DiTe 2 v1.0
User Guide

Jonathan Campeggio, Antonino Polimeno, Mirco Zerbetto
Theoretical Chemistry Group – University of Padua – Italy

Index

1. Installation

To install DiTe2, just unpack the tgz package file

```
tar zxf DiTe2-v1.0.tgz
```

step into the DiTe2-v1.0/ folder, and run the build.sh script

```
cd DiTe2-v1.0
./build.sh
```

following the instructions printed on screen. The script will first compile the armadillo, zmatlib, and dite2lib libraries located in the lib sub-folder, and then will build a standalone version of DiTe2, located in the src sub-folder.

Two scripts will be generated in the DiTe2-v1.0/ directory: build_zmatrix and dite2_run, which are those to be used in running the standalone program.

Lapack and BLAS (or OPENBLAS, ATLAS, CBLAS, ...) libraries are required.

If the Rotne-Prager tensor and/or the constraints matrix are needed, they can be printed on the standard output by adding the options -DOUTPUT_SMALL_D and/or -DOUTPUT_BB to the end of line 13 of the Makefile located in DiTe2-v1.0/lib/dite2lib/ folder.

1.1 Troubelshooting

If the build.sh script fails in compiling DiTe2 and the errors refer to BLAS libraries, modify line 66 of the build.sh script removing the '#' character, i.e.

```
LIBS=-std=c++11 -ldite2 -lzmat -larmadillo -D_ARMA_USE_LAPACK # -llapack -lblas
```

should read

```
LIBS=-std=c++11 -ldite2 -lzmat -larmadillo -D_ARMA_USE_LAPACK -llapack -lblas
```

2. Usage

DiTe2 is used via command line. The work-flow is typically divided in three steps:

1. Run the build_zmatrix script passing a minimal input file: `build_zmatrix input.inp`
2. Use the Z-matrix produced in the previous step to define the internal coordinates and complete the input file.
3. Calculate the friction and diffusion tensors using the dite2_run script: `dite2_run input.inp`

The structure of the `input.inp` file is the following, explained line by line:

Line 1: PDB file name with molecular structure

```
pdb file.pdb
```

Line 2: ID of reference atoms with numbering based on the original PDB file. At present, 4 atoms should be indicated, even if only the first 3 atoms are used to define the frame.

```
refatoms 1 2 3 4
```

Lines 3-4: hydrodinamic data, i.e. effective beads radius in Å, boundary conditions, viscosity in Pa s, and temperature in K.

```
reff 2.0
C 6.0
viscosity 8.9e-4
temperature 298.15
```

Line 5: additional bonds can be specified if atoms are too far away (e.g., in a PDB with two or more molecules) and the Zmatlib library is not able to automatically connect them. Just specify the number of manual bonds, Nb, followed by Nb couples of atoms ID's in the original numbering of the PDB file.

```
bonds Nb ID(1,1) ID(1,2) ... ID(Nb,1) ID(Nb,2)
```

These first 5 lines are sufficient for the minimal input file to run the build_zmatrix script.

Line 6: specify the coordinate `type = simple` or `collective`, and the number of such coordinates.

```
coordinate type N
```

If type = simple, lines 7-(N+6): this part is usually prepared after the build_zmatrix script has been run with the information above. At this point, atoms are renumbered and the Z-Matrix is provided in output. To specify an internal coordinate, insert the atom ID in the new numbering scheme and the type of simple coordinate: 1 = bond length, 2 = bond angle, 3 = dihedral angle. Moreover, the user must specify if a scan along that coordinate should be performed or not.

```
atomID(1) qType(1) [0 | n(1) first(1) last(1)]
...
atomID(N) qType(N) [0 | n(N) first(N) last(N)]
```

where n(j) is the number of points of a scan of the j-th coordinate from the first to the last value.

In case of a scan, the fist and last values of the coordinates can be different from that in the PDB file specified in the first line of the input file.

WARNING: DO NOT INSERT TWO DIHEDRAL ANGLES THAT SHARE THE SAME CENTRAL ATOMS.

