Science and Technology Facilities Council

DL_FIELD

Force field conversion tool for molecular simulations

Chin Yong
Classical Materials & Molecular Modelling Group.
Daresbury Laboratory.

https://www.ccp5.ac.uk/DL_FIELD/

Engineering and Physical Sciences Research Council

CoSeC Computational Science Centre for Research Communities

HEC MCC

CCP5 Collaborative Computational Project Computer Simulation of Condensed Phases

ada lovelace centre

DL SOFTWARE Enabling Discovery

Functions

DL_FIELD acts like the "translator" between molecular systems and several simulation engines, to run molecular simulations.

DL_FIELD version 4.12 – released now! May 2025

(1) Convert molecular structures for DL_POLY (4.11), GROMACS (4.11) and LAMMPS (4.12) molecular/particulate simulation packages.

(2) Force field editor: modify, customise, addition of new and existing FF schemes.

(3) Standardised atom annotation (DL_F Notation): Identification of the chemical nature of every atom in a system model, and from such, to assign a universal expression that is independent of FF schemes.

(4) Force field format conversions: translation of third-party FF structure files and produces FF files for DL_POLY, GROMACS and LAMMPS.

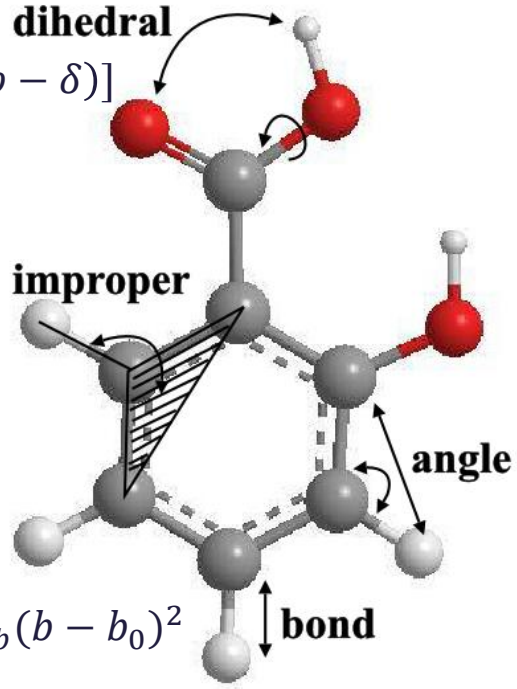
Why DL_FIELD

Tricky to setup force field models for molecular simulations, especially for complex models.

The atomic interactions are represented as a collection of simple mathematical functions (both inter- and intra-), also known as the potentials, collectively form a force field (FF).

Bonded: intramolecular - maintaining the geometrical structure of the molecule.

$$V(\varphi) = \sum_n k_n[1 + \cos(n\varphi - \delta)]$$



Salicylic acid

$$V(b) = k_b(b - b_0)^2$$

Basic components:

Bond (two-body)
Angle (three-body)
Dihedral (four-body)
Improper (four body)
Urey-Bradley (1-3 two body)

Can also have cross-components:

Bond-bond,
Bond-angle,
Angle-angle, etc.

Nonbonded: intermolecular - van-der-Waals and Coulombic

Science and Technology Facilities Council

FF model – requires all the necessary parameters to model Various interaction components

Intramolecular

$$\sum_{bonds} K_b(b - b_o)^2 + \sum_{angles} K_\theta(\theta - \theta_o)^2 + \sum_{torsions} K\phi(1 + \cos(n\phi - \delta))$$

$$+ \sum_{impropers} K_\varphi(\varphi - \varphi_o)^2 + \sum_{Urey-Bradley} K_{UB}(r_{1,3} - r_{1,3,o})^2$$

Intermolecular

$$\sum_{electrostatics} \frac{q_i q_j}{r_{ij}} + \sum_{VDW} \varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right]$$

These potential parameters are empirically derived or **adjusted** such that the potential function **approximately** represent the chemical behaviour of the interaction component.

The first derivatives of these functions give the forces applied on each atom and, from such, velocities and positions are derived.

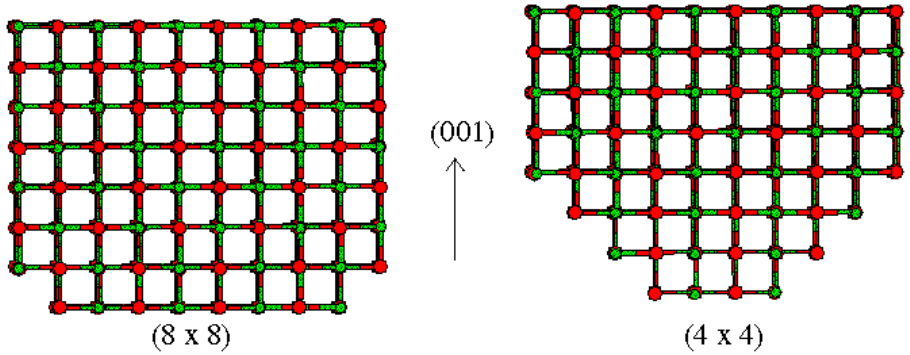
Force fields for

MgO, NaCl, simple ionic crystals - easy

MgO – rigid ion model

E(total) = E(vdw) + E(coulomb)

