d_z^2 \rangle A_n \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} A_n^{\tau} \tag{4.2}$$

For any atom in the rigid group the elements of the matrix in the right hand of expression (4.1) are actual atomic coordinates (as found in PDB file, for example; a better choice will be discussed below) and the random displacement of the rigid group is presented by a common factor $\langle d_z^2 \rangle$ that shows the amplitude of librations.

4.2. Roation axis parallel to k

When the rotation axis is parallel to **k** and passes through a point $\mathbf{w}^k = \left(w_x^k, w_y^k, 0\right)$ different from the origin, the corresponding shift (section 2.3) is

$$\mathbf{q}_n \approx \mathbf{v}_n = d_z A_n \mathbf{k} - d_z \mathbf{k} \times \mathbf{w} = d_z A_n \mathbf{k} - d_z W_k \mathbf{k} = d_z A_n \mathbf{k} + d_z W_k^\tau \mathbf{k} \tag{4.3}$$

with

$$W_k = \begin{pmatrix} 0 & 0 & -w_y^k \\ 0 & 0 & w_x^k \\ w_y^k & -w_x^k & 0 \end{pmatrix} \tag{4.4}$$

Here matrix W_k is introduced similarly to matrix A_n in (2.6)-(2.7), section 2.3. Similarly to (4.2), expression (4.3) leads to

$$U_n = \tag{4.5}$$

$$= \left\langle d^2 \right\rangle W_k^\tau \left(\mathbf{k}\mathbf{k}^\tau\right) W_k + \left\langle d^2 \right\rangle A_n \left(\mathbf{k}\mathbf{k}^\tau\right) A_n^\tau + \left\langle d^2 \right\rangle A_n \left(\mathbf{k}\mathbf{k}^\tau\right) W_k + \left\langle d^2 \right\rangle W_k^\tau \left(\mathbf{k}\mathbf{k}^\tau\right) A_n^\tau$$

Here

$$\mathbf{k}\mathbf{k}^\tau = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{4.6}$$

and

$$\mathbf{k}\mathbf{k}^\tau W_k = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ w_y^k & -w_x^k & 0 \end{pmatrix}, \qquad W_k^\tau \mathbf{k}\mathbf{k}^\tau = \begin{pmatrix} 0 & 0 & w_y^k \\ 0 & 0 & -w_x^k \\ 0 & 0 & 0 \end{pmatrix} \tag{4.7}$$

$$W_k^\tau \mathbf{k}\mathbf{k}^\tau W_k = \begin{pmatrix} \left(w_y^k\right)^2 & -w_x^k w_y^k & 0 \\ -w_x^k w_y^k & \left(w_x^k\right)^2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{4.8}$$

The first term in (4.5) is independent of the point $\mathbf{w}^k$ and corresponds to an apparent translation even when it was no translation in the initial description of the motion (see section 2.3 for similar examples).

5. Special case 3: rotation around k correlated with translation

5.1. Several examples

When in addition to the rotation around **k** (section 4.1) the body is undergoing a translation as described in section 3, the total distribution of the displacement of each atom depends on the correlation between rotation and translation, as well as on the direction of the translation. We start from a couple of simple illustrations. The displacement distribution can be represented by a surface on which the points have the same probability distribution; for harmonic oscillations this surface is called a

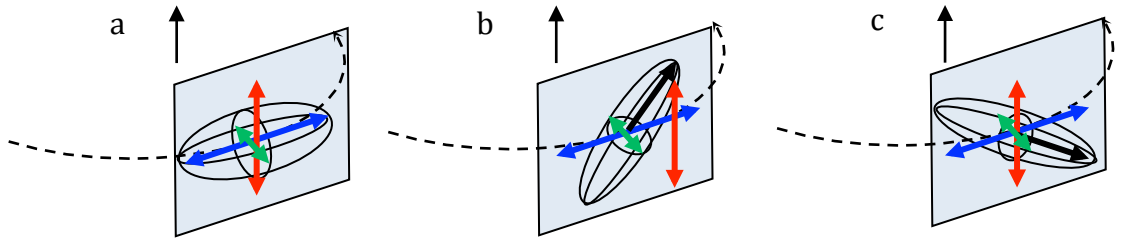


Figure 5.1. Schematic illustration of a libration around the vertical axis (an arrow at the center of circle) and a random translation. The motions are uncorrelated in the rotation plane (horizontal). Ellipses show surfaces of atomic displacement with the same probability. (a) The rotational displacement (blue) and translational displacement along the rotation axis (red) are uncorrelated; main elliptic axis is horizontal. (b) The displacements are positively correlated; the main elliptic axis (black bold) is in the plane formed by blue and red arrows and follows a right-hand helix. (c) The displacements are negatively correlated; similar to (b) but the main elliptic axis follows a left-hand helix. See Section 5.1 for more detail.

thermal ellipsoid.

Let's suppose that the translational displacement is isotropic. When rotation and translation are not correlated, the thermal motion ellipsoid in our example has one axis parallel to the axis **k** and two other axes normal to it (figure 5.1a).

When the displacement $(-y_n d_z, x_n d_z, 0)$ of a point (x_n, y_n, z_n) due to rotation is coupled (correlated) with its shift $(0,0,s_z d_z)$ *along the rotation axis*, the total linear displacement $(-y_n d_z, x_n d_z, s_z d_z)$ approximates an arch of a helix. The parameter s_z defines the slope of the trajectory with respect to the axis (figures 5.1b, c).

A superposition of a rotation and a correlated displacement in the direction *normal to the rotation axis* generates an apparent rotation axis shifted with respect to the original one, as discussed above in section 2.3. As an example (figure 5.2), let's consider again a rotation around the **k** axis, that generates

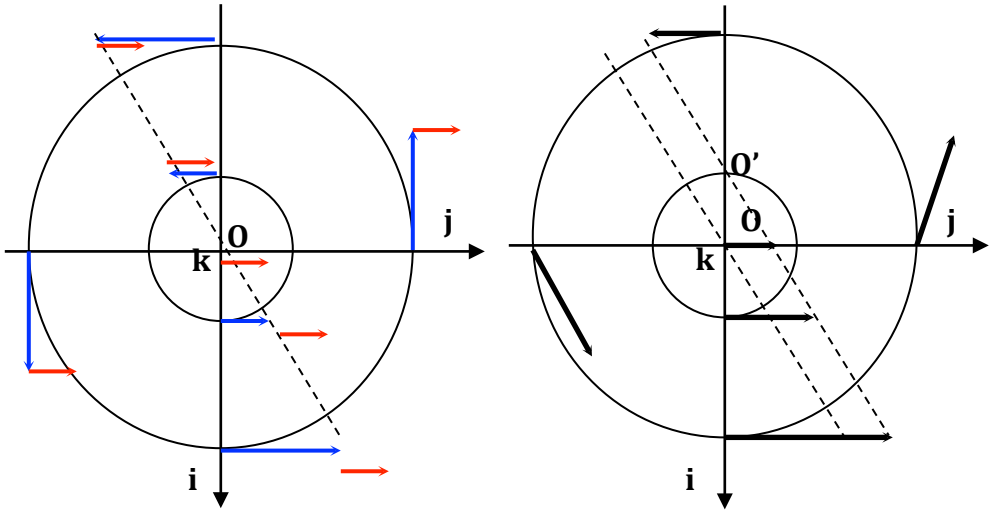


Figure 5.2. Schematic illustration of a correlation of a rotation around an axis **k** normal to the projection shown and a translation in the direction **j**. (a) displacements due to rotation (linear approximation, blue arrows) and translation (red arrows) are shown for several points; (b) total displacement (black arrows) and the shift of the rotation axis to its apparent position **O'**. See section 5.1 for more detail.

the shift $(-y_n d_z, x_n d_z, 0)$ for the points (x_n, y_n, z_n). If additionally we add a translation $(0, s_y d_z, 0)$, where s_y is some number, the total shift becomes $(-y_n d_z, (x_n + s_y) d_z, 0)$. This corresponds to a rotation around the axis parallel to **k** and crossing the point $(-s_y, 0, 0)$.

5.2. Screw axes along k

Let us analyse a screw rotation around an axis **k** more formally. As discussed in the previous section, the corresponding displacement of a point (x_n, y_n, z_n) is

$$\mathbf{q}_n = (-y_n d_z, x_n d_z, s_z d_z) = d_z A_n \mathbf{k} + s_z d_z \mathbf{k} \qquad (5.1)$$

The component $s_z d_z \mathbf{k}$ is the same for all points of the rigid body and it is a translation component of the group. Accordingly (2.3) the matrix U_n is

$$U_n = s_z^2 \langle d_z^2 \rangle (\mathbf{k}\mathbf{k}^\tau) + A_n \langle d_z^2 \rangle (\mathbf{k}\mathbf{k}^\tau) A_n^\tau + A_n (s_z \langle d_z^2 \rangle \mathbf{k}\mathbf{k}^\tau) + (s_z \langle d_z^2 \rangle \mathbf{k}\mathbf{k}^\tau) A_n^\tau \qquad (5.2)$$

One may note a similarity between (5.2) and (4.5). The first term in (5.2) is independent of the atomic coordinates and stands for the translation of the group. The second term depends, through matrix A_n, on the atomic coordinates quadratically and corresponds to the group rotation. Two last terms in (5.2) depend on atomic coordinates linearly and are due to the screw component. Following Schomaker & Trueblood (1968) we associate these terms with the matrices T, L and S that we define here as

$$T = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & s_z^2 \langle d_z^2 \rangle \end{pmatrix} \qquad (5.3)$$

$$L = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \langle d_z^2 \rangle \end{pmatrix} \qquad (5.4)$$

$$S = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & s_z \langle d_z^2 \rangle \end{pmatrix} \qquad (5.5)$$

With these matrices,

$$U_n = T + A_n L A_n^\tau + A_n S + (A_n S)^\tau \qquad (5.6)$$

5.3. *TLS* presentation

Now let's generalise the examples of sections 5.1 and 5.2. We keep the same notation and use **u** for translational displacement (section 3) and $\mathbf{v}_n$ for the displacement due to libration, always in a linear approximation (section 2.3).

For an atom *n* presented by its Cartesian coordinates (x_n, y_n, z_n) in the same basis as above the total

displacement vector

$$\mathbf{q}_n = \mathbf{u} + \mathbf{v}_n = \mathbf{u} + d_z A_n \mathbf{k} \tag{5.7}$$

has the coordinates

$$\begin{pmatrix} q_{nx} \\ q_{ny} \\ q_{nz} \end{pmatrix} = \begin{pmatrix} u_x \\ u_y \\ u_z \end{pmatrix} + \begin{pmatrix} -y_n d_z \\ x_n d_z \\ 0 \end{pmatrix} = \begin{pmatrix} u_x - y_n d_z \\ u_y + x_n d_z \\ u_z \end{pmatrix} \tag{5.8}$$

where d_z, as previously, defines the linear approximation $(0, d_z, 0)$ to the displacement of the point (1,0,0). Note that it follows from section 5.2, translation vector $\mathbf{u}$ may include the screw component and therefore be correlated with the rotation. Components of the matrix U_n (2.3) in the same coordinate system (**i**, **j**, **k**) are:

$$U_{nxx} = \left\langle \left(u_x - y_n d_z\right)^2 \right\rangle = \left\langle u_x^2 + y_n^2 d_z^2 - 2 y_n u_x d_z \right\rangle = \left\langle u_x^2 \right\rangle + y_n^2 \left\langle d_z^2 \right\rangle - 2 y_n \left\langle u_x d_z \right\rangle$$

$$U_{nyy} = \left\langle \left(u_y + x_n d_z\right)^2 \right\rangle = \left\langle u_y^2 + x_n^2 d_z^2 + 2 x_n u_y d_z \right\rangle = \left\langle u_y^2 \right\rangle + x_n^2 \left\langle d_z^2 \right\rangle + 2 x_n \left\langle u_y d_z \right\rangle$$

$$U_{nzz} = \left\langle u_z^2 \right\rangle$$

$$U_{nxy} = U_{nyx} = \left\langle \left(u_x - y_n d_z\right)\left(u_y + x_n d_z\right) \right\rangle = \left\langle u_x u_y - y_n u_y d_z + x_n u_x d_z - x_n y_n d_z^2 \right\rangle$$

$$= \left\langle u_x u_y \right\rangle - y_n \left\langle u_y d_z \right\rangle + x \left\langle u_x d_z \right\rangle - x_n y_n \left\langle d_z^2 \right\rangle$$

$$U_{nxz} = U_{nzx} = \left\langle \left(u_x - y_n d_z\right) u_z \right\rangle = \left\langle u_x u_z - y_n u_z d_z \right\rangle = \left\langle u_x u_z \right\rangle - y_n \left\langle u_z d_z \right\rangle$$

$$U_{nyz} = U_{nzy} = \left\langle \left(u_y + x_n d_z\right) u_z \right\rangle = \left\langle u_y u_z + x_n u_z d_z \right\rangle = \left\langle u_y u_z \right\rangle + x_n \left\langle u_z d_z \right\rangle \tag{5.9}$$

Similarly to section 5.2, we can obtain the *TLS* presentation

$$U_n = T + A_n L A_n^\tau + A_n S + S^\tau A_n^\tau \tag{5.10}$$

similar to (5.6). Currently,

$$T = U_T = \begin{pmatrix} \left\langle u_x^2 \right\rangle & \left\langle u_x u_y \right\rangle & \left\langle u_x u_z \right\rangle \\ \left\langle u_x u_y \right\rangle & \left\langle u_y^2 \right\rangle & \left\langle u_y u_z \right\rangle \\ \left\langle u_x u_z \right\rangle & \left\langle u_y u_z \right\rangle & \left\langle u_z^2 \right\rangle \end{pmatrix} \tag{5.11}$$

$$L = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \left\langle d_z^2 \right\rangle \end{pmatrix} \tag{5.12}$$

$$S = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \langle u_x d_z \rangle & \langle u_y d_z \rangle & \langle u_z d_z \rangle \end{pmatrix} \qquad (5.13)$$

and

$$A_n = \begin{pmatrix} 0 & z_n & -y_n \\ -z_n & 0 & x_n \\ y_n & -x_n & 0 \end{pmatrix} \qquad (5.14)$$

In presentation (5.10-5.14) matrices *T*, *L* and *S* are the same for all points of the rigid body while A_n is expressed through the point coordinates. This presentation can be obtained either by decomposition of (5.8) or by applying (2.3) to (5.7).

