Tutorial 5 - Rigid Bodies

Overview

In this tutorial you will model data of a molecular solid, paracetamol, at room temperature and low temperature. A "rigid body" will be constructed and used to describe the well-defined geometry of the molecule. Different ways of describing the differences between the PDF peak shapes from intermolecular (between molecules) and intramolecular (within molecular) correlations will be investigated.

Required files

- para_300K_pdf.xy
- para_100K_pdf.xy
- para_partial.rgd
- para.cif

Introduction to rigid bodies in TOPAS

TOPAS allows for the description of molecular fragments as rigid bodies, which can then be translated and rotated within the unit cell. In the case where a molecular geometry is well defined and known, the use of rigid bodies can dramatically improve the quality of a refinement over the free refinement of x, y and z coordinates for each atom. There are two main ways in which molecular fragments can be described in TOPAS; `point_for_site` and `z_matrix`.

point_for_site

`point_for_site` allow fragments described in Cartesian coordinates to be moved around within a unit cell. This can work well where there is a very fixed geometry that you need to move within a unit cell. However, refining things like bond lengths are not easy to do. We will not address `point_for_site` further in this tutorial, but more details can be found in the 'Rigid bodies and bond length restraints' section of the TOPAS Technical Reference manual.

z_matrix

A more flexible way of describing molecules using more chemically descriptive internal coordinates for the molecular is to use `z_matrix`. This allows for the description of a molecule in terms of its bond lengths and angles. The lengths and angles can be parameters which form part of the refinement or used within penalties. While the use of `z_matrix` can take some time to get used to, it is ultimately a much more powerful method of describing and refining (flexible) rigid bodies.

The keyword `rigid` starts the definition of a rigid body in TOPAS. Each atom site that you want to include in the rigid body requires a `z_matrix` line to describe its position. For example, a simple tetrahedra is defined and described below:

`rigid`	This defines the start of a rigid body
`z_matrix C`	The first site defines the origin of the rigid body. Translations will act directly on this point, and rotations will act around this point.
`z_matrix H1 C 1.0`	The second site requires a distance to be defined (along positive the z-axis). This states that H1 is 1.0 Å away from atom C.
`z_matrix H2 C 1.0 H1 109.5`	The third site requires a distance and an angle to be defined (in the x-z plane). This states that H2 is also 1.0 Å away from atom C, and the H2–C–H1 angle is 109.5°.
`z_matrix H3 C 1.0 H1 109.5 H2 120`	The fourth and all subsequent sites also require a dihedral (torsion) angle to be defined. Here the angle between the planes defined by H3-C-H1 and H3-C-H2 is 120°.
`z_matrix H4 C 1.0 H1 109.5 H3 120`	Note that this could equally be described as `z_matrix H4 C 1.0 H1 109.5 H2 240`.

As molecular become more complicated, in particular in the case of largely planar molecules, the requirement of a unique description of the dihedral angle between sites can become problematic. For this reason, it is common to use 'dummy atoms' to define a local axis system around which to describe the rigid body. These dummy atoms are sites with zero scattering from which the rigid body is defined.

Some more general information on Z matrices (not specific to TOPAS) with some examples can be found here.

Tutorial instructions

Theses PDF data of paracetamol were collected on the XPDF beamline at Diamond. The structure of paracetamol can be found here.

Fitting without the use of a rigid body

1. Start a fresh PDF refinement using `"para_100K_pdf.xy"` as the filename *(TOPASforPDF > 1. PDF data > Select PDF Data File)* and in jEdit save the file as 'para_100K_without_rigid.inp'.
2. Enter a dQ damping with a value of 0.08 *(TOPASforPDF > 2. Instrumental parameters > dQ damping)*.
3. Load structure from para.cif *(TOPASforPDF > 3. Phase information > 3b. add new phase from CIF > i. Read a .CIF File)*.
4. Allow the lattice parameters to refine within the monoclinic symmetry *(TOPASforPDF > 3. Phase information > 3b. add new phase from CIF > ii. constrain lattice parameters > convert to monoclinic)*.