If type = collective, lines 7-(N+6): this part is usually prepared after the build_zmatrix script has been run with the information above. At this point, atoms are renumbered and the Z-Matrix is provided in output. To specify a collective coordinate, insert the number, m, of simple coordinates on which the collective one depends, followed by m sequences of three numbers corresponding to: the atomID (from the Z-Matrix, see the simple coordinate), the type of single coordinate (1 = bond length, 2 = bond angle, 3 = dihedral angle), and the coefficient of the linear combination. Moreover, the user must specify if a scan along that coordinate should be performed or not.

```
m(1) atomID(1,1) qID(1,1) C(1,1) ... atomID(m,1) qID(m,1) C(m,1) [0 | n(1) first(1) last(1)]
...
m(N) atomID(1,N) qID(1,N) C(1,N) ... atomID(m,N) qID(m,N) C(m,N) [0 | n(N) first(N) last(N)]
```

WARNING: differently from simple coordinates, in case of a scan along a collective coordinate the first point is assumed as the value of the coordinate in the PDB file.

3. Output

If the PDB file mol.pdb was specified in the input file, DiTe2 will output the following files:

- mol.pdb.zmatrix.zmt: the Z-Matrix built from the Zmatilb library
- mol.pdb.rebuild.xyz: Cartesian coordinates in the molecular frame (see below)
- mol.pdb.j.xyz: Cartesian coordinates of the j-th conformation of the scan process
- mol.pdb.diffusion.j: diffusion tensor in the j-th conformation of the scan process
- mol.pdb.friction.j: friction tensor in the j-th conformation of the scan process
- mol.pdb.scan.log: values of the coordinates of the conformations generated along the scan

3.1 Geometry

The friction and diffusion tensors are given in a molecular frame that has its origin on the 2nd reference atom, the X axis is defined along the line connecting the 1st and 2nd reference atoms, with the 1st reference atom laying in the negative X. The Z axis is perpendicular to the plane defined by the 1st, 2nd, and 3rd reference atoms. Finally, the Y axis is obtained from X and Z.

3.2 Units

The friction and diffusion tensors are divided into pure translational (TT), rotational (RR), and internal-conformational (II) diagonal blocks, and mixed TR, TI, and RI out of diagonal blocks.

Friction tensor

	T	**R**	**I (bonds)**	**I (angles, dihedrals)**
T	N s m^{-1}	10^{10} N s	N s m^{-1}	10^{10} N s
R		10^{20} N m s	10^{10} N s	10^{20} N m s
I (bonds)			N s m^{-1}	10^{10} N s
I (angles, dihedrals)				10^{20} N m s

Diffusion tensor

	T	**R**	**I (bonds)**	**I (angles, dihedrals)**
T	m^2 s^{-1}	10^{-10} m s^{-1}	m^2 s^{-1}	10^{-10} m s^{-1}
R		10^{-20} s^{-1}	10^{-10} m s^{-1}	10^{-20} s^{-1}
I (bonds)			m^2 s^{-1}	10^{-10} m s^{-1}
I (angles, dihedrals)				10^{-20} s^{-1}

4. Examples

In the DiTe2-v1.0/exmaples sub-folder, a number of examples are provided. They are briefly commented in what follows.

4.1 Dialanine simple coordinates scan

This example is in the DiTe2-v1.0/examples/2alanine/2alanine_simple/ folder. The input file reads

```
pdb 2ala.pdb
refatoms 7 2 1 17
reff 2.0
C 6.0
viscosity 8.9e-4
temperature 298.15
coordinates simple 2
4 3 3 -99 -59
16 1 3 1.4 1.6
```

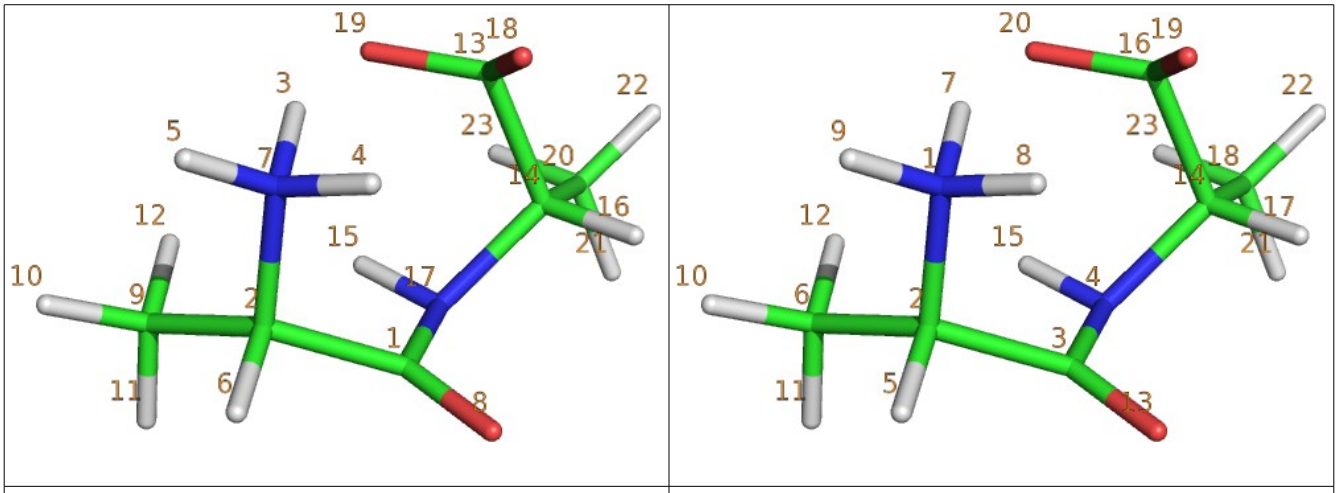
The first line specifies the 2ala.pdb file contained in the folder. The atoms N(7), C(2), C(1), and N(17) are used to build the molecular frame. It is here recalled that the numbering here is that of the PDB file. Hydrodynamic parameters are: effective radius of 2 Å, stick boundary conditions (C = 6), viscosity 8.9 mPa s, and temperature 298.15 K.