Expressions (5.11-5.13) show the 10 parameters common for the rigid group: 6, 1 and 3 parameters associated with *T*, *L* and *S*, respectively. Together with the matrices A_n, they fully define U_n for all atoms in the rigid group.

5.4. Origin shift

Obviously, matrices U_n should not depend on the choice of the origin of the coordinate system while matrices A_n do. This means that *T*, *L*, *S* or at least some of them, vary with the origin. Inversely, this means that some combinations of *T*, *L* or *S* may correspond to the same U_n but expressed in coordinate systems with different origins. We will demonstrate this relation that is very important for further analysis.

In a new coordinate system with the origin shifted by vector $\mathbf{p} = (p_x, p_y, p_z)$

$$\mathbf{O}' = \mathbf{O} + \mathbf{p} \qquad (5.15)$$

the new coordinates are $x_n - p_x, y_n - p_y, z_n - p_z$ defining the matrix

$$A'_n = \begin{pmatrix} 0 & z_n - p_z & -(y_n - p_y) \\ -(z_n - p_z) & 0 & x_n - p_x \\ y_n - p_y & -(x_n - p_x) & 0 \end{pmatrix} =$$

$$= \begin{pmatrix} 0 & z_n & -y_n \\ -z_n & 0 & x \\ y_n & -x_n & 0 \end{pmatrix} - \begin{pmatrix} 0 & p_z & -p_y \\ -p_z & 0 & p_x \\ p_y & -p_x & 0 \end{pmatrix} = A_n - P \qquad (5.16)$$

(we define the shift vector in the opposite way as Tickle & Moss (1999) do). Matrix *P* in (5.16) is the same for all points of the group. Accordingly to (5.10) matrix U_n becomes

$$U_n = T + (A'_n + P) L (A'_n + P)^{\tau} + (A'_n + P)\ S + S^{\tau} (A'_n + P)^{\tau}$$

$$= T + \left(A'_n L A'^{\tau}_n + A'_n L P^{\tau} + P L A'^{\tau}_n + P L P^{\tau} \right) + \left(A'_n S + P S \right) + \left(S^{\tau} A'^{\tau}_n + S^{\tau} P^{\tau} \right)$$

$$= \left(T + P L P^{\tau} + P S + S^{\tau} P^{\tau} \right) + A'_n L A'^{\tau}_n + A'_n \left(S + L P^{\tau} \right) + \left(S^{\tau} + P L^{\tau} \right) A'^{\tau}_n$$

$$= T' + A'_n L' A'^{\tau}_n + A'_n S' + S'^{\tau} A'^{\tau}_n = U'_n \qquad (5.17)$$

(at the last transition we substituted $P L A'^{\tau}_n$ by $P L^{\tau} A'^{\tau}_n$ using the symmetry $L = L^{\tau}$). Comparison with (5.10) defines new matrices T', L' and S' as

$$T' = T + \left(P L P^{\tau} + P S + S^{\tau} P^{\tau} \right)$$

$$L' = L \qquad (5.18)$$

$$S' = S + L P^{\tau}$$

It is very important that the expressions (5.18) were obtained for a general case of matrices *T*, *L* and *S* in (5.10) with no use of their specific form (5.11-5.13). A simplest example illustrating (5.18) is presented below.

Let's suppose that a rigid group oscillates around the axis **k**. In this case the only non-zero matrix is *L* (5.6) while *T* and *S* are zero. For the point **M** = (0,0,0) its $A_n = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ giving zero matrix U_n (a point sits at the rotation axis). Now let's choose another coordinate system shifting the origin by **p** = (1,0,0). The new coordinates of the point **M** are (-1,0,0), its new matrix $A'_n = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$ and matrix $P = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}$. New matrices are

$$S' = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \langle d_z^2 \rangle \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \langle d_z^2 \rangle & 0 \end{pmatrix}$$

$$A'_n S' = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \langle d_z^2 \rangle & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -\langle d_z^2 \rangle & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$T' = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \langle d_z^2 \rangle \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \langle d_z^2 \rangle & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$A'_n L' A'^{\tau}_n = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \langle d_z^2 \rangle \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \langle d_z^2 \rangle & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

New matrix $U'_n = T' + A'_n L' A'^{\tau}_n + A'_n S' + S'^{\tau} A'^{\tau}_n$ is a zero matrix, as it should be.

5.5. Search for the apparent rotation axis

Applying (5.18) to the special case (5.11-5.13) of a rotation around **k** gives

$$LP^{\tau} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -p_y \langle d_z^2 \rangle & p_x \langle d_z^2 \rangle & 0 \end{pmatrix} \tag{5.19}$$

$$S + LP^{\tau} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \langle u_x d_z \rangle - p_y \langle d_z^2 \rangle & \langle u_y d_z \rangle + p_x \langle d_z^2 \rangle & \langle u_z d_z \rangle \end{pmatrix} \tag{5.20}$$

$$PLP^{\tau} = \begin{pmatrix} p_y^2 \langle d_z^2 \rangle & -p_x p_y \langle d_z^2 \rangle & 0 \\ -p_x p_y \langle d_z^2 \rangle & p_x^2 \langle d_z^2 \rangle & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{5.21}$$

$$PS = \begin{pmatrix} -p_y \langle u_x d_z \rangle & -p_y \langle u_y d_z \rangle & -p_y \langle u_z d_z \rangle \\ p_x \langle u_x d_z \rangle & p_x \langle u_y d_z \rangle & p_x \langle u_z d_z \rangle \\ 0 & 0 & 0 \end{pmatrix} \tag{5.22}$$

$$T + PLP^{\tau} + PS + S^{\tau} P^{\tau} = \tag{5.23}$$

$$= \begin{pmatrix} p_y^2 \langle d_z^2 \rangle - 2p_y \langle u_x d_z \rangle + \langle u_x^2 \rangle & \left\{ \begin{matrix} p_x \langle u_x d_z \rangle - p_y \langle u_y d_z \rangle \\ p_x p_y \langle d_z^2 \rangle - \langle u_x u_y \rangle \end{matrix} \right\} & -p_y \langle u_z d_z \rangle + \langle u_x u_z \rangle \\ \left\{ \begin{matrix} p_x \langle u_x d_z \rangle - p_y \langle u_y d_z \rangle \\ -p_x p_y \langle d_z^2 \rangle + \langle u_x u_y \rangle \end{matrix} \right\} & p_x^2 \langle d_z^2 \rangle + 2p_x \langle u_y d_z \rangle + \langle u_y^2 \rangle & p_x \langle u_z d_z \rangle + \langle u_y u_z \rangle \\ -p_y \langle u_z d_z \rangle + \langle u_x u_z \rangle & p_x \langle u_z d_z \rangle + \langle u_y u_z \rangle & \langle u_z^2 \rangle \end{pmatrix}$$

Expression (5.23) shows existence of a special origin

$$p_y = \frac{\langle u_x d_z \rangle}{\langle d_z^2 \rangle}, \quad p_x = -\frac{\langle u_y d_z \rangle}{\langle d_z^2 \rangle} \qquad . \qquad (5.24)$$

that minimises the diagonal elements of T' making this matrix equal to

$$T' = \begin{pmatrix} \langle u_x^2 \rangle - \frac{\langle u_x d_z \rangle^2}{\langle d_z^2 \rangle} & \langle u_x u_y \rangle - \frac{\langle u_x d \rangle \langle u_y d_z \rangle}{\langle d_z^2 \rangle} & \langle u_x u_z \rangle - \frac{\langle u_x d_z \rangle \langle u_z d_z \rangle}{\langle d_z^2 \rangle} \\ \langle u_x u_y \rangle - \frac{\langle u_x d_z \rangle \langle u_y d_z \rangle}{\langle d_z^2 \rangle} & \langle u_y^2 \rangle - \frac{\langle u_y d_z \rangle^2}{\langle d_z^2 \rangle} & \langle u_y u_z \rangle - \frac{\langle u_y d_z \rangle \langle u_z d_z \rangle}{\langle d_z^2 \rangle} \\ \langle u_x u_z \rangle - \frac{\langle u_x d_z \rangle \langle u_z d_z \rangle}{\langle d_z^2 \rangle} & \langle u_y u_z \rangle - \frac{\langle u_y d_z \rangle \langle u_z d_z \rangle}{\langle d_z^2 \rangle} & \langle u_z^2 \rangle \end{pmatrix} \qquad (5.25)$$

The minimisation of all diagonal elements T_{xx}, T_{yy}, T_{zz} at a time can be reformulated as the minimisation of the trace of the matrix, $tr(T) = T_{xx} + T_{yy} + T_{zz} \rightarrow \min\limits_{p_x, p_y, p_z}$, which stays non-negative due to the Cauchy-Schwarz inequality. In fact, this minimisation means that the new origin is at the apparent rotation axis, as discussed in the examples of sections 2.3, 4.2 and 5.1 (see also Schomaker & Trueblood, 1968; Pawley, 1970; Tickle & Moss, 1999). This is also confirmed by (5.20) that becomes

$$S' = S + LH^{\tau} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \langle u_z d_z \rangle \end{pmatrix} \qquad (5.26)$$

showing no correlation between rotation-translation displacements in the plane normal to the rotation axis (we remind the reader that in this example case it is the axis $\mathbf{k}$). The results above are independent of the choice of p_z (a shift along the rotation axis).

Relocation of the translation component by including it into the displacement of the rotation axis is opposite to an operation discussed in section 3: an apparent translation component for the axes different from the coordinate ones. Comparison of (5.24) with (4.4) shows their similarity in determination of the position of the rotation axis.

5.6. Parameters with a physical meaning

Following from section 3, one may define the elements of *T* through its eigenvalues (uncorrelated translations) $\langle t_x^2 \rangle, \langle t_y^2 \rangle, \langle t_z^2 \rangle$ and the rotation matrix R_t. One may note also that due to relation (3.3)

$$\begin{pmatrix} \langle u_x d_z \rangle \\ \langle u_y d_z \rangle \\ \langle u_z d_z \rangle \end{pmatrix} = R_t \begin{pmatrix} \langle t_x d_z \rangle \\ \langle t_y d_z \rangle \\ \langle t_z d_z \rangle \end{pmatrix} \qquad (5.27)$$

where $\langle t_x d_z \rangle, \langle t_y d_z \rangle, \langle t_z d_z \rangle$ describe correlation of mutually uncorrelated random displacements (t_x, t_y,

t_z) with the displacement due to libration around the axis **k** and the matrix R_t is the same as before.

As a next step, one may introduce the correlations

$$-1 \le c_x = \frac{\langle t_x d_z \rangle}{\sqrt{\langle t_x^2 \rangle}\sqrt{\langle d_z^2 \rangle}} \le 1\,,\ -1 \le c_y = \frac{\langle t_y d_z \rangle}{\sqrt{\langle t_y^2 \rangle}\sqrt{\langle d_z^2 \rangle}} \le 1,\ -1 \le c_z = \frac{\langle t_z d_z \rangle}{\sqrt{\langle t_z^2 \rangle}\sqrt{\langle d_z^2 \rangle}} \le 1 \tag{5.28}$$

as independent parameters instead of $\langle u_x d_z \rangle, \langle u_y d_z \rangle, \langle u_z d_z \rangle$, resulting in another set of parameters: $\langle t_x^2 \rangle, \langle t_y^2 \rangle, \langle t_z^2 \rangle, \langle d_z^2 \rangle$, c_x, c_y, c_z and 3 mutually orthogonal directions of uncorrelated translations described through R_t by three Euler angles; this makes ten parameters in total as previously. Such a choice has an advantage that all corresponding parameters have a clear physical meaning.

6. Special case 4: three rotation axes parallel to ijk

6.1. Uncorrelated pure rotations

Now we will extend the analysis done in section 4.2. When a rotation axis is parallel to **i** or **j** instead of **k**, the resulting matrix has always the form (4.5) in which W_k and $\mathbf{k}\mathbf{k}^\tau$ (4.4 and 4.6) are replaced by

$$W_i = \begin{pmatrix} 0 & w_z^i & -w_y^i \\ -w_z^i & 0 & 0 \\ w_y^i & 0 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{i}\mathbf{i}^\tau = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{6.1}$$

for rotations around **i** or by

$$W_j = \begin{pmatrix} 0 & w_z^j & 0 \\ -w_z^j & 0 & w_x^j \\ 0 & -w_x^j & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{j}\mathbf{j}^\tau = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{6.2}$$

for rotations around **j**, respectively.

When three rotations around the axes parallel to **i**, **j** and **k** with the corresponding amplitudes d_x, d_y and d_z are executed simultaneously the resulting shift is

$$\mathbf{q}_n \approx \mathbf{v}_n = \left(d_x A_n \mathbf{i} + d_x W_i^\tau \mathbf{i}\right) + \left(d_y A_n \mathbf{j} + d_y W_j^\tau \mathbf{j}\right) + \left(d_z A_n \mathbf{k} + d_z W_k^\tau \mathbf{k}\right) \tag{6.3}$$

For uncorrelated rotations, *i.e.* such that

$$\langle d_x d_y \rangle = \langle d_x d_z \rangle = \langle d_y d_z \rangle = 0, \tag{6.4}$$

a calculation similar to (4.3)-(4.8) gives the matrix U_n in the form (5.10) with

$$T = \tag{6.5}$$

$$= \begin{pmatrix} \left(w_z^j\right)^2\left\langle d_y^2\right\rangle + \left(w_y^k\right)^2\left\langle d_z^2\right\rangle & -w_x^k w_y^k\left\langle d_z^2\right\rangle & -w_x^j w_z^j\left\langle d_y^2\right\rangle \\ -w_x^k w_y^k\left\langle d_z^2\right\rangle & \left(w_z^i\right)^2\left\langle d_x^2\right\rangle + \left(w_x^k\right)^2\left\langle d_z^2\right\rangle & -w_y^i w_z^i\left\langle d_x^2\right\rangle \\ -w_x^j w_z^j\left\langle d_y^2\right\rangle & -w_y^i w_z^i\left\langle d_x^2\right\rangle & \left(w_y^i\right)^2\left\langle d_x^2\right\rangle + \left(w_x^j\right)^2\left\langle d_y^2\right\rangle \end{pmatrix}$$

$$L = \begin{pmatrix} \left\langle d_x^2\right\rangle & 0 & 0 \\ 0 & \left\langle d_y^2\right\rangle & 0 \\ 0 & 0 & \left\langle d_z^2\right\rangle \end{pmatrix} \tag{6.6}$$

$$S = \begin{pmatrix} 0 & w_z^i\left\langle d_x^2\right\rangle & -w_y^i\left\langle d_x^2\right\rangle \\ -w_z^j\left\langle d_y^2\right\rangle & 0 & w_x^j\left\langle d_y^2\right\rangle \\ w_y^k\left\langle d_z^2\right\rangle & -w_x^k\left\langle d_z^2\right\rangle & 0 \end{pmatrix} \tag{6.7}$$

As previously, T corresponds to the apparent translation term, the same for all points.