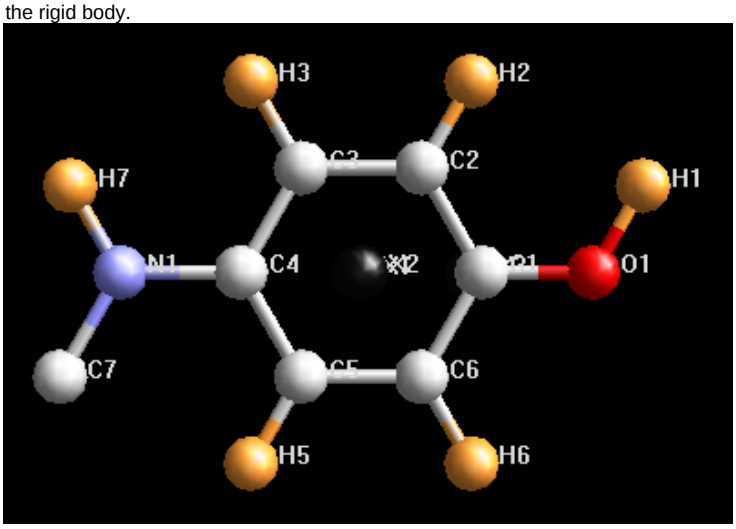
5. Delete the `beq 0.0789` for each atom, and instead add a beq_rcut_rlo_spherical peak shape with the default values; we will look at different peak shape functions in more details later on in the tutorial *(TOPASforPDF > 3. Phase information > beq peak shape functions > beq spherical with min r and low r cutoffs)*.
6. Add view_structure to open a structure viewer where you can watch the structure as the refinement progresses *(TOPASforPDF > 3. Phase information > view_structure).*
7. Include output to CIF so you can compare the results in VESTA later *(TOPASforPDF > 3. Phase information > outputs > output .CIF file)*.
8. Allow all of the atomic coordinates to refine (place an `@` sign after `x`, `y` and `z` in each site) and run the refinement in TOPAS, but when asked if you want to update the .inp file with the .out file at the end of the refinement click **No**. Use the structure viewer to look at how the atoms move during the refinement.
    a. How has the molecular geometry changed?
    b. Use VESTA to look at the .cif file created, and compare it to the starting para.cif.
9. Edit the sites of the structure by adding `Damp(5)` after each fractional coordinate value (use Rectangular Selection in jEdit to edit over multiple lines). Each site line should now look something like: `site C1 x @ 0.1286 Damp(5) y @ 0.1743 Damp(5) z @ -0.0080 Damp(5) occ C 1.0`.

> The Damp macro is defined in local.inc. It only applies a small proportion of the total calculated value change from the least squares refinement. This can help to keep atomic positions near their starting positions in the early stages of a refinement, but should be removed before finalising a refinement as it can prevent a refinement from converging.
>
> ```
> macro Damp(DampFactor) { update = Val + (Change*(1/Exp(DampFactor/2))); }
> ```

10. Change the filename for the output .cif to `Out_CIF_STR("para_100K_without_rigid_damp.cif")` so the previous one doesn't get overwritten.
11. Re-refine the structure and use the structure viewer to look at how the atoms move during the refinement.
    a. Use VESTA to compare the output .cif files.

Creating a rigid body

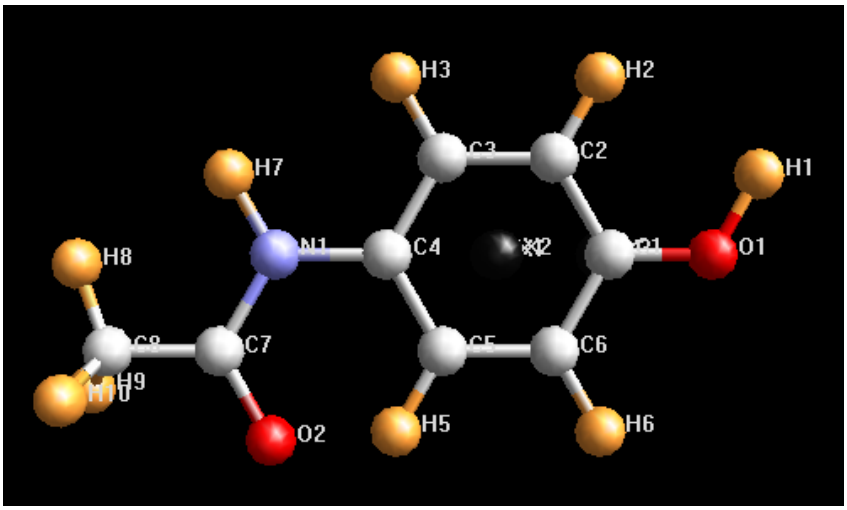
TOPAS includes a 'rigid-body editor' which can be used to create and test rigid bodies. There are some examples of rigid bodies in the 'C:\TOPAS-6\rigid' folder; if at any point you get stuck while editing the `z_matrix` for this section, it may help to look at some of the examples in this folder.