With this first part of the input, it is possible to run the build_zmatrix script. The Z-Matrix reads

```
1   0  0.00000  0  0.00000  0  0.00000
2   1  1.47645  0  0.00000  0  0.00000
3   2  1.52868  1  109.049  0  0.00000
4   3  1.33908  2  118.054  1 -79.4839
5   2  1.08866  1  107.992  3  114.470
6   2  1.53751  1  111.771  3 -126.872
7   1  1.01889  2  108.978  3  68.6990
8   1  1.01417  2  109.930  3 -43.7973
9   1  1.00481  2  112.631  3 -169.016
10  6  1.09074  2  109.644  1 -61.6072
11  6  1.09036  2  109.586  1  178.994
12  6  1.09142  2  110.573  1  58.4945
```

```
13  3  1.22481  2  116.862  1   84.6037
14  4  1.46753  3  119.522  2   106.468
15  4  1.01024  3  113.556  2 -18.9831
16 14  1.54119  4  108.236  3 -71.4542
17 14  1.09202  4  110.089  3  46.2716
18 14  1.53611  4  109.024  3  165.980
19 16  1.24859 14  111.727  4  93.2348
20 16  1.25333 14  112.289  4 -41.4179
21 18  1.09081 14  109.878  4 -59.6381
22 18  1.09031 14  109.449  4 -179.307
23 18  1.09112 14  110.045  4  61.0655
```

The numbering scheme changed, as shown in the next figure. Panel (a) shows the atoms numbering in the original PDB file, while panel (b) shows the numbering in the 2ala.pdb.rebuild.xyz file. As it can be seen, atoms 7, 2, 1, 17 are changed into 1, 2, 3, 4.



(a) **(b)**

With reference to the Z-Matrix and the new numbering scheme in panel (b), the input file 2ala.inp defines the following 2 simple coordinates:

- the torsion around the dihedral angle defined by atoms 4-3-2-1: defined in the 8th line of the input file with atomID = 4 (4th line of the Z-Matrix), qType = 3 (dihedral angle). A scan from -99 deg to -59 deg is requested dividing the range in 3 intervals. A total of 4 conformations will be generated for this coordinate;
- the stretching between atoms 16-14: defined in the 9th line of the input file with atomID = 16 (16th line of the Z-Matrix), qType = 1 (bond length). A scan from 1.4 Å to 1.6 Å is requested dividing the range in 3 intervals. A total of 4 conformations will be generated for this coordinate.

A total of 16 conformations will be generated while running DiTe2 with the provided input file. The geometry (in XYZ format), friction, and diffusion tensors will be generated for each of the 16 conformations. A scan log file will also be given in output and would read

```
0    -99       1.4
1    -85.6667  1.4
2    -72.3333  1.4
3    -59       1.4
4    -99       1.46667
5    -85.6667  1.46667
6    -72.3333  1.46667
7    -59       1.46667
8    -99       1.53333
9    -85.6667  1.53333
10   -72.3333  1.53333
11   -59       1.53333
12   -99       1.6
13   -85.6667  1.6
14   -72.3333  1.6
15   -59       1.6
```

where the first column is the number counting the configurations, the second column reports the dihedral angle in deg, and the third column reports the bond length in Å.