6.2. Screw rotations around the coordinate axes

Following from the calculations in section 5.2, one can derive that three simultaneous uncorrelated rotations around the three coordinate axes with the amplitudes d_x, d_y, d_z and screw components s_x, s_y, s_z give U_n in the form (5.10) with the matrices T, L and S:

$$T = \begin{pmatrix} s_x^2\left\langle d_x^2\right\rangle & 0 & 0 \\ 0 & s_y^2\left\langle d_y^2\right\rangle & 0 \\ 0 & 0 & s_z^2\left\langle d_z^2\right\rangle \end{pmatrix} \tag{6.8}$$

$$L = \begin{pmatrix} \left\langle d_x^2\right\rangle & 0 & 0 \\ 0 & \left\langle d_y^2\right\rangle & 0 \\ 0 & 0 & \left\langle d_z^2\right\rangle \end{pmatrix} \tag{6.9}$$

$$S = \begin{pmatrix} s_x\left\langle d_x^2\right\rangle & 0 & 0 \\ 0 & s_y\left\langle d_y^2\right\rangle & 0 \\ 0 & 0 & s_z\left\langle d_z^2\right\rangle \end{pmatrix} \tag{6.10}$$

It may be useful to compare (6.8)-(6.10) with (6.5)-(6.7).

7. Rotation around an axis in a general position

7.1. Rotation around a fixed bond

A libration of an atomic group around a given axis plays a special role in macromolecular modeling where dihedral angles are relatively flexible compared to bond angles and lengths. This may be a libration of a peptide side chain around $C_\alpha C_\beta$ bond (see figure 7.1 for illustration) or, in general, a

libration of an atomic group or a domain around a bond between two given atoms. Detailed studies of a rotation around a bond can be found in Prince & Finger (1972), Dunitz & White (1973), Sygusch (1976) and Schomaker & Trueblood (1998).

In this section, let's consider a rotation around the vector **g** between two fixed points $\mathbf{G}_1 = (G_{1x}, G_{1y}, G_{1z})$ and $\mathbf{G}_2 = (G_{2x}, G_{2y}, G_{2z})$, thus **g** being fixed as well. In figure 7.2 point $\mathbf{G}_1$ correspond to C_α and $\mathbf{G}_2$ corresponds to C_β when the peptide group is fixed. It is trivial to express a unit vector $\boldsymbol{l} = (l_x, l_y, l_z)$ along the rotation axis through the coordinates of the two chosen points:

$$\mathbf{g} = (g_x, g_y, g_z) = (G_{2x} - G_{1x}, G_{2y} - G_{1y}, G_{2z} - G_{1z}) \tag{7.1}$$

$$\boldsymbol{l} = \mathbf{g} / \|\mathbf{g}\| = \mathbf{g} / \sqrt{g_x^2 + g_y^2 + g_z^2} \tag{7.2}$$

The point **w** at the rotation axis can be taken for example as

$$\mathbf{w} = (w_x, w_y, w_z) = (G_{1x}, G_{1y}, G_{1z}) \tag{7.3}$$

We remind that the result is independent of the choice of a point at the rotation axis, see section 2.4. We remind the reader also that 4 parameters and not 6 (coordinates of the two points) are sufficient to define a rotation axis (section 2.2).

The shift $\mathbf{q}_n$ of a point (x_n, y_n, z_n) is defined now as (see section 2.3)

$$\begin{pmatrix} q_{nx} \\ q_{ny} \\ q_{nz} \end{pmatrix} = dA_n \begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix} - dW \begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix} = dA_{nw} \begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix} \tag{7.4}$$

where

$$W = \begin{pmatrix} 0 & w_z & -w_y \\ -w_z & 0 & w_x \\ w_y & -w_x & 0 \end{pmatrix}$$

$$A_{nw} = \begin{pmatrix} 0 & z_n - w_z & -(y_n - w_y) \\ -(z_n - w_z) & 0 & x_n - w_x \\ y_n - w_y & -(x_n - w_x) & 0 \end{pmatrix} \tag{7.5}$$

and d is a random parameter describing the amplitude of libration. Similarly to (4.2) if in (7.4) we use the single-term presentation through A_{nw}, the averaging of $\mathbf{q}_n \mathbf{q}_n^\tau$ gives

$$U_n = \left\langle \mathbf{q}_n \mathbf{q}_n^\tau \right\rangle = A_{nw} L A_{nw}^\tau = A_{nw} L_d A_{nw}^\tau \tag{7.6}$$

with

$$L = \left\langle d^2 \right\rangle \begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix} \begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix}^{\tau} = \left\langle d^2 \right\rangle \begin{pmatrix} l_x^2 & l_x l_y & l_x l_z \\ l_x l_y & l_y^2 & l_y l_z \\ l_x l_z & l_y l_z & l_z^2 \end{pmatrix} = \left\langle d^2 \right\rangle L_d \qquad (7.7)$$

Here $\left\langle d^2 \right\rangle$ characterises the random distribution and size of the libration angle. This is the single *random* parameter *to be adjusted* to the experimental data. As mentioned above, six *fixed* parameters, 4 of them being independent, are used to define matrices L_d.

In this simplest situation using the two-terms presentation in (7.4) requires more computing than that with A_{nw}.

7.2. Coordinate system aligned with the bond

In this trivial case of a libration around a single fixed axis a direct approach (7.5-7.7) with no intermediate coordinate systems seems to be preferable for practical applications. However, the procedure descried in this section may be useful to understand more complex situations.

Using matrix A_{nw} (7.5) instead of A_n (7.4) in fact means a "hidden" change of the origin of the coordinate system. This eliminates matrices T and S which are unnecessary in this trivial case. We may further modify the coordinate system by choosing new base vectors $(\mathbf{i}_l, \mathbf{j}_l, \mathbf{k}_l)$ such that the new vector $\mathbf{k}_l = \boldsymbol{l}$. To do so, we define the angle ψ between $\boldsymbol{l}$ and the axis $\mathbf{k}$

$$\psi = \arccos(\mathbf{kl}) = \arccos(l_z) \qquad (7.8)$$

and the angle φ between the axis $\mathbf{i}$ and the projection $\mathrm{l}_{xy} = (l_x, l_y, 0)$ of $\boldsymbol{l}$ into the plane **Oij** :

$$\cos(\varphi) = l_x \Big/ \sqrt{l_x^2 + l_y^2} \quad , \quad \sin(\varphi) = l_y \Big/ \sqrt{l_x^2 + l_y^2} \qquad (7.9)$$

(for illustration see figure 2.1b). Now matrix

$$R_l = \begin{pmatrix} \cos\varphi\cos\psi & -\sin\varphi & \cos\varphi\sin\psi \\ \sin\varphi\cos\psi & \cos\varphi & \sin\varphi\sin\psi \\ -\sin\psi & 0 & \cos\psi \end{pmatrix} \qquad (7.10)$$

describes the transformation of the original coordinate system (Appendix A1) with the base vectors $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ into an intermediate system with the base vectors $(\mathbf{i}_l, \mathbf{j}_l, \mathbf{k}_l)$, in particular

$$\begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix} = R_l \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \qquad (7.11)$$

and inversely

$$\begin{pmatrix}0\\0\\1\end{pmatrix} = R_l^{-1}\begin{pmatrix}l_x\\l_y\\l_z\end{pmatrix} \qquad (7.12)$$

The coordinates in the new system are calculated from the original coordinates of a point as

$$\begin{pmatrix}x'_n\\y'_n\\z'_n\end{pmatrix} = R_l^{-1}\begin{pmatrix}x_n - C_{1x}\\y_n - C_{1y}\\z_n - C_{1z}\end{pmatrix} = \begin{pmatrix}\cos\psi\cos\varphi & \cos\psi\sin\varphi & -\sin\psi\\ -\sin\varphi & \cos\varphi & 0\\ \sin\psi\cos\varphi & \sin\psi\sin\varphi & \cos\psi\end{pmatrix}\begin{pmatrix}x_n - C_{1x}\\y_n - C_{1y}\\z_n - C_{1z}\end{pmatrix} \qquad (7.13)$$

and matrix U'_n (see (4.2)) is

$$U'_n = \langle d^2\rangle A'_n\begin{pmatrix}0&0&0\\0&0&0\\0&0&1\end{pmatrix}A'^{\tau}_n = \langle d^2\rangle\begin{pmatrix}y'^2_n & -x'_n y'_n & 0\\ -x'_n y'_n & x'^2_n & 0\\ 0&0&0\end{pmatrix} \qquad (7.14)$$

resulting in (Appendix A2)

$$U_n = \langle d^2\rangle R_l\begin{pmatrix}y'^2_n & -x'_n y'_n & 0\\ -x'_n y' & x'^2_n & 0\\ 0&0&0\end{pmatrix}R_l^{\tau} \qquad (7.15)$$

7.3. Axis with the fixed direction

Now let's suppose that the two points defining the libration axis $\boldsymbol{l}$ (*e.g.* C_α and C_β atoms in figure 7.1) oscillate around their central positions. If (for simplicity) we consider the direction of the axis being fixed (for example when the distance between the atoms is much larger than their displacements; see figure 7.2 for another illustration), the displacement (7.4) with the definitions (7.5) becomes

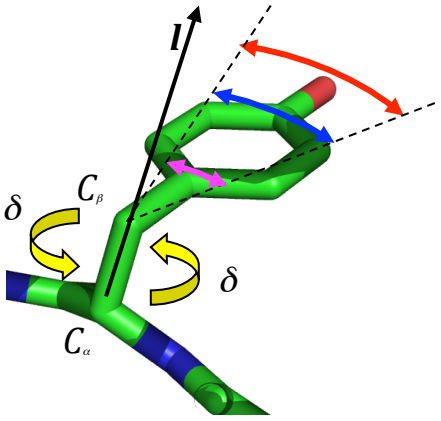


Figure 7.1. Schematic illustration of a libration around a bond. Angle δ is the parameter describing random oscillations around the bond $C_\alpha C_\beta$. A shift of each atom is proportional to its distance to this rotation axis.

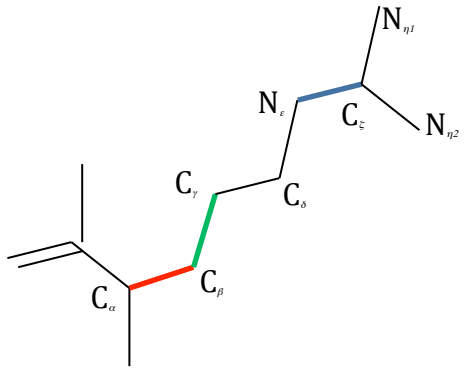


Figure 7.2. Schematic illustration of different kinds of libration axes associated with interatomic bonds. The peptide group is considered as a fixed. The rotation axis $C_\alpha C_\beta$ is a fixed axis (see Sections 7.1-7.2). The rotation axis $C_\beta C_\gamma$ changes its orientation (Section 7.5). If $N_\varepsilon C_\xi$ and $C_\alpha C_\beta$ are roughly parallel, the libration around $C_\alpha C_\beta$ translates the bond $N_\varepsilon C_\xi$ rather than changes its orientation (Sections 7.3-7.4).

$$\begin{pmatrix} q_{nx} \\ q_{ny} \\ q_{nz} \end{pmatrix} = \begin{pmatrix} u_x \\ u_y \\ u_z \end{pmatrix} + dA_n \begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix} - dW \begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix} = \begin{pmatrix} \hat{u}_x \\ \hat{u}_y \\ \hat{u}_z \end{pmatrix} + dA_n \begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix} \qquad (7.16)$$

where we introduce a new random vector, the same for all points of the rigid group

$$\begin{pmatrix} \hat{u}_x \\ \hat{u}_y \\ \hat{u}_z \end{pmatrix} = \begin{pmatrix} u_x \\ u_y \\ u_z \end{pmatrix} - dW \begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix} \qquad (7.17)$$

Differently from section 7.1, here the two-terms presentation is preferable and the averaging of $\mathbf{q}_n\mathbf{q}_n^\tau$ gives a sum

$$U_n = \left\langle \mathbf{q}_n\mathbf{q}_n^\tau \right\rangle = T + A_n L A_n^\tau + \left(A_n S + S^\tau A_n^\tau \right) \qquad (7.18)$$

always in the same form as (5.10) where the matrix A_n in the base $(\mathbf{i},\mathbf{j},\mathbf{k})$ is always (2.7) and other matrices are

$$T = \begin{pmatrix} \left\langle \hat{u}_x^2 \right\rangle & \left\langle \hat{u}_x\hat{u}_y \right\rangle & \left\langle \hat{u}_x\hat{u}_z \right\rangle \\ \left\langle \hat{u}_x\hat{u}_y \right\rangle & \left\langle \hat{u}_y^2 \right\rangle & \left\langle \hat{u}_y\hat{u}_z \right\rangle \\ \left\langle \hat{u}_x\hat{u}_z \right\rangle & \left\langle \hat{u}_y\hat{u}_z \right\rangle & \left\langle \hat{u}_z^2 \right\rangle \end{pmatrix}$$

$$L = \left\langle d^2 \right\rangle \begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix} \begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix}^\tau = \left\langle d^2 \right\rangle \begin{pmatrix} l_x^2 & l_x l_y & l_x l_z \\ l_x l_y & l_y^2 & l_y l_z \\ l_x l_z & l_y l_z & l_z^2 \end{pmatrix} \qquad (7.19)$$

$$S = \begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix} \begin{pmatrix} \left\langle d\hat{u}_x \right\rangle \\ \left\langle d\hat{u}_y \right\rangle \\ \left\langle d\hat{u}_z \right\rangle \end{pmatrix}^\tau = \begin{pmatrix} l_x\left\langle d\hat{u}_x \right\rangle & l_x\left\langle d\hat{u}_y \right\rangle & l_x\left\langle d\hat{u}_z \right\rangle \\ l_y\left\langle d\hat{u}_x \right\rangle & l_y\left\langle d\hat{u}_y \right\rangle & l_y\left\langle d\hat{u}_z \right\rangle \\ l_z\left\langle d\hat{u}_x \right\rangle & l_z\left\langle d\hat{u}_y \right\rangle & l_z\left\langle d\hat{u}_z \right\rangle \end{pmatrix}$$

This presentation shows ten *random* parameters required to define U_n: six independent elements of the matrix *T*, libration scale $\left\langle d^2 \right\rangle$ in *L* and three parameters $\left\langle d\hat{u}_x \right\rangle, \left\langle d\hat{u}_y \right\rangle, \left\langle d\hat{u}_z \right\rangle$ for the correlations of the rotation and translation components in *S*. All other values necessary to calculate U_n (7.18) for all points are defined through the coordinates of $\mathbf{G}_1$ and $\mathbf{G}_2$ (7.1-7.2) and the coordinates of atoms "hidden" in A_n.