1. In TOPAS, select 'Tools' > 'New Rigid-body editor Window'.
2. Click '**Load / Hide**' to bring up the file load dialogue, navigate to the file 'para_partial.rgd' and double click on it to open it in the editor. This is a partially completed `z_matrix` description of an idealised paracetamol molecule. Notice how bond lengths (`prm !rCC 1.30`, `prm !rCO 1.20` etc.) and dihedral angles (`prm !dCOH 90`, `prm !dCNC 90` etc.) are defined as parameters at the top of the rigid body. These parameters also have sensible `min` and `max` values applied to them. Try changing a few of these parameters and click **Update** to see the effect on the rigid body.
    a. 

3. Edit the `z_matrix` to add the C=O and CH3 groups on the end of the molecule. Use the naming scheme shown below (i.e. add O2, C8, H8, H9 and H10).

> It is best to add one site at a time, then click **Update** to see the effect on the rigid body. This will allow you to fix problems as you go.

The completed structure should look like this:

a. 

4. If you are having problems adding atoms to the rigid body, there are some hints here.

   First we need to add the O2 atom, which is connected to C7 by a bond distance with a parameter name of `rCO`. It is at an angle of 240° from atom N1, and we can use the dummy atom X2 to define a dihedral angle of 180 degrees.

   Next we need to add the C8 atom, which is connected to C7 by a bond distance with a parameter name of `rCC`. It is at an angle of 120° from atom N1, and we can again use the dummy atom X2 to define a dihedral angle of 180 degrees.

   Finally we need to add three H atoms (H8, H9 and H10), all connected to C8 by a distance of parameter `rCH`. They form roughly a tetrahedron, so they will all be at an angle of 109.5° from atom C7. It is useful to refine a rotation angle for this CH3 group (e.g. parameter name `aCH3`) and have the free dihedral angles all relative to X2 with values of `dCH3`, `dCH3+120` and `dCH3+240`.

5. Click '**Save As**' and save your new .rgd file of the completed paracetamol rigid body on your computer.

Matching a rigid body to a known structure

Although we now have a rigid body description of the paracetamol molecule, we don't yet know the translations / rotations needed to get the molecule where it should be within the unit cell. We can use TOPAS to optimise the position of the paracetamol rigid body to that of the known atomic positions from a .cif file; thanks to Prof Simon Parsons, University of Edinburgh, for this tip! In order for us to do this, it is easiest if the naming conventions in our rigid body are the same as those from our .cif. The site names in the rigid body described above already match those of the .cif file, so in this case we can carry out.

1. Start a fresh PDF refinement using "para_300K_pdf.xy" as the filename (TOPASforPDF > 1. PDF data > Select PDF Data File) and save the input file as 'para_300K_optimise.inp'. We will use this file to optimise the geometry of the rigid body.
2. Enter a dQ damping with a value of 0.08 *(TOPASforPDF > 2. Instrumental parameters > dQ damping).*
3. Load structure from para.cif *(TOPASforPDF > 3. Phase information > 3b. add new phase from CIF > i. Read a .CIF File).*
4. Add the rigid body information to the file, which will be used to calculate the fractional coordinates of the sites *(TOPASforPDF > 3. Phase information > rigid bodies > read an .RGD File)*.
   a. Also allow the translation and rotation of the rigid body to be refined (*TOPASforPDF > 3. Phase information > constraints and restraints > rigid bodies > translate rigid body* and *rotation rigid body*).
5. If you run a TOPAS refinement now, you will get an error of 'Cannot find site: X1'. That is because the dummy atoms X1, X2 and X3 are in the rigid body description, but aren't declared as sites. Add one site line for each dummy atom, but with zero site occupancy.