4.2 Dialanine collective coordinate scan

This example is in the DiTe2-v1.0/examples/2alanine/2alanine_coll/ folder. The input file reads

```
pdb 2ala.pdb
refatoms 7 2 1 17
reff 2.0
C 6.0
viscosity 8.9e-4
temperature 298.15
coordinates collective 1
2 4 3 1.0 14 3 1.0 10 0.0 1.0
```

The first part of the input file is equivalent to the previous example. The difference is in the type of

internal coordinate, which is collective in this example. In particular, it is built combining 2 simple coordinates, which are the dihedral angle 4-3-2-1 (atomID = 4, qType = 3), and the dihedral angle 14-4-3-2 (atomID = 14, qType = 3). The corfficients are both equal to 1.0 and the collective coordinate is defined as 0 in the input PDB file. A scan is performed until the coordinate assumes value 1, dividing the range in 10 intervals.

4.3 Dialanine + water

This example is in the DiTe2-v1.0/examples/2alanine/2alanine_h2o/ folder. The input file reads

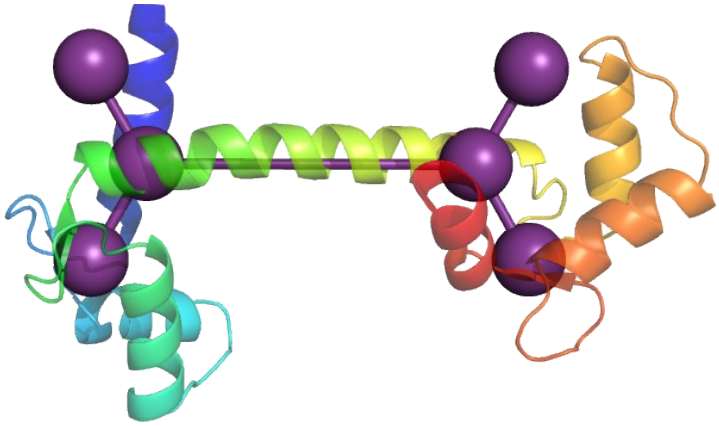
```
pdb 2ala.pdb
refatoms 7 2 1 17
reff 2.0
C 6.0
viscosity 8.9e-4
temperature 298.15
coordinates simple 1
4 3 0
bonds 1 17 24
```

In this case, the calculation is run on a single simple coordinate, i.e. the dihedral angle 4-3-2-1, with no scan (note the “0” after the qType). Since the PDB file contains a molecule of water together with the dialanine, a bond is manually added between atoms 17 and 24 (in the PDB numbering) to allow DiTe2 to build the Z-Matrix.

4.4 Coarse-grained Calmodulin simple coordinates scan

Two further examples of scan along simple coordinates are provided in the sub-folders:
- DiTe2-v1.0/examples/calmodulin/simple/stretching
- DiTe2-v1.0/examples/calmodulin/simple/stretching-torsion

A coarse-grained (CG) version of calmodulin is here analyzed. In particular, a 6-beads approximation of the volume occupied by the two domains of the protein. The CG model is shown in the next figure, together with a cartoon representation of the secondary structure of calmodulin.



In the first example, the input file reads

```
pdb cg.pdb
refatoms 1 2 4 5
reff 5.0
C 6.0
viscosity 8.9e-4
temperature 298.15
bonds 5 1 2 2 3 2 4 4 5 4 6
coordinates simple 1
3 1 1.0 10 50 500
```

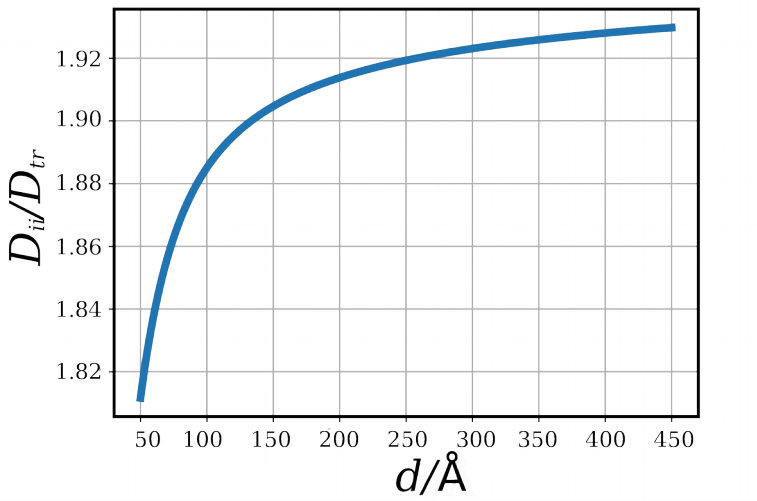
As it can be seen, since the 6 beads are far away (intra-domain distances are of 30 Å, while the domains are 50 Å apart), all of the 5 bonds need to be manually defined. Beads on the left domain are numbered 1, 2, and 3 in the cg.pdb file, while the other three beads ID's are 4, 5, and 6. Beads 2 and 4 are those connected by the long 50 Å bond. After running the build_zmatrix scirpt, the Z-Matrix reads

```
1 0 0 0 0 0 0
2 1 30.001  0 0 0 0
3 2 49.9992 1 120.626 0 0
4 3 29.9997 2 115.485 1 -180
5 2 29.9998 1 125.465 3 -180
6 3 29.9999 2 120.669 1 0
```