7.4. Axis with the fixed direction – modified coordinate systems

As above, we can switch to an equivalent set of parameters that have clearer physical interpretation. First, we diagonalise *T* as discussed in section 3.1 and get matrix R_t that describes the transition (3.3) from the common system $(\mathbf{i},\mathbf{j},\mathbf{k})$ to another Cartesian coordinate system $(\mathbf{i}_t,\mathbf{j}_t,\mathbf{k}_t)$ with the axes along the three principal axes of vibration (by its construction, this new coordinate system has nothing to do with the geometry of the rigid body but is defined by the nature of its movement). This leads to

$$\begin{pmatrix} q_{nx} \\ q_{ny} \\ q_{nz} \end{pmatrix} = \begin{pmatrix} \hat{u}_x \\ \hat{u}_y \\ \hat{u}_z \end{pmatrix} + dA_n \begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix} = R_t \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} + dA_n \begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix} \tag{7.20}$$

and

$$T = \left\langle R_t \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}^{\tau} R_t^{\tau} \right\rangle = R_t \begin{pmatrix} \langle t_x^2 \rangle & \langle t_x t_y \rangle & \langle t_x t_z \rangle \\ \langle t_x t_y \rangle & \langle t_y^2 \rangle & \langle t_y t_z \rangle \\ \langle t_x t_z \rangle & \langle t_y t_z \rangle & \langle t_z^2 \rangle \end{pmatrix} R_t^{\tau} = R_t \begin{pmatrix} \langle t_x^2 \rangle & 0 & 0 \\ 0 & \langle t_y^2 \rangle & 0 \\ 0 & 0 & \langle t_z^2 \rangle \end{pmatrix} R_t^{\tau}$$

$$L = \langle d^2 \rangle \begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix} \begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix}^{\tau} = \langle d^2 \rangle \begin{pmatrix} l_x^2 & l_x l_y & l_x l_z \\ l_x l_y & l_y^2 & l_y l_z \\ l_x l_z & l_y l_z & l_z^2 \end{pmatrix} \tag{7.21}$$

$$S = \begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix} \begin{pmatrix} \langle dt_x \rangle \\ \langle dt_y \rangle \\ \langle dt_z \rangle \end{pmatrix}^{\tau} R_t^{\tau} = \begin{pmatrix} l_x \langle dt_x \rangle & l_x \langle dt_y \rangle & l_x \langle dt_z \rangle \\ l_y \langle dt_x \rangle & l_y \langle dt_y \rangle & l_y \langle dt_z \rangle \\ l_z \langle dt_x \rangle & l_z \langle dt_y \rangle & l_z \langle dt_z \rangle \end{pmatrix} R_t^{\tau}$$

We can also consider the Cartesian coordinate system $(\mathbf{i}_l, \mathbf{j}_l, \mathbf{k}_l)$ where $\mathbf{k}_l$ is aligned with the rotation axis, the corresponding matrix R_l being (7.10). Then (7.21) can be presented as

$$T = R_{lt} \begin{pmatrix} \langle t_x^2 \rangle & 0 & 0 \\ 0 & \langle t_y^2 \rangle & 0 \\ 0 & 0 & \langle t_z^2 \rangle \end{pmatrix} R_{lt}^{\tau}$$

$$L = \langle d^2 \rangle R_l \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} R_l^{\tau} \tag{7.22}$$

$$S = R_l \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \langle dt_x \rangle & \langle dt_y \rangle & \langle dt_z \rangle \end{pmatrix} R_{lt}^{\tau}$$

with

$$R_{lt} = R_t R_l^{-1} \tag{7.23}$$

Working with diagonalised matrices is more convenient and as previously shown aids in a better understanding of parameters of the *TLS* model. Also, we will see below in section 7.6 that there exists a special origin shift that diagonalises matrix *S*. Section 8.3 shows that in fact it is more convenient to start from diagonalisation of *L* and *S* and only then diagonalise *T*.

7.5. Libration axis that may change its direction

At the next level of generalisation we suppose that the direction of the rotation axis can vary. From now on let's assume that the coordinates (l_x, l_y, l_z) of the vector $\boldsymbol{l}$ be also random values. This may correspond to a general case of a group motion and not necessarily to a rotation around a covalent bond (see also an illustration in figure 7.2). If we introduce a new vector

$$\mathbf{d} = \begin{pmatrix} d_x \\ d_y \\ d_z \end{pmatrix} = d \begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix} = \begin{pmatrix} dl_x \\ dl_y \\ dl_z \end{pmatrix} \tag{7.24}$$

then matrices (7.19) in (7.18) become

$$T = \begin{pmatrix} \langle \hat{u}_x^2 \rangle & \langle \hat{u}_x \hat{u}_y \rangle & \langle \hat{u}_x \hat{u}_z \rangle \\ \langle \hat{u}_x \hat{u}_y \rangle & \langle \hat{u}_y^2 \rangle & \langle \hat{u}_y \hat{u}_z \rangle \\ \langle \hat{u}_x \hat{u}_z \rangle & \langle \hat{u}_y \hat{u}_z \rangle & \langle \hat{u}_z^2 \rangle \end{pmatrix} \tag{7.25}$$

$$L = \begin{pmatrix} \langle d_x^2 \rangle & \langle d_x d_y \rangle & \langle d_x d_z \rangle \\ \langle d_x d_y \rangle & \langle d_y^2 \rangle & \langle d_y d_z \rangle \\ \langle d_x d_z \rangle & \langle d_y d_z \rangle & \langle d_z^2 \rangle \end{pmatrix} \tag{7.26}$$

$$S = \begin{pmatrix} \langle d_x \hat{u}_x \rangle & \langle d_x \hat{u}_y \rangle & \langle d_x \hat{u}_z \rangle \\ \langle d_y \hat{u}_x \rangle & \langle d_y \hat{u}_y \rangle & \langle d_y \hat{u}_z \rangle \\ \langle d_z \hat{u}_x \rangle & \langle d_z \hat{u}_y \rangle & \langle d_z \hat{u}_z \rangle \end{pmatrix} \tag{7.27}$$

Calculating explicitly the elements of U_n through elements of the composite matrices (7.25-7.27) and atomic coordinates gives (we use the symmetry of the L matrix)

$$U_{nxx} = T_{xx} + \left(z_n^2 L_{yy} + y_n^2 L_{zz} - 2y_n z_n L_{yz}\right) + 2\left(z_n S_{yx} - y_n S_{zx}\right) \tag{7.28}$$

$$U_{nyy} = T_{yy} + \left(z_n^2 L_{xx} + x_n^2 L_{zz} - 2x_n z_n L_{xz}\right) + 2\left(x_n S_{zy} - z_n S_{xy}\right)$$

$$U_{nzz} = T_{zz} + \left(y_n^2 L_{xx} + x_n^2 L_{yy} - 2x_n y_n L_{xy}\right) + 2\left(y_n S_{xz} - x_n S_{yz}\right)$$

$$U_{nxy} = T_{xy} + \left(y_n z_n L_{xz} + x_n z_n L_{yz} - x_n y_n L_{zz} - z_n^2 L_{xy}\right) + \left(x_n S_{zx} - y_n S_{zy} + z_n \left(S_{yy} - S_{xx}\right)\right)$$

$$U_{nxz} = T_{xz} + \left(y_n z_n L_{xy} + x_n y_n L_{yz} - x_n z_n L_{yy} - y_n^2 L_{xz}\right) + \left(z_n S_{yz} - x_n S_{yx} + y_n \left(S_{xx} - S_{zz}\right)\right)$$

$$U_{nyz} = T_{yz} + \left(x_n z_n L_{xy} + x_n y_n L_{xz} - y_n z_n L_{xx} - x_n^2 L_{yz}\right) + \left(y_n S_{xy} - z_n S_{xz} + x_n \left(S_{zz} - S_{yy}\right)\right)$$

(note the "+" sign at $x_n y_n L_{xz}$ term in the last equation in comparison with 1.2.11.5 in Coppens, 2006; this agrees with Table 1 in *TLSView Manual* (Merritt, pymmlib.sourceforge.net/tlsview/tlsview.html).

Clearly this general situation (7.25-7.27) is characterised by 21 parameters: 6 to describe *T*, 6 to describe *L* and 9 to describe an asymmetric matrix *S*. It is ubiquitously pointed out in the literature (see for example Schomaker & Trueblood (1968) or Coppens (2006)) that the linear combinations (7.28) of the *T*, *L* and *S* elements use only differences between S_{xx}, S_{yy}, S_{zz} and not these values themselves. Knowledge of two of these differences defines the third one. Simultaneous increasing or decreasing of $\langle d_x\hat{u}_x\rangle, \langle d_y\hat{u}_y\rangle, \langle d_z\hat{u}_z\rangle$ by the same quantity does not change U_n. Some consequences of this are discussed below in section 7.8.

7.6. Symmetrisation of S

Following from section 5.5, we will show that there exists a special origin for *S* which makes *S* symmetric (for example see Brenner (1967)). Accordingly to section 5.4, since *P* is antisymmetric (see (5.16)), $P^\tau = -P$ and *L* is symmetric, $L^\tau = L$, then

$$\left(S'\right)^\tau = S^\tau - \left(LP\right)^\tau \tag{7.29}$$

and

$$\left(S'\right)^\tau - S' = S^\tau - S + LP - \left(LP\right)^\tau \tag{7.30}$$

If with some choice of the origin matrix S' is symmetric, the last expression is equal to 0 giving

$$S^\tau - S = \left(LP\right)^\tau - LP = D \tag{7.31}$$

where *D* by construction is antisymmetric with diagonal elements equal to zero and off-diagonal elements equal to

$$\begin{aligned} D_{xy} &= -D_{yx} = p_x L_{xz} + p_y L_{yz} - p_z\left(L_{xx} + L_{yy}\right) = S_{yx} - S_{xy} \\ D_{zx} &= -D_{xz} = p_x L_{xy} - p_y\left(L_{xx} + L_{zz}\right) + p_z L_{yz} = S_{xz} - S_{zx} \\ D_{yz} &= -D_{zy} = -p_x\left(L_{yy} + L_{zz}\right) + p_y L_{xy} + p_z L_{xz} = S_{zy} - S_{yz} \end{aligned} \tag{7.32}$$

Here p_x, p_y, p_z are unknown parameters and the right-hand expressions contain corresponding elements of (7.27). The determinant of this system of linear equations, after insertion of the values from (7.26), is equal to

$$\begin{aligned} &\left(\langle d_x^2\rangle + \langle d_y^2\rangle\right)\left(\langle d_x^2\rangle\langle d_y^2\rangle - \langle d_x d_y\rangle^2\right) + \left(\langle d_x^2\rangle + \langle d_z^2\rangle\right)\left(\langle d_x^2\rangle\langle d_z^2\rangle - \langle d_x d_z\rangle^2\right) + \\ &+\left(\langle d_y^2\rangle + \langle d_z^2\rangle\right)\left(\langle d_y^2\rangle\langle d_z^2\rangle - \langle d_y d_z\rangle^2\right) + \\ &+2\left(\langle d_x^2\rangle\langle d_y^2\rangle\langle d_z^2\rangle - \langle d_x d_y\rangle\langle d_x d_z\rangle\langle d_y d_z\rangle\right) \end{aligned} \tag{7.33}$$

and is non-negative by the Cauchy-Schwarz inequality. It is equal to zero only in the hypotheticial case of a full correlation of all three motions, a case that does not happen in practice. This means that the system of equations (7.32) has always a unique solution. The corresponding point $\left(p_x, p_y, p_z\right)$ is called centre of libration (Pawley (1963), Hirshfeld *et al*. (1963) Schomaker & Trueblood (1968) and

Scheringer (1973)), "centre of diffusion" or "centre of reaction" (Brenner (1967), Tickle & Moss (1999)).

In fact, a possibility to symmetrise *S* is evident because it was demonstrated in case of a rotation around the **k** axis (section 5.4) and because the symmetry property is conserved with the rotation of the coordinate system (Appendix A4). This proves also that such a choice of the origin simultaneously minimises the trace of *T*, $tr(T) \rightarrow \min_{p_x, p_y, p_z}$, since the trace is also invariant with the coordinate system (Appendix A4).

When symmetrizing the matrix *S*, the number of the independent elements in it is reduced from 9 to 6. There is no contradiction since 3 'disappearing' parameters are now converted into *a priori unknown* coordinates of the reaction center.