   ```
   site X1 x 0 y 0 z 0 occ H 0
   ```

6. Create a copy of all of the real atom sites (not X1, X2 and X3), but re-name these sites `dummyC1`, `dummyC2` , `dummyC3` etc. and set their occupancy to zero. These will be dummy sites which we will use to move the rigid body sites onto the original atomic positions from the cif file.
7. Add a distance restraint (*TOPASforPDF > 3. Phase information > constraints and restraints > distance restraint*). Edit the line to apply the restraint between sites `dummyC1` and `C1` for a distance of `0` and tolerance of `0`.

   > Distance_Restrain creates a penalty for when two sites are beyond a given distance. Here we are using to minimise the distance between two sites, but normally it would be used to force a bond length towards a known value.

8. Use Rectangular Selection in jEdit to copy and edit the file to give you one Distance_Restrain line per dummy site.
9. Add `view_structure` so you can watch how the refinement progresses.
10. Add the keyword only_penalties to your input file. The tells TOPAS not to fit to the data, and instead only minimise the distance penalties.
11. Run the refinement, and watch the rigid body move onto the atomic coordinates from the .cif file using the structure viewer. When the refinement finishes, click **Yes** to update the .inp file with the values from the refinement.
    a. How well does the rigid body fit to the published crystal structure from the .cif file?
    b. Are there other angles within the rigid body that you could refine to improve the fit?
12. The C7-N1-C4 bond angle is the source of the largest error between the published crystal structure and the idealised rigid body where it is fixed to 120°. Add a new refined bond angle, e.g. `aCNC`, and set the angle between C7, N1 and C4 to be equal to this parameter.

    > Note that you can refine rigid body parameters by putting an @ sign before the value, but when you may re-use the parameter elsewhere it is best to provide your own name for it.

13. Run the refinement again, and click **Yes** to update the .inp file and complete this refinement.

You now have the translate and rotate values needed for a good starting point for your PDF refinement in the next part of this tutorial.

PDF fitting using a rigid body

You will now have all the pieces you need to fit a PDF using a rigid body.

1. Start a fresh PDF refinement using "para_100K_pdf.xy" as the filename (*TOPASforPDF > 1. PDF data > Select PDF Data File*) and save the file as 'para_100K_rigid.inp'.
2. Enter a dQ damping with a value of 0.08 *(TOPASforPDF > 2. Instrumental parameters > dQ damping).*
3. Load structure from para.cif *(TOPASforPDF > 3. Phase information > 3b. add new phase from CIF > i. Read a .CIF File).*
4. Allow the lattice parameters to refine within the monoclinic symmetry *(TOPASforPDF > 3. Phase information > 3b. add new phase from CIF > ii. constrain lattice parameters > convert to monoclinic).*
5. Delete the `beq 0.0789` for each atom, and instead add a beq_rcut_rlo_spherical peak shape with the default values; we will look at different peak shape functions in the next section of this tutorial *(TOPASforPDF > 3. Phase information > beq peak shape functions > beq spherical with min r and low r cutoffs).*
6. Add one site line for each dummy atom in the rigid body (X1, X2 and X3), but with zero site occupancy.
7. Copy the rigid body information from 'para_300K_optimise.inp', including everything from the keyword `rigid` down to the refined values for `Translate` and `Rotate_about_axies`, and paste it into your current .inp file below your sites.
8. Add `view_structure` so you can watch how the refinement progresses.
9. Run the refinement and observe the fit to the PDF. You should get an ok fit with with an `Rwp` of approximately 25.6.

Thermal parameters and molecules

In the previous section you obtained a rigid body fit to paracetamol (it's not completely rigid because you refined some bond distances and angles). In this section we will improve on the fit using thermal parameters.

1. Open 'para_100K_rigid.inp' and save it as something like 'para_100K_rigid_beq_spherical.inp'. So far we have only used the default values for the `beq_rcut_rlo_spherical` peak shape, and all of the atoms have the same peak shape.
2. Check the `rcut` and `rlo` values. These are usefully fixed to distances after the first and second peaks in your PDF respectively. The `rcut` value of 1.0 looks ok, but change `rlo` to 1.7 and run the refinement. Since the starting values were already pretty good it should give a negligible improvement to the Rwp.
3. Currently all of the sites have the same peak shape function (`beq_rcut_rlo_spherical`) with the same parameters, but different atoms are likely to be moving by different amounts. Allow different values for `beqlo` and `beqhi` parameters for each of the different atom types (C, N, O and H) by giving them different parameter names (`beqloC`, `beqloN`, `beqloO` etc.). Run the refinement, and it should improve the fit a bit more and give an Rwp of around 25.3.