where now atom 2 keeps its ID, while atom 4 has become atom 3. With this new numbering, it is possible to interpret the coordinate: it is the stretching (qType = 1) between atom 3 and 2 (atomID = 3, i.e. 3rd row of the Z-Matrix). A scan is set up for this coordinate, changing the inter-domain distance

from 50 Å to 500 Å. The calculation is performed with an effective radius of the beads of 5 Å.

The figure below reports how the ratio between the internal part of the diffusion tensor and the isotropic part of the translational part of the diffusion tensor change while increasing the inter-domains distance.



In the second example, a second internal variable is simply added, i.e. the dihedral angle providing the mutual orientation of the two domains. The input file for this calculation reads

```
pdb cg.pdb
refatoms 1 2 4 5
reff 5.0
C 6.0
viscosity 8.9e-4
temperature 298.15
bonds 5 1 2 2 3 2 4 4 5 4 6
coordinates simple 2
4 3 10 0  90
3 1 10 50 60
```

As can be seen, here a bi-dimensional scan is performed changing the dihedral angle from 0 deg to 90 deg, and the inter-domains distance from 50 Å to 60 Å.

4.5 Coarse-grained Calmodulin collective coordinate scan

A further example with the CG calmodulin is found in the DiTe2-v1.0/examples/calmodulin/collective/generalized_screw sub-folder of the DiTe2 package. The input file reads

```
pdb cg.pdb
refatoms 1 2 4 5
reff 5.0
C 6.0
viscosity 8.9e-4
temperature 298.15
bonds 5 1 2 2 3 2 4 4 5 4 6
coordinates collective 1
2 4 3 1.0 3 1 1.0 10 0.0 1.0
```

A simple coordinate is built combining the stretching and torsion internal motions as defined in the previous examples. This provides a screw motion of one domain with respect to the other, mimicking a partial winding/unwinding of the α-helix connecting the two domains.

4.5 Atomistic Calmodulin first PCA coordinate

The last example of this package is found in the DiTe2-v1.0/examples/calmodulin/collective/pca sub-folder. The input file reads

```
pdb backbone.pdb
refatoms 1 2 3 5
reff 2.0
C 6.0
viscosity 8.9e-4
temperature 298.15
coordinates collective 1
28    283 3 -0.0656987    285 3 -0.0468999    287 3 -0.0122000    289 3 0.0809979
291 3 0.165901    293 3 -0.462899    295 3 0.194001    297 3 0.238385    299 3
-0.208699   301 3 0.0390000   303 3 0.370206   305 3 -0.392023   307 3 0.357603
309 3 -0.163599    311 3 -0.0877963    313 3 -0.0307996    315 3 -0.0175000    317 3
0.107699    319 3 -0.0534988    321 3 0.180999    323 3 -0.185000    325 3 0.198799
327 3 -0.051401   329 3 -0.0239000   331 3 -0.0444998   333 3 0.112300   3
35 3 -0.0986972   337 3 0.0421994 0
```

Here, an all-atom representation of the calmodulin backbone is used. As it can be seen, a collective variable is here defined with 28 dihedral angles. They are the (ϕ, ψ) couples of the 14 aminoacids composing the α-helix connecting the two domains of calmodulin.

The coefficients of the linear combination have been obtained from a principal components analysis (PCA) of the 28 dihedral angles from a molecular dynamics trajectory. The first PCA coordinate (i.e., the one with the highest eigenvalue) has been considered as the internal free degree of freedom. The coefficients are thus the elements of the eigenvector corresponding to the largest eigenvalue.

In the example, no scan is performed along the collective coordinate (note the “0” at the end of the long line defining the coordinate). The figure below reports the value of the internal part of the diffusion tensor, i.e. the diffusion of the collective change of conformation due by the concerted motion of the dihedral angles in the PCA coordinate, along a molecular dynamics trajectory.