7.7. T and L parameterisation

As previously, both symmetric matrices *T* and *L* can be diagonalised

$$T = R_t \begin{pmatrix} t_x^2 & 0 & 0 \\ 0 & t_y^2 & 0 \\ 0 & 0 & t_z^2 \end{pmatrix} R_t^\tau \qquad (7.34)$$

$$L = R_d \begin{pmatrix} d_{lx}^2 & 0 & 0 \\ 0 & d_{ly}^2 & 0 \\ 0 & 0 & d_{lz}^2 \end{pmatrix} R_d^\tau \qquad (7.35)$$

(see also Painter & Merritt (2006) about the diagonalisation of *L*). A difference with section 7.2 is that now R_l defined in (7.10) cannot be used anymore and is substituted by R_d that is built from the coordinates of the eigenvectors of *L*, similarly to R_t (see section 3). These eigenvectors describe three mutually orthogonal axes around which the rigid body has uncorrelated librations with the parameters d_{lx}, d_{ly}, d_{lz}, similarly to (6.6). So both *T* and *L* are characterised each by three mutually orthogonal directions and by three displacements corresponding to these directions. Other types of efficient parameterisation can be suggested (see for example Pawley (1970) or Rae (1975a, b)). In particular, the parameterisation suggested by Rae allows easy and efficient introducing of constraints on the *TLS* parameters.

7.8. S parameterisation

A non-symmetric matrix *S* is defined by nine its elements. However, only the differences between its diagonal elements are used in (7.28). The relation

$$(S_{yy} - S_{xx}) + (S_{xx} - S_{zz}) + (S_{zz} - S_{yy}) = 0 \qquad (7.36)$$

means that only 2 of these differences are sufficient to define U_n unambiguously reducing the total number of parameters to 20 = 6 + 6 + 8. Traditionally starting from Schomaker and Trueblood (1968) this constraint on the diagonal elements of *S* is presented differently. Equations (7.28) mean that the knowledge of U_n cannot unambiguously define the diagonal elements of *S* and an arbitrary constant *h* can be added to all of them simultaneously. The resulting diagonal elements cannot be too large since they should satisfy the Cauchy-Schwarz inequalities

$$S_{xx}^2 \le \left\langle \hat{u}_x^2 \right\rangle \left\langle d_x^2 \right\rangle,\ S_{yy}^2 \le \left\langle \hat{u}_y^2 \right\rangle \left\langle d_y^2 \right\rangle, S_{zz}^2 \le \left\langle \hat{u}_z^2 \right\rangle \left\langle d_z^2 \right\rangle \qquad (7.37)$$

Historically, the convention of setting the trace of the S matrix equal to 0,

$$S_{xx} + S_{yy} + S_{zz} = 0 \qquad (7.38)$$

is used to assert that (7.37) holds true.

Such a condition corresponds to (6.6) obtained for three rotation axes (we saw in section 6.7 above that the three resulting axes are always mutually orthogonal). The condition (7.38) can be also understood from the following equality:

$$\begin{aligned} tr(S) &= \left\langle d_x \hat{u}_x \right\rangle + \left\langle d_y \hat{u}_y \right\rangle + \left\langle d_z \hat{u}_z \right\rangle = \left\langle d_x \hat{u}_x + d_y \hat{u}_y + d_z \hat{u}_z \right\rangle \\ &= \left\langle dl_x \left[u_x - \left(l_y w_z - l_z w_y \right) \right] + dl_y \left[u_y - \left(l_z w_x - l_x w_z \right) \right] + dl_z \left[u_z - \left(l_x w_y - l_y w_x \right) \right] \right\rangle \\ &= \left\langle \left(dl_x \right) u_x + \left(dl_y \right) u_y + \left(dl_z \right) u_z \right\rangle + \\ &+ \left\langle d \left[l_x \left(l_y w_z - l_z w_y \right) + dl_y \left(l_z w_x - l_x w_z \right) + dl_z \left(l_x w_y - l_y w_x \right) \right] \right\rangle \\ &= \left\langle \left(dl_x \right) u_x + \left(dl_y \right) u_y + \left(dl_z \right) u_z \right\rangle \end{aligned} \qquad (7.39)$$

The second term in (7.39) disappeares because the expression in $[\ ]$ is a scalar product of two orthogonal vectors, $\boldsymbol{l}$ and $\boldsymbol{l} \times \mathbf{w}$. Therefore, condition (7.38) becomes

$$\left\langle \left(dl_x \right) u_x + \left(dl_y \right) u_y + \left(dl_z \right) u_z \right\rangle = 0, \qquad (7.40)$$

a requirement that among of all possible decompositions of U_n we choose that with no correlation between the translation and rotation. This agrees with the remarks by some authors that the difficulty of finding individual S_{xx}, S_{yy}, S_{zz} "*arises from incomplete knowledge about the correlation of atomic motions*" (Bürgi (1989), see also Scheringer (1973)).

8. General case

8.1. Several axes in a general position

Let's suppose now that a rigid body participates simultaneously in several librations, K in total, of different amplitudes (see for example Stuart & Phillips (1985)). These librations are defined by axes $\boldsymbol{l}_k$, k = 1,K, by the points $\mathbf{w}_k$ at the corresponding axis and by the elementary shifts d_k, due to rotations (section 2.3). The axes are not necessarily mutually intersecting. The displacement vector $\mathbf{v}$ due to rotations becomes (see (2.9))

$$\mathbf{v}_n = \sum_k d_k A_n \boldsymbol{l}_k - \sum_k d_k \boldsymbol{l}_k \times \mathbf{w}_k = A_n \sum_k d_k \boldsymbol{l}_k - \sum_k d_k \boldsymbol{l}_k \times \mathbf{w}_k = dA_n \hat{\boldsymbol{l}} - \hat{\mathbf{d}}_w \qquad (8.1)$$

Now, with $d = \left\| \hat{\mathbf{d}} \right\|$ and $\hat{\boldsymbol{l}} = \hat{\mathbf{d}} / d$, both

$$\hat{\mathbf{d}} = \sum_k d_k \boldsymbol{l}_k = \begin{pmatrix} \sum_k d_k l_{kx} \\ \sum_k d_k l_{ky} \\ \sum_k d_k l_{kz} \end{pmatrix} = \begin{pmatrix} \hat{d}_x \\ \hat{d}_y \\ \hat{d}_z \end{pmatrix} = d\hat{\boldsymbol{l}} \tag{8.2}$$

and

$$\hat{\mathbf{d}}_w = \sum_k d_k \boldsymbol{l}_k \times \mathbf{w}_k = \sum_k d_k W_k \boldsymbol{l}_k = \sum_k d_k \begin{pmatrix} 0 & w_{kz} & -w_{ky} \\ -w_{kz} & 0 & w_{kx} \\ w_{ky} & -w_{kx} & 0 \end{pmatrix} \begin{pmatrix} l_{kx} \\ l_{ky} \\ l_{kz} \end{pmatrix} =$$

$$= \begin{pmatrix} \sum_k d_k \left(w_{kz} l_{ky} - w_{ky} l_{kz} \right) \\ \sum_k d_k \left(w_{kx} l_{ky} - w_{kz} l_{kx} \right) \\ \sum_k d_k \left(w_{ky} l_{kx} - w_{kx} l_{ky} \right) \end{pmatrix} = \begin{pmatrix} \hat{d}_{wx} \\ \hat{d}_{wy} \\ \hat{d}_{wz} \end{pmatrix} \tag{8.3}$$

are random vectors depending on random variables d_k and on the parameters of the axes. The second term in (8.1) is common for all points and acts as an apparent (random) translation of the rigid body (see section 4.2). Expression (8.1) shows that a multiple-axes rotation may be considered as a rotation around a single random rotation axis (section 7.5). The normalised vector $\hat{\boldsymbol{l}}$, $\left\| \hat{\boldsymbol{l}} \right\| = 1$, defines the rotation axis and d is the parameter for the rotation angle as discussed in section 2.3.

8.2. General formulae

When a translational displacement complements rotations, the total displacement $\mathbf{q}_n = \mathbf{u} + \mathbf{v}_n$ of the atom n at a point $\mathbf{r}_n$ is the sum of $\mathbf{u}$ and $\mathbf{v}_n$ due to rigid body translation and libration, respectively:

$$\begin{pmatrix} q_{xn} \\ q_{yn} \\ q_{zn} \end{pmatrix} = \begin{pmatrix} u_x \\ u_y \\ u_z \end{pmatrix} + A_n \begin{pmatrix} \hat{d}_x \\ \hat{d}_y \\ \hat{d}_z \end{pmatrix} - \begin{pmatrix} \hat{d}_{wx} \\ \hat{d}_{wy} \\ \hat{d}_{wz} \end{pmatrix} = \begin{pmatrix} u_x - \hat{d}_{wx} \\ u_y - \hat{d}_{wy} \\ u_z - \hat{d}_{wz} \end{pmatrix} + A_n \begin{pmatrix} \hat{d}_x \\ \hat{d}_y \\ \hat{d}_z \end{pmatrix} = \begin{pmatrix} \hat{u}_x \\ \hat{u}_y \\ \hat{u}_z \end{pmatrix} + A_n \begin{pmatrix} \hat{d}_x \\ \hat{d}_y \\ \hat{d}_z \end{pmatrix} \tag{8.4}$$

where random values are $\hat{u}_x, \hat{u}_y, \hat{u}_z, \hat{d}_x, \hat{d}_y, \hat{d}_z$. For the atom n, the components of U_n expressed in the original Cartesian coordinate system $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ as (7.19) are

$$T = \begin{pmatrix} \langle \hat{u}_x^2 \rangle & \langle \hat{u}_x \hat{u}_y \rangle & \langle \hat{u}_x \hat{u}_z \rangle \\ \langle \hat{u}_x \hat{u}_y \rangle & \langle \hat{u}_y^2 \rangle & \langle \hat{u}_y \hat{u}_z \rangle \\ \langle \hat{u}_x \hat{u}_z \rangle & \langle \hat{u}_y \hat{u}_z \rangle & \langle \hat{u}_z^2 \rangle \end{pmatrix} \tag{8.5}$$

$$L = \begin{pmatrix} \langle \hat{d}_x^2 \rangle & \langle \hat{d}_x \hat{d}_y \rangle & \langle \hat{d}_x \hat{d}_z \rangle \\ \langle \hat{d}_x \hat{d}_y \rangle & \langle \hat{d}_y^2 \rangle & \langle \hat{d}_y \hat{d}_z \rangle \\ \langle \hat{d}_x \hat{d}_z \rangle & \langle \hat{d}_y \hat{d}_z \rangle & \langle \hat{d}_z^2 \rangle \end{pmatrix} \quad (8.6)$$

$$S = \begin{pmatrix} \langle \hat{d}_x \hat{u}_x \rangle & \langle \hat{d}_x \hat{u}_y \rangle & \langle \hat{d}_x \hat{u}_z \rangle \\ \langle \hat{d}_y \hat{u}_x \rangle & \langle \hat{d}_y \hat{u}_y \rangle & \langle \hat{d}_y \hat{u}_z \rangle \\ \langle \hat{d}_z \hat{u}_x \rangle & \langle \hat{d}_z \hat{u}_y \rangle & \langle \hat{d}_z \hat{u}_z \rangle \end{pmatrix} \quad (8.7)$$

Equations (8.5)-(8.7) are literally similar to (7.25)-(7.27) derived and analysed previously.

8.3. Analysis of the *TLS* matrices

Both simple examples above and the demonstration in section 8.2 show that for all kinds of rigid-body oscillations, considered in a harmonic approximation, the set of matrices U_n for all atoms of the group can be expressed as a sum (5.10) of four terms where T, L and S are common for all atoms and each antisymmetric matrix A_n contains individual coordinates of the atom n. Matrices T and L are symmetric while S is not necessarily symmetric.

Now let's suppose that crystallographic calculations, for example as discussed below in section 9, gave us some T, L & S matrices and we want to find which rigid-body motion produces these matrices. To answer this question several steps should be performed. We remind the reader that rotation axes that do not pass through the origin as well as the rotation-translation correlation contribute to an apparent translation and that these contributions should be removed properly in order to define the pure translation component.

a) Origin shift. Shift of the origin into the reaction center p_x, p_y, p_z (section 7.6) is the solution of the system of linear equations (7.32). The new matrices T', L', S' in this new coordinate system are obtained accordingly to (5.18). In this coordinate system matrix S' is symmetric, trace of T' is the smallest possible; both these properties are retained in further rotation of the coordinate system (Appendix A4). Obtain new atomic coordinates (x'_n, y'_n, z'_n) by subtracting (p_x, p_y, p_z), $\mathbf{r}'_n = \mathbf{r}_n - \mathbf{p}$. In fact here and later we do not need atomic coordinates for the T, L, or S interpretation; the transformation is done simply for completeness.

b) Diagonalisation of L. Find 3 non-negative eigenvalues of matrix L' and three mutually orthogonal eigenvectors; rotation matrix R_d (7.35) is composed from the coordinates of these eigenvectors. Choose a new coordinate system with the new axes along the three eigenvectors; R_d is the transformation matrix to this system. Recalculate the matrices $L'' = R_d L' R_d^\tau$ (7.35), $T'' = R_d T' R_d^\tau$, $S'' = R_d S' R_d^\tau$ and new atomic coordinates as $\mathbf{r}''_n = R_d^{-1}\mathbf{r}'_n = R_d^\tau \mathbf{r}'_n$ in this new system (Appendices A1-A2). In the new coordinate system matrix L'' is diagonal with the elements $L''_{xx}, L''_{yy}, L''_{zz}$. The rotation axes are parallel to the new coordinate axes and pass through the points $(0, w_y^i, w_z^i)$, $(w_x^j, 0, w_z^j)$, $(w_x^k, w_y^k, 0)$ to be defined.