   > You should only ever refine a single `beqcut` value for all atom types.

4. Refining just the atom types may not be enough for a flexible molecule. Remember, the peak width is due to how much the two atoms contributing to that peak are moving relative to each other. Explore whether allowing different parts of the molecule to have different `beqlo` and `beqhi` values improves the fit significantly.

The `beq_spherical` peak shapes don't distinguish between inter- and intra-molecular distances, which will often have quite radically different peak shapes. TOPAS allows you to use `pdf_for_pairs` to specify the width of certain pairs and to differentiate between the first instance of an atom and subsequent instances of that atom. How we will use `pdf_for_pairs` and see if you can improve on the above fit.

1. Rename and save your previous .inp file as 'para_100K_pdf_for_pairs.inp'.
2. Comment out all of the `beq_rcut_rlo_spherical` parts of each site.
3. Add the line `pdf_for_pairs C* C* pdf_only_eq_0 pdf_gauss_fwhm @ 0.1 min 0.1 max 2` for each atom type, i.e. C* C*, C* N*, C* O* etc. There should be 10 types in total. The keyword `pdf_only_eq_0` tells TOPAS that this peak shape only applies to the first position for that site, i.e. if there are multiple sites with the unit cell (such as ones belonging to other molecules) it will have a different peak shape.
4. Add the line `pdf_for_pairs C* C* pdf_gauss_fwhm @ 0.5 min 0.1 max 10` for each atom type. This is the peak shape that will be used for all other correlations beyond the first.
5. Add the keyword `pdf_info` at the bottom of the file and run the refinement. After running the refinement, this will print out a list of all of the peak shapes defined using `pdf_for_pairs` and is very useful for error checking.
   a. How do the peaks at low *r* compare with those at high *r*?
   b. How does this relate to inter- and intra-molecular interactions?
6. The fit is not great, because currently we only have a single value for all inter-molecular bonds, rather than one that varies as a function of *r*. Set up the following parameters:

```
prm loC 0.1 min 0 max 10 del 0.001
prm loN 0.1 min 0 max 10 del 0.001
prm loO 0.1 min 0 max 10 del 0.001
prm loH 0.1 min 0 max 10 del 0.001
prm hiC 0.1 min 0 max 10 del 0.001
prm hiN 0.1 min 0 max 10 del 0.001
prm hiO 0.1 min 0 max 10 del 0.001
prm hiH 0.1 min 0 max 10 del 0.001
```

7. Now change each of the '`pdf_for_pairs C* C* pdf_gauss_fwhm`' lines (the ones without `pdf_only_eq_0`) for:

```
pdf_for_pairs C* C* pairs_spherical(!rlo,1.7, loC+loC , hiC+hiC ,beqradius,10)
```

   `pairs_spherical` is a similar function to `beq_rlo_spherical`, only it is for use within `pdf_for_pairs`.
8. Run the refinement and look at the results.
   a. How does this result compare to the one obtained solely with the `beq_spherical` peak shape?

b. The example of paracetamol is possibly not the best to demonstrate the use of `pdf_for_pairs` as it is a moderately spherical molecule. How might this be different for something like anthracene or naftacene?

A dataset at 300 K is also provided called para_300K_pdf.xy. Run a refinement using this data and compare the results with that of the 100 K dataset. How does reducing the temperature help to resolve structural details in organic systems?

Final comments

Rigid bodies don't need to be used just for organic molecules. If you have well defined tetrahedra or octahedra in an inorganic structure it can sometimes help to describe them as rigid bodies. It also allows for a much more direct way of refining geometries - bond lengths and angles become refinable parameters, rather than just xyz coordinates of individual atoms.

For researchers modelling rigid bodies, it is worth getting very familiar with the different ways of modelling peak shapes using the `beq_spherical` and `pdf_for_pairs` methods. Remember that the best refinement is one which achieves the best fit to the data with as few parameters as possible.