c) Position of rotation axes. Obtain estimates $\bar{d}_x = \sqrt{L''_{xx}}$, $\bar{d}_y = \sqrt{L''_{yy}}$, $\bar{d}_z = \sqrt{L''_{zz}}$ (6.6) of the

rotation parameters around the three axes defined at the previous step; calculate the positions of the rotation axes (4.16):

$$w_y^i = -\frac{S''_{xz}}{L''_{xx}},\ w_z^i = \frac{S''_{xy}}{L''_{xx}},\ w_x^j = \frac{S''_{yz}}{L''_{yy}},\ w_z^j = -\frac{S''_{yx}}{L''_{yy}},\ w_x^k = -\frac{S''_{zy}}{L''_{zz}},\ w_y^k = \frac{S''_{zx}}{L''_{zz}} \qquad (8.8)$$

d) Contribution of rotations to T due to axes displacement. Calculate contribution (6.5) of the rotations to the translation of the group due to the displacement of the axes

$$\Delta_T = \begin{pmatrix} \left(w_z^j\right)^2 L''_{yy} + \left(w_y^k\right)^2 L''_{zz} & -w_x^k w_y^k L''_{zz} & -w_x^j w_z^j L''_{yy} \\ -w_x^k w_y^k L''_{zz} & \left(w_z^i\right)^2 L''_{xx} + \left(w_x^k\right)^2 L''_{zz} & -w_y^i w_z^i L''_{xx} \\ -w_x^j w_z^j L''_{yy} & -w_y^i w_z^i L''_{xx} & \left(w_y^i\right)^2 L''_{xx} + \left(w_x^j\right)^2 L''_{yy} \end{pmatrix}; \qquad (8.9)$$

the residual translation matrix after removal this contribution is

$$T''' = T'' - \Delta_T \qquad (8.10)$$

e) Minimisation of correlation between translation and rotation. Calculate the trace $trS'' = S''_{xx} + S''_{yy} + S''_{zz}$ of S'' and get a new $\widetilde{S}$ (section 7.8) with the minimal correlation between translation and rotation (7.40)

$$\widetilde{S} = S'' - \begin{pmatrix} \frac{1}{3} trS'' & 0 & 0 \\ 0 & \frac{1}{3} trS'' & 0 \\ 0 & 0 & \frac{1}{3} trS'' \end{pmatrix} \qquad (8.11)$$

f) Contribution of screwing to T. Obtain estimates $\bar{s}_x = \frac{\widetilde{S}_{xx}}{L''_{xx}}$, $\bar{s}_y = \frac{\widetilde{S}_{yy}}{L''_{yy}}$, $\bar{s}_z = \frac{\widetilde{S}_{zz}}{L''_{zz}}$ of the screw parameters (6.9)-(6.10) following the rotation axes currently aligned with the coordinate axes; remove the contribution of the screwing from the translation matrix (6.8) :

$$\widetilde{T} = T''' - \begin{pmatrix} \bar{s}_x \widetilde{S}_{xx} & 0 & 0 \\ 0 & \bar{s}_y \widetilde{S}_{yy} & 0 \\ 0 & 0 & \bar{s}_z \widetilde{S}_{zz} \end{pmatrix} \qquad (8.12)$$

The resulting matrix $\widetilde{T}$ stands for the pure translation of the rigid group with the contribution of other movements removed.

g) Three uncorrelated translations. Find 3 non-negative eigenvalues of matrix $\widetilde{T}$ and three mutually orthogonal eigenvectors (section 3). The three eigenvectors give the directions of the uncorrelated translations and the corresponding eigenvalues are means square displacement along these axes. Do not forget that these vectors are given in the coordinate system with the axes parallel to the rotation axes and not in the original one.

9. Search for the optimal *TLS* decomposition

9.1. Optimal *TLS* decomposition and refinement with *TLS*

Summarizing the analysis above, we conclude that for all kinds of harmonic oscillation of a rigid group, the matrices U_n for all its points can be presented as (5.10) through three common matrices *T*, *L* & *S* and through matrices A_n specific for each point. Sections 3-8 above show how to calculate the contribution $U_{TLS,n}$ of a rigid group motion into atomic displacement parameters U_n of the corresponding atoms when the movement is known. A frequent task (see for example Sternberg *et al*. (1979)) is the inverse problem: given a set of matrices for a group of atoms (let's call these matrices $\overline{U}_n$, differently from U_n that are matrices calculated from the atomic parameters), present them in the *TLS* form considering that this group oscillates as a rigid body, *i.e.* find all the elements of the three composing matrices (8.5-8.7) reproducing as close as possible the whole set of $\overline{U}_n$. We may note that such a problem is important not only for their *a posteriori* analysis (see for example Holbrook & Kim (1984), Kuriyan & Weiss (1991), Stec *et al.*(1995) while some more applications can be found in the additional list of references) but also to obtain initial values of *TLS* parameters for their further refinement as discussed below.

The problem of decomposing the set of $\overline{U}_n$ into *TLS* (find *TLS* such that the corresponding U_n are close as much as possible to $\overline{U}_n$) is more complicated in a real situation when the matrices $\overline{U}_n$ contain contribution of individual (independent) atomic vibrations $U_{ind,n}$ and other contributions and errors, when rigid groups are an idealisation and when a composition of these rigid groups is *a priori* unknown. Looking for optimal values of the *TLS* parameters means to minimise some target function with respect to these parameters. Here we do not discuss how to decide which atoms belong to which rigid group. This problem was studied by, for example, Winn *et al*. (2001) and Painter & Merritt (2005, 2006a, 2006b). See also references therein for further details outside the scope of this article.

To find an optimal solution of the problem, a formal measure of the solution quality shall be introduced first. Traditionally, it might be a least-squares target for a difference between the elements of all $\overline{U}_n$ and corresponding calculated U_n (for example, Painter & Merritt (2006)):

$$\sum_{n=1,N}\left\{\left(U_{TLS,nxx}-\bar{U}_{nxx}\right)^2+\left(U_{TLS,nyy}-\bar{U}_{nyy}\right)^2+\left(U_{TLS,nzz}-\bar{U}_{nzz}\right)^2+\right. \tag{9.1}$$

$$\left.+\left(U_{TLS,nxy}-\bar{U}_{nxy}\right)^2+\left(U_{TLS,nxz}-\bar{U}_{nxz}\right)^2+\left(U_{TLS,nyz}-\bar{U}_{nyz}\right)^2\right\}\rightarrow \min$$

However, this target may be more sensible to errors for atoms that are far from the reaction center (since the matrices A_n are "larger"), does not distinguish atomic types (a proper modeling of U_n for a heavy atom may be more important considering its contribution to the electron density and structure factors) and others. Also the sum over all elements of the matrix U_n is probably non-optimal since not all of them contribute equally to the form of electron clouds. From that point of view, the target function (9.1) is an intuitive, but has no real physical background. Merritt (1999) suggested a more sophisticated density-correlation-based target expressed through anisotropic atomic displacement parameters

$$\frac{\left(\det \bar{U}^{-1} \det \bar{U}_{TSL}^{-1}\right)^{1/4}}{\left[\det\left(\bar{U}^{-1} + \bar{U}_{TSL}^{-1}\right)/8\right]} \rightarrow \max \qquad (9.2)$$

In fact, some more general function of the form

$$f\left(TLSparameters; \left\{\bar{U}_{nxx}, \bar{U}_{nxx}, \bar{U}_{nxx}, \bar{U}_{nxx}, \bar{U}_{nxx}, \bar{U}_{nxx}, n = 1, N\right\}\right) \rightarrow \min \qquad (9.3)$$

can be introduced. Here in the simplest case '*TLS parameters*' mean the elements of the matrices (8.5-8.7); some other ways of parameterisation have been discussed in the text. Formally speaking, one should control that the residual matrices $\bar{U}_n - U_{TLS,n}$ are positive definite since they suppose to correspond to $U_{ind,n}$. At the same time, practical studies show that neglecting this restriction allows to improve the *R*-factors (Afonine, phenix-online.org/newsletter/CCN_2010_07.pdf).

Another task, different from reinterpretation of $\bar{U}_n$, is to consider directly *TLS parameters* as parameters of the model during model refinement

$$f(\{atomic\ coordinates\}, \{TLS\ parameters\}, \{F_{obs}(hkl)\}) \rightarrow \min \qquad (9.4)$$

as it was implemented in *CORELS* (Sussman *et al*., 1977), *RESTRAIN* (Driessen *et al*., 1989) and later in *REFMAC* (Winn *et al*., 2001), *phenix.refine* (Afonine *et al*., 2005) and *BUSTER* (Bricogne *et al*., 2009). See also Moss et. al. (1996). Here f can be any appropriate function of model parameters and experimental diffraction data.

9.2. Practical scheme

As stated above, the formal goal is to find the parameters of the *T*, *L* and *S* matrices that minimise the target (9.3) or (9.4). One can note that the elements of $U_{TLS,n}$ are linear functions of the elements of the *TLS* matrices. When the target is a quadratic function of elements of U_n like (9.1), a close solution of the problem can be suggested solving the system of linear equations. However, in practice, even in this case iterative optimisation methods may be required where knowledge of best possible initial parameter values facilitates solving the problem.

Given a set of $\bar{U}_n$ values, the *TLS* matrices intuitively should include "as much as possible" of common atomic movement leaving the rest to individual atomic movements. To start the procedure, one can try to assign all possible common motion to the T matrix making L and S equal to zero unless they are known from previous refinement cycles. When all atomic displacement factors are isotropic with $\bar{B}_n$ instead of $\bar{U}_n$, the search for the "maximal" T is trivial. This matrix is diagonal, with all three diagonal elements equal to the minimal $\bar{B}_n$ value over all atoms of the group. When the displacement parameter is anisotropic for some of atoms of the group, the "maximal" T can be found following the algorithm described by Afonine & Urzhumtsev (2007). Then the parameters of all three TLS matrices are refined.

9.3. Once more about the origin at the reaction center

When using *TLS* parameterisation, the choice of group origin is arbitrary and does not affect the matrices $U_{TLS,n}$ calculated from obtained *TLS*. Typically the origin is taken as center of mass or center of geometry of a *TLS* group (the difference is the use or lack of use of the molecular weights in averaging

the atomic coordinates; usually this difference is insignificant) as suggested by Rae (1975b) and used for example in *RESTRAIN* (Driessen *et al.*, 1989). Citing Tickle & Moss (1999), "*...for a molecule constrained by intermolecular forces in a crystalline environment, the centre of gravity loses the special significance that it has for freely moving rigid bodies.*" The natural origin for this model is a reaction center and the idea that the body is oscillated around some bond(s) suggests that it is rather at a periphery of the body and not at its centre. Unlikely for the center of mass, the reaction center is initially unknown.

If a hypothetical rotation axis is known in advance, one can initially choose any its point as the origin instead of the center of mass. In any case, it seems useful to find the coordinates of the reaction center from diagonalisation of S (section 7.6) when it starts to be known and to reassign there the origin for further calculations (see for example *TLSANL* by Howlin *et al.*, 1993).

Appendix A. Changing the coordinate system

A1. Transformation matrix

Let (x, y, z) be coordinates of a vector $\mathbf{q}$ in some coordinate system with the basis vectors $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ and (x', y', z') be coordinates of the same vector in another coordinate system with the basis vectors $(\mathbf{i}', \mathbf{j}', \mathbf{k}')$. Let (i'_x, i'_y, i'_z), (j'_x, j'_y, j'_z) and (k'_x, k'_y, k'_z) be coordinates of the vectors of the new base in the initial coordinate system. We define the transformation matrix

$$Q = \begin{pmatrix} i'_x & j'_x & k'_x \\ i'_y & j'_y & k'_y \\ i'_z & j'_z & k'_z \end{pmatrix} \qquad \text{(A1.1)}$$

It is easy to see the rule

$$\begin{pmatrix} i'_x \\ i'_y \\ i'_z \end{pmatrix} = Q \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} j'_x \\ j'_y \\ j'_z \end{pmatrix} = Q \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} k'_x \\ k'_y \\ k'_z \end{pmatrix} = Q \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \qquad \text{(A1.2)}$$

linking the initial coordinates of the base vectors $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ with their coordinates in the new system. The same rule can be applied to any vector $\mathbf{q}$:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = Q \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} \qquad \text{(A1.3)}$$

To demonstrate this, it is sufficient to note that

$$\mathbf{q} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k} = x'\mathbf{i}' + y'\mathbf{j}' + z'\mathbf{k}' \qquad \text{(A1.4)}$$

That gives, accordingly to (A1.2), in the original coordinate system

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = x' Q \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + y' Q \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + z' Q \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = Q \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} \qquad \text{(A1.5)}$$

A2. Relation between coordinates

Let (q_x, q_y, q_z) and (q'_x, q'_y, q'_z) be coordinates of a vector $\mathbf{q}$ in some coordinate systems with the base vectors $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ and $(\mathbf{i}', \mathbf{j}', \mathbf{k}')$, respectively. If Q is the transformation matrix as defined in Appendix A1, then accordingly to (A1.3) the coordinates of a vector in these coordinate systems are linked by relations:

$$\begin{pmatrix} q_x \\ q_y \\ q_z \end{pmatrix} = Q \begin{pmatrix} q'_x \\ q'_y \\ q'_z \end{pmatrix} \quad , \quad \begin{pmatrix} q'_x \\ q'_y \\ q'_z \end{pmatrix} = Q^{-1} \begin{pmatrix} q_x \\ q_y \\ q_z \end{pmatrix}, \tag{A2.1}$$

$$(q_x\ q_y\ q_z) = (q'_x\ q'_y\ q'_z) Q^{\tau} \quad , \quad (q'_x\ q'_y\ q'_z) = (q_x\ q_y\ q_z)(Q^{\tau})^{-1}$$

Let's have a quadratic function of the coordinates

$$f(\mathbf{q}) = \left(-\mathbf{q}^{\tau} U^{-1} \mathbf{q}\right) = (q_x\ q_y\ q_z) U^{-1} \begin{pmatrix} q_x \\ q_y \\ q_z \end{pmatrix} \tag{A2.2}$$

such that this function is independent of the choice of the coordinate system. This means that, using (A2.1),

$$f(\mathbf{q}) = (q'_x\ q'_y\ q'_z)\left[U'\right]^{-1} \begin{pmatrix} q'_x \\ q'_y \\ q'_z \end{pmatrix} = (q_x\ q_y\ q_z) U^{-1} \begin{pmatrix} q_x \\ q_y \\ q_z \end{pmatrix}$$

$$= \left[(q'_x\ q'_y\ q'_z) Q^{\tau}\right] U^{-1} \left[Q \begin{pmatrix} q'_x \\ q'_y \\ q'_z \end{pmatrix} \right] = (q'_x\ q'_y\ q'_z) \left[Q^{\tau} U^{-1} Q\right] \begin{pmatrix} q'_x \\ q'_y \\ q'_z \end{pmatrix} \tag{A2.3}$$

requiring the relation

$$U'^{-1} = Q^{\tau} U^{-1} Q \tag{A2.4}$$

or

$$U' = Q^{-1} U_n (Q^{\tau})^{-1} \tag{A2.5}$$

The last relation is satisfied for U defined as (2.3) since

$$U' = \left\langle \mathbf{q}\mathbf{q}^{\tau} \right\rangle = \left\langle \begin{pmatrix} q'_x \\ q'_y \\ q'_z \end{pmatrix} \begin{pmatrix} q'_x & q'_y & q'_z \end{pmatrix} \right\rangle$$

$$= \left\langle Q^{-1} \begin{pmatrix} q_x \\ q_y \\ q_z \end{pmatrix} \begin{pmatrix} q_x & q_y & q_z \end{pmatrix} (Q^{\tau})^{-1} \right\rangle = Q^{-1} U (Q^{\tau})^{-1} \qquad \text{(A2.6)}$$

Quite often both old and new bases are composed from the unit vectors mutually orthogonal to each other. This means that the transformation matrix Q is nothing else but a rotation matrix. For example, a rotation by angle φ around the axis $\mathbf{k}$ is defined by the matrix

$$Q = R_z(\varphi) = \begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \text{(A2.7)}$$

Note also that here the inverse transformation is a rotation by angle $-\varphi$ giving

$$Q^{-1} = R_z^{-1}(\varphi) = R_z(-\varphi) = \begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} = R_z^{\tau}(\varphi) = Q^{\tau} \qquad \text{(A2.8)}$$

This property $R^{-1}(\varphi) = R^{\tau}(\varphi)$ is true for any rotation matrix given in the orthonormal coordinate system. With this property the transformations (A2.4-A2.5) become

$$U'_n = Q^{-1} U_n Q \qquad \text{(A2.9)}$$

$$U'^{-1}_n = Q^{-1} U_n^{-1} Q$$

A3. Matrices of linear operators

A transformation of vectors of the three-dimensional space, $\mathbf{q} \rightarrow \mathbf{p}$, is called linear (a linear operator) if for any vectors $\mathbf{q}_1 \rightarrow \mathbf{p}_1$, $\mathbf{q}_2 \rightarrow \mathbf{p}_2$ and for any number λ there are

$$(\mathbf{q}_1 + \mathbf{q}_2) \rightarrow \mathbf{p}_1 + \mathbf{p}_2 \qquad \text{(A3.1)}$$

$$(\lambda \mathbf{q}_1) \rightarrow \lambda \mathbf{p}_1$$

A particular example of a linear transformation is rotation.

In a given coordinate system with the base vectors $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ a linear transformation can be defined by a matrix

$$R = \begin{pmatrix} i_x & j_x & k_x \\ i_y & j_y & k_y \\ i_z & j_z & k_z \end{pmatrix} \qquad \text{(A3.2)}$$

which columns are coordinates of the transformed base vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$, respectively. It is easy to see that the coordinates (p_x, p_y, p_z) of each transformed vector are related to the coordinates (q_x, q_y, q_z) of the initial vector by relation

$$\begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix} = R \begin{pmatrix} q_x \\ q_y \\ q_z \end{pmatrix} \qquad \text{(A3.3)}$$

Note that here (A3.3) links coordinates of different vectors in the same coordinate system while (A1.3) and (A2.1) link coordinates of the same vector but in different coordinate system.

It can be noted that the matrix of the inverse transformation $\mathbf{p} \rightarrow \mathbf{q}$, when this transformation exists, is just the inverse matrix R^{-1}. It follows from (A3.3) and (A2.1) that

$$\begin{pmatrix} p'_x \\ p'_y \\ p'_z \end{pmatrix} = Q^{-1} \begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix} = Q^{-1} R \begin{pmatrix} q_x \\ q_y \\ q_z \end{pmatrix} = Q^{-1} R Q \begin{pmatrix} q'_x \\ q'_y \\ q'_z \end{pmatrix} \qquad \text{(A3.4)}$$

and when changing the coordinate system, the matrix of a linear operator, not necessary a rotation one, is transformed following the rule similar to (A2.10) :

$$R' = Q^{-1} R Q \qquad \text{(A3.6)}$$

Just as a remark we remind the reader that the property $R^{-1}(\varphi) = R^{\tau}(\varphi)$ of the rotation matrices is not necessary conserved for non-orthonormal coordinate systems. For example, for a rotation by $\pi/3$ in the hexagonal coordinate system

$$R_{zH}(\pi/3) = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \; R_{zH}^{-1}(\pi/3) = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \neq R_{zH}^{\tau}(\pi/3) \qquad \text{(A3.7)}$$

A4. Properties of matrices U (trace and symmetry)

Let's consider two square matrices, A with the coefficients α_{jk} and B with the coefficients β_{jk}, $j,k=1,\ldots,K$. We start from a trivial exercise

$$tr(AB) = \sum_{j=1,K} (AB)_{jj} = \sum_{j=1,K} \sum_{k=1,K} \alpha_{jk} \beta_{kj} = \sum_{k=1,K} \sum_{j=1,K} \alpha_{jk} \beta_{kj} =$$

$$= \sum_{k=1,K} \sum_{j=1,K} \beta_{kj} \alpha_{jk} = \sum_{k=1,K} (BA)_{jj} = tr(BA) \qquad \text{(A4.1)}$$

showing a property of the trace of a product of two matrices. As a consequence, for the matrices with the

property (A2.9) or (A3.6)

$$tr(U') = tr(Q^{-1}UQ) = tr(UQQ^{-1}) = tr(U) \quad , \qquad \text{(A4.2)}$$

the trace is conserved when changing the coordinate system.

Another property used in the main text is that for such matrices the symmetry of the matrix $U^{\tau} = U$ is conserved with the rotation of the coordinate system :

$$(U')^{\tau} = (Q^{-1}UQ)^{\tau} = Q^{\tau}U^{\tau}\left(Q^{-1}\right)^{-1} = Q^{-1}UQ = U' \qquad \text{(A4.3)}$$

REFERENCES

Afonine, P.V. & Grosse-Kunstleve, R.W. & Adams, P.D. (2005). "The Phenix refinement framework". *CCP4 Newsletter on Protein Crystallography,* **42**, contribution 8.

Afonine, P.V. & Urzhumtsev, A. (2007). "On determination of *T* matrix in *TLS* modelling". *CCP4 Newsletter on Protein Crystallography,* **45**, http://www.ccp4.ac.uk/newsletters/

Afonine, P.V., Urzhumtsev, A., Grosse-Kunstleve, R.W. & Adams, P.D. (2010). "Atomic displacement parameters (ADPs), their parameterization and refinement in PHENIX".*Cryst.Comput. Newsletters*, **1**, 24-31.

Becka, L.N. & Cruickshank, D.W.J. (1961). "Coordinate errors due to rotational oscillations of molecules". *Acta Cryst.* **14**, 1092-1092.

Birnbaum, G.I. (1972). "The crystal and molecular structure of the trans-syn photodimer of methyl orotate". *Acta Cryst.* B**28**, 1248-1254.

Brenner, H. (1967). "Coupling between the translational and rotational Brownian motions of rigid particles of arbitrary shape". *J.Colloid Interface Chem.* **23**, 407-435.

Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Smart, O., Vonrhein, C. & Womack, T.O. (2009). *BUSTER v.2.8.0.* Global Phasing Ltd, Cambridge.

Bürgi, H.B. (1989). "Interpretation of atomic displacement parameters: intramolecular translational oscillation and rigid-body motion". *Acta Cryst.* B**45**, 383-390.

Burns, D.M., Ferrier, W.G. & McMullan, J.T. (1967). "The rigid-body vibrations of molecules in crystals". *Acta Cryst.* **22**, 623-629.

Busing, W.R., Levy, H.A. (1964). "The effect of thermal motion on the estimation of bond lengths from diffraction measurements". *Acta Cryst.* **17**, 142-146.

Coppens, P. (2006). "The structure factor". *International Tables for Crystallography*, Vol. B, ed. U.Shmueli; Kluwer Academic Publishers; Dordrecht/Boston/London, 10-24.

Cruickshank, D.W.J. (1956a). "The determination of the anisotropic thermal motion of atoms in crystals". *Acta Cryst.* **9**, 747-753.

Cruickshank, D.W.J. (1956b). "The analysis of the anisotropic thermal motion of molecules in crystals". *Acta Cryst.* **9**, 754-756.

Cruickshank, D.W.J. (1956c). "Errors in bond lengths due to rotational oscillations of molecules". *Acta*

Cryst. **9**, 757-758.

Cruickshank, D.W.J. (1961). "The variation of apparent bond lengths with temperature in molecular crystals". *Acta Cryst.* **14**, 896-897.

Downs, R.T., Gibbs, G.V., Barletmehs, K.L. & Boisen, M.B., Jr. (1992). "Variation of bond lengths and volumes of silicate tetrahedra with temperature". *American Mineral.* **77**, 751-757.

Driessen, H., Haneef, M.I.J., Harris, G.W., Howlin, B., Khan, G. & Moss, D.S. (1989). "RESTRAIN: restrained structure-factor least-squares refinement program for macromolecules". *J.Appl. Cryst.,* **22,** 510-516.

Dunitz, J.D. & White, D.N.J. (1973). "Non-rigid-body thermal-motion analysis". *Acta Cryst.* A**29**, 93-94.

Dunitz, J.D. (1979). *X-ray analysis and the structure of organic molecules*, Cornell University Press, Ithaca and London.

Dunitz, J.D., Schomaker, V. & Trueblood, K.N. (1988). "Interpretation of atomic displacement parameters from diffraction studies of crystals". *J.Phys.Chem.* **92,** 856-867.

Dunitz, J.D. (1999). "A curiously short carbon-carbon double bond ?". *Chem.Communication,* 2574-2574.

Goldstein, H. (1950). *Classical Mechanics*. Cambridge, Massachusetts: Addison-Wesley.

Haestier, J., Sadki, M., Thompson, A. & Watkin, D. (2008). "Error estimates on bond-length and angle corrections from *TLS* analysis". *J.Appl. Cryst.* **41**, 531-536.

Haneef, I., Moss, D.S., Stanford, M.J. & Borkakoti, N. (1985). "Restrained structure-factor least-squares refinement of protein structures using a vector-processing computer". *Acta Cryst.* A**41**, 426-433.

Hirshfeld, F.L., Sandler, S. & Schmidt, G.M.J. (1963). "The structure of overcrowded aromatic compounds. VI. The crystal structure of benzo[c]phenanthrene and of 1,12-dimethylbenzo[c]phenanthrene". *J.Chem.Soc.*, 2108-2125.

Holbrook, S.R. & Kim, S.-H. (1984). "Local mobility of nucleic acids as determined from crystallographic data. I. RNA and B form DNA". *J.Molec.Biol.* **173**, 361-388.

Howlin, B., Moss, D.S. & Harris, G.W. (1989). "Segmented anisotropic refinement of bovine ribonuclease A by the application of the rigid-body *TLS* model". *Acta Cryst.* A**45**, 851-861.

Howlin, B., Butler, S.A., Moss, D.S., Harris, G.W. & Driessen, H.P.C. (1993). "TLSANL: *TLS* parameter-analysis program for segmented anisotropic refinement of macromolecular structures". *J.Appl. Cryst.* **26**, 622-626.

Johnson, C.K. (1970a). "The Effect of Thermal Motion on Interatomic Distances and Angles". In *Crystallographic Computing*, ed. F.R.Ahmed, Munksgaard, Copenhagen, 220-226.

Johnson, C.K. (1970b). "Generalized treatments for Thermal Motion". In *Thermal Neutron Diffraction*, ed. B.T.M.Willis, Oxford University Press: London, 132-136.

Johnson, C.K. & Levy, H.A. (1974). "Thermal-Motion Analysis Using Bragg Diffraction Data". *International Tables for X-ray* Crystallography, Vol. IV, eds. J.A.Ibers and W.C.Hamilton; Birmingham: Kynoch Press, 311-336.

Johnson, C.K. (1980). "Thermal motion analysis". In *Computing in Crystallography*, eds. R.Diamond, S.Ramaseshan, K.Venkatesan, Indian Academy of Sciences, Bangalor, India. pp.14.01-14.19

Kuriyan, J. & Weis, W.I. (1991). "Rigid protein motion as a model for crystallographic temperature

factors". *Proc.Natl.Acad.Sci. USA,* **88**, 2773-2777.

Merritt, E.A. (1999). "Comparing anisotropic displacement parameters in protein structres". *Acta Cryst.* D**55**, 1997-2004.

Moore, P.B. (2009) "On the relationship between diffraction Patterns and motion in macromolecular crystals". *Structure,* **17,** 1307-1315.

Moss, D.S., Tickle, I.J., Theis, O. & Wostrack, A. (1996). "X-ray analysis of domain motions in protein crystals". *Proceeding of the CCP4 Study Weekend. Macromolecular refinement*, eds. E.Dodson, M.Moore, A.Ralph, S.Bailey, Warrington: Daresbury Laboratory, 105-114.

Painter, J. & Merritt, E.A. (2005). "A molecular viewer for the analysis of *TLS* rigid-body motion in macromolecules". *Acta Cryst.* D**62**, 439-450.

Painter, J. & Merritt, E.A. (2006a). "Optimal description of a protein structure in terms of multiple groups undergoing *TLS* motion". *Acta Cryst.* D**61**, 465-471.

Painter, J. & Merritt, E.A. (2006b). "TLSMD web server for the generation of multi-group *TLS* models". *J.Appl. Cryst.* **39**, 109-111.

Pawley, G.S. (1963). "On the least-squares analysis of the rigid body vibrations of non-centrosymmetrical molecules". *Acta Cryst.* **16**, 1204-1208.

Pawley, G.S. (1964). "Least-squares structure refinement assuming molecular rigidity". *Acta Cryst.* **17**, 457-458.

Pawley, G.S. (1970). "Rigid-body molecular motion in crystals. The centre of libration". *Acta Cryst.* A**26**, 289-292.

Prince, E. & Finger, L.M. (1972). "Use of constraints on thermal motion in structure refinement of molecules with librating side groups". *Acta Cryst.* B**29**, 179-183.

Rae, A.D. (1975a). "Crystal structure refinement using a number of orthogonal axial systems". *Acta Cryst.* A**31**, 560-570.

Rae, A.D. (1975b). "Rigid-body motion in crystals - the application of constraints on the *TLS* model". *Acta Cryst.* A**31**, 570-574.

Scheringer, C. (1972a). "A lattice-dynamical treatment of the thermal-motion bond-length correction". *Acta Cryst.* A**29**, 616-619.

Scheringer, C. (1972b). "A lattice-dynamical bond-length correction for diatomic and triatomic molecules". *Acta Cryst.* A**29**, 619-628.

Scheringer, C. (1973). "A lattice-dynamics interpretation of molecular rigid-body vibration tensonrs". *Acta Cryst.* A**29**, 554-570.

Scheringer, C. (1978a). "The thermal-motion correction for bond angles". *Acta Cryst.* A**34**, 428-431.

Scheringer, C. (1978b). "Temperature factors for large librations of molecules. Expression in a general crystal metric and for any site symmetry". *Acta Cryst.* A**34**, 702-709.

Scheringer, C. (1978c). "Dynamic density and structure factors for rigid molecules with large librations". *Acta Cryst.* A**34**, 905-908.

Schomaker, V. & Trueblood, K.N. (1968). "On the rigid-body motion of molecules in crystals". *Acta Cryst.*

B**24**, 63-76.

Schomaker, V. & Trueblood, K.N. (1984). *Acta Cryst.* A**40**, C339.

Schomaker, V. & Trueblood, K.N. (1998). "Correlation of internal torsional motion with overall molecular motion in crystals". *Acta Cryst.* B**54**, 507-514.

Stec, B., Zhou, R. & Teeter, M.M. (1995). "Full-matrix refinement of the protein crambin at 0.83 A and 130 K". *Acta Cryst.* D**51**, 663-681.

Steiner, T. & Seanger, W. (1993). "Distribution of observed C-H bond lengths in neutron crystal structures and temperateure dependence of the mean values". *Acta Cryst.* A**49**, 379-384.

Sternberg, M.J.E., Grace, D.E.P. & Phillips, D.C. (1979). "Dynamic information from protein crystallography. An analysis of temperature factors from refinement of the hen egg-white lysozyme structure". *J.Molec.Biol.* **130**, 231-253.

Stuart, D.I. & Phillips, D.C. (1985). "Description of overall anisotropy in diffraction from macromolecular crystals". *Methods in Enzymology*, **115**, 117-142.

Sussman, J.L., Holbrook, S.R., Church, G.M. & Kim, S.-H. (1977). "A structure-factor least-squares refinement procedure for macromolecular structures using constrained and restrained parameters". *Acta Cryst.* A**33**, 800-804.

Sygusch, J. (1976). "Constrained thermal motion refinement for a rigid molecule with librating side groups". *Acta Cryst.* B**32**, 3295-3298.

Tickle, I. & Moss, D.S. (1999). "Probabilistic approach and geometric interpretation of the model and consequences". Notes from IUCr Cryst. Computing School, *http://public-1.cryst.bbk.ac.uk/~tickle/iucr99/iucrcs99.htm.*

Trueblood, K.N., Bürgi, H.-B., Burzlaff, H., Dunitz, J.D., Grammaccioli, C.M., Schulz, H.H., Shmueli, U., Abrahams, S.C. (1996). "Atomic displacement parameter nomenclature. Report of a subcomittee on atomic displacement parameter nomenclature". *Acta Cryst.* A**52**, 770-781.

Urzhumtseva, L.M. & Urzhumtsev, A.G. (1997) "Tcl/Tk based programs. II. CONVROT: program to recalculate different rotation descriptions". *J.Appl. Cryst.,* **30**, 402-410.

Winn, M.D., Isupov, M.N. & Murshudov, G.N. (2001) "Use of *TLS* parameters to model anisotropic displacements in macromolecular refinement". *Acta Cryst.,* **D57,** 122-133.

Zucker, F., Champ, P.C. & Merritt, E.A. (2010). "Validation of crystallographic models contaning *TLS* or other descriptors of anysotropy". *Acta Cryst.* D**66**, 889-900.

SOME OTHER RELEVANT ARTICLES

Aragao, D., Frazao, C., Sieker, L., Sheldrick, G.M., LeGall, J. & Carrondo, M.A. (2003). "Structure of dimeric cytochrome c3 from Desulfovibrio gigas at 1.2 A resolution". *Acta Cryst.* D**59**, 644-653.

Arnoux, B., Ducruix, A. & Prange, T. (2002). "Anisotropic behavious of the C-terminal Kunitz-type domain of the a3 chain of human type VI collagen at atomic resolution (0.9 A)". *Acta Cryst.* D**58**, 1252-1254.

Artymiuk, P.J., Blake, C.C.F., Grace, D.E.P., Oatley, S.J., Phillips, D.C., Sternberg, M.J.E. (1979). "Crystallographic studies of the dynamic properties of lysozyme". *Nature*, **280**, 563-568.

Blessing, R. (1986). "Hydrogen bonding and thermal vibrations in crystalline phosphate salts of histidine

and imidazole". *Acta Cryst.* B**42**, 613-621.

Bloch, F. (1932). "Zur Theorie des Austauschproblems und der Remanenzerscheinung der Ferromagnetika" ("On the theory of the exchange problem and the remanence phenomenon of ferromagnets"). *Zeitschrift für Physika.* **74**, 295-335.

Bürgi, H.B. & Capelli, S.C. (2000) "Dynamics of molecules in crystals from multitemperature anisotropic displacement parameters. I. Theory". *Acta Cryst.* A**56**, 403-412.

Byrom, P.G., Hoffmann, S.E. & Lucas, B.W. (1989). "MORGUE, a new powder diffraction profile refinement program with control-file factility to include structural and rigid-body thermal-motion constraints". *J.Appl. Cryst.* **22**, 629-633.

Capelli, S.C., Fortsch, M. & Bürgi, H.B. (2000) "Dynamics of molecules in crystals from multitemperature anisotropic displacement parameters. II. Application to benzene (C6D6) and urea [OC(NH)2]". *Acta Cryst.* A**56**, 413-424.

Chaudhry, C., Horwich, A.L., Brunger, A.T. & Adams, P.D. (2004). "Exploring the structural dynamics of the E.coli chaperonin GroEL using translation-libration-screw crystallographic refinement of intermediate states". *J.Molec.Biol.* **342**, 229-245.

Cochran, W. (1951a). "The structures of pyrimidines and purines. V. The electron density distribution in adenine hydrochloride". *Acta Cryst.* **4**, 81-92.

Cochran, W. (1951b). "Some Properties of the (Fo-Fc)-Synthesis". *Acta Cryst.* **4**, 408-411.

Diamond, R. (1990). "On the Use of Normal Modes in Thermal Parameter Refinement: Theory and Application to the Bovine Pancreatic Trypsin Inhibitor". *Acta Cryst.* A**46**, 425-435.

Ducros, V.M.-A., Lewis, R., Verna, C.S., Dodson, E.J., Leonard, G., Turkenburg, J.P., Murshudov, G.N., Wilkinson, A.J. & Brannigan, J.A. (2001). "Crystal structure of GerE, the ultimate transcriptional regulator of spore formation in Bacillus subtilis". *J.Molec.Biol.* **306**, 759-771.

Garcia, P., Dahaoui, S., Fertey, P., Wenger, & E., Lecomte, C. (2005). "Crystallographic investigation of temperature-induced phase transition of the tetrathiafulvalene-p-bromanil, TTF-BA charge transfer complex". *Phys.Rev.* B, **72**, 104115.

Harata, K., Abe, Y. & Muraki, M. (1999). "Crystallographic evaluation of internal motion of human alpha-lactalbumin refined by full-matrix least-squares method". *J.Molec.Biol.* **287**, 347-358.

Harata, K., Abe, Y. & Muraki, M. (1998). "Full-matrix least-squares refinement of lysozymes and analysis of anisotropic thermal motion". *Proteins Struct. Funct. Genet.* **30**, 232-243.

Harata, K. (2003). "Crystallographic analysis of the thermal motion of the inclusion complex of cyclomaltoheptaose (beta-cyclodextrin) with hexamethylentetramine". *Carbohydrate Res.* **338**, 353-359.

Harris, G.W., Pickersgill, R.W., Howlin, B. & Moss, D.S. (1992). "The segmented anisotropic refinement of monoclinic papain by the application of the rigid-body *TLS* model and comparison to bovine ribonuclease A". *Acta Cryst.* B**48**, 67-75.

Hirshfeld, F.L. & Shmueli, U. (1972). "Covariances of thermal parameters and their effect on rigid-body calculations". *Acta Cryst.* A**28**, 648-652.

Holbrook, S.R., Dickerson, R.E. & Kim, S.-H. (1985). "Anisotropic thermal parameter refinement of the DNA dodecamer CGCGAATTCGCG by the segmented rigid-body method". *Acta Cryst.* B**41**, 255-262.

Hummel, W., Raselli, A. & Burgi, H.-B. (1990). "Analysis of atomic displacement parameters and molecular motion in crystals". *Acta Cryst.* B**46**, 683-692.

Johnson, C.K. (1970). "An Introduction to Thermal Motion Analysis". In *Crystallographic Computing*, ed. F.R.Ahmed, Munksgaard, Copenhagen, 207-219.

Kidera, A. & Go, N. (1990). "Refinement of protein dynamic structure: normal mode refinement". *Proc.Natl.Acad.Sci. USA,* **87**, 3718-3722.

Kidera, A. & Go, N. (1992). "Normal mode refinement: crystallographic refinement of protein dynamic structure. I. Theory and test by simulated diffraction data". *J.Molec.Biol.* **225**, 457-475.

Kidera, A., Matsushima, M. & Go, N. (1994). "Dynamic structure of human lysozyme derived from X-ray crystallography: normal mode refinement". *Biophys.Chem.* **50**, 25-31.

Merritt, E.A. (1999). "Expanding the model: anisotropic displacement parameters in protein structure refinement". *Acta Cryst.* D**55**, 1109-1117.

Moroz, O.V., Antson, A.A., Murshudov, G.N., Maitland, N.J., Dodson, G.G., Wilson, K.S., Skibshoj, I., Lukanidin, E.M., Bronstein, I.B. (2001) "The three-dimensional structure of human S100A12". *Acta Cryst.,* **D57,** 20-29.

Murshudov, G.N., Vagin, A.A., Lebedev, A., Wilson, K.S. & Dodson, E.J. (1999) "Efficient anisotropic refinement of macromolecular structures using FFT". *Acta Cryst.,* **D55,** 247-255.

Papiz, M.Z. & Prince, S.M. (1996). "Group anisotropic thermal parameter refinement of the light-harvestingcomplex from purple bacteria *Rhodopseudomonas acidophila*". *Proceeding of the CCP4 Study Weekend. Macromolecular refinement*, eds. E.Dodson, M.Moore, A.Ralph, S.Bailey, Warrington: Daresbury Laboratory, 115-123.

Papiz, M.Z. & Prince, S.M. (2003). "The structure and thermal motion of the B800-850 LH2 complex from Rps.acidophila at 2.0 A resolution and 100 K: new structural features and functionally relevant motions". *J.Molec.Biol.* **326**, 1523-1538.

Pawley, G.S. (1965). "Refinement of azulene assuming rigid-body thermal motion". *Acta Cryst.* **18**, 560-561.

Pawley, G.S. (1966). "Further refinements of some rigid boron compounds". *Acta Cryst.* **20**, 631-638.

Pawley, G.S. (1968). "Anisotropic temperature factors and screw rotation coefficients from a lattice dynamical viewpoint". *Acta Cryst.* B**24**, 485-486.

Pawley, G.S. (1970). "The use of molecular lattice dynamical motion". *Crystallographic Computing*, eds. Ahmed, F.R., Hall, S.R., Huber, C.P., Munksgaard, Copenhagen, 243-249.

Perez, J., Faure, P. & Benoit, J.-P. (1996). "Molecular rigid-body displacements in a tetragonal lysozyme crystal confirmed by X-ray diffuse scattering". *Acta Cryst.* D**52**, 722-729.

Phillips, C., Gover, S. & Adams, M.J. (1995). "Structure of 6-phosphogluconate dehydrogenase refined at 2 A resolution". *Acta Cryst.* D**51**, 290-304.

Raaijmakers, H., Toro, I., Birkenbihl, R., Kemper, B. & Suck, D. (2001). "Conformational flexibility in T4 endonuclease VII revealed by crystallography: implications for substrate binding and cleavage". *J.Molec.Biol.* **308**, 311-323.

Sali, A., Veerapandian, B., Cooper, J.B., Moss, D.S., Hofmann, T. & Blundell, T.L. (1992) "Domain flexibility in aspartic proteinases". *Proteins Struct. Funct. Genet.* **12**, 158-170.

Sarma, G.N., Savvidis, S.N., Becker, K., Schirmer, M., Schirmer, R.H. & Karplus, P.A. (2003). "Glutathione reductase of the malarial parasite Plasmodium falciparum: crystal structure and inhibitor development". *J.Molec.Biol.* **328**, 893-907.

Schneider, T.R. (1996). "What we can learn from anisotropic temperature factors?". *Proceeding of the CCP4 Study Weekend. Macromolecular refinement*, eds. E.Dodson, M.Moore, A.Ralph, S.Bailey, Warrington: Daresbury Laboratory, 133-144.

Sternberg, M.J.E., Grace, D.E.P. & Phillips, D.C. (1979). "Dynamic information from protein crystallography. An analysis of temperature factors from refinement of the hen egg-white lysozyme structure". *J.Molec.Biol.* **130**, 231-253.

Trueblood, K.N. (1978). "Analysis of molecular motion with allowance for intramolecular torsion". *Acta Cryst.* A**34**, 950-955.

Verlinde, C.L. & De Ranter, C.J. (1989) "Furan revisited : when to avoid ab initio studies on crystal structures". *J.Molec.Struct. (Therochem.)*, **187**, 161-167.

Wilson, C. (2000) "Single crystal neutron diffraction from molecular materials". *Singapore : World Scientific Publishing*.

Wilson, M.A. & Brunger, A.T. (2000) "The 1.0 A crystal structure of Ca2+ -bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity". *J.Molec.Biol.* **301**, 1237-1256.

Wilson, M.A. & Brunger, A.T. (2003) "Macromolecular *TLS* refinement in REFMAC at moderate resolutions". *Acta Cryst.* D**59**, 1782-1792.

Winn, M.D., Murshudov, G.N. & Papiz, M.Z. (2003) "Macromolecular *TLS* refinement in REFMAC at moderate resolutions". *Methods in Enzymology,* **374,** 300-321.

Yousef, M.S., Fabiola, F., Gattis, J.L., Somasundaram, T. & Chapman, M.S. (2002) "Refinement of the arginine kinase transition-state analogue complex at 1.2 A resolution: mechanistic insights". *Acta Cryst.* D**58**, 2009-2017.