E(valence) = 0

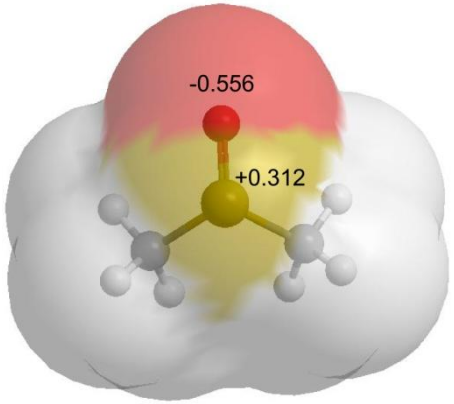


DL_POLY FIELD file

```
atoms 24
Mg3          24.30500     2.00000  12   0
O10          15.99940    -2.00000  12   0
finish
vdw 3
Mg3    Mg3    buck       0.0000      0.0000      0.0000
O10    Mg3    buck   26566.2720      0.3065      0.0000
O10    O10    buck  524868.3600      0.1490    667.8466
close
```

Science and Technology Facilities Council

Covalent molecules have a lot more parameters, even for a simple molecule



For example: dimethyl sulphoxide

E(total) = E(valence) + E(cross-terms) + E(nonbonded)

E(valence) = E(bond) + E(angle) + E(torsion) + E(oop) + E(ub)

$$U(b) = k(b - b_0)^2$$

atoms 10

HD	1.00797	0.09000	1
HD	1.00797	0.09000	1
HD	1.00797	0.09000	1
CD	12.01115	-0.14800	1
HD	1.00797	0.09000	1
HD	1.00797	0.09000	1
HD	1.00797	0.09000	1
CD	12.01115	-0.14800	1
SD	32.06400	0.31200	1
OD	15.99940	-0.55600	1

bonds 21

harm	4	1	644	1.11000
harm	4	2	644	1.11000
harm	4	3	644	1.11000
harm	8	5	644	1.11000
harm	8	6	644	1.11000
harm	8	7	644	1.11000
harm	4	9	480	1.80000
harm	8	9	480	1.80000
harm	9	10	1080	1.53000
-126	1	8	96195	129.03
-126	1	10	33433	84.716

angles 15

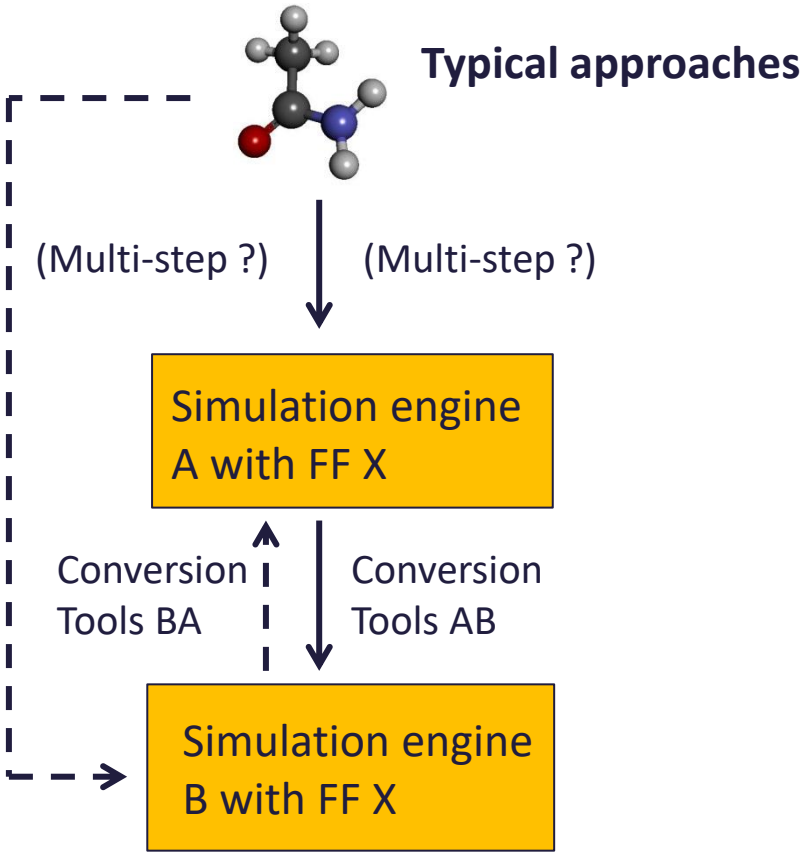
harm	1	4	2	71.00000	108.40000
harm	1	4	3	71.00000	108.40000
harm	1	4	9	92.20000	111.30000
harm	2	4	3	71.00000	108.40000
harm	2	4	9	92.20000	111.30000
harm	3	4	9	92.20000	111.30000
harm	5	8	6	71.00000	108.40000
harm	5	8	7	71.00000	108.40000
harm	5	8	9	92.20000	111.30000
harm	6	8	7	71.00000	108.40000
harm	6	8	9	92.20000	111.30000
harm	7	8	9	92.20000	111.30000
harm	4	9	8	68.00000	95.00000
harm	4	9	10	158.00000	106.75000
harm	8	9	10	158.00000	106.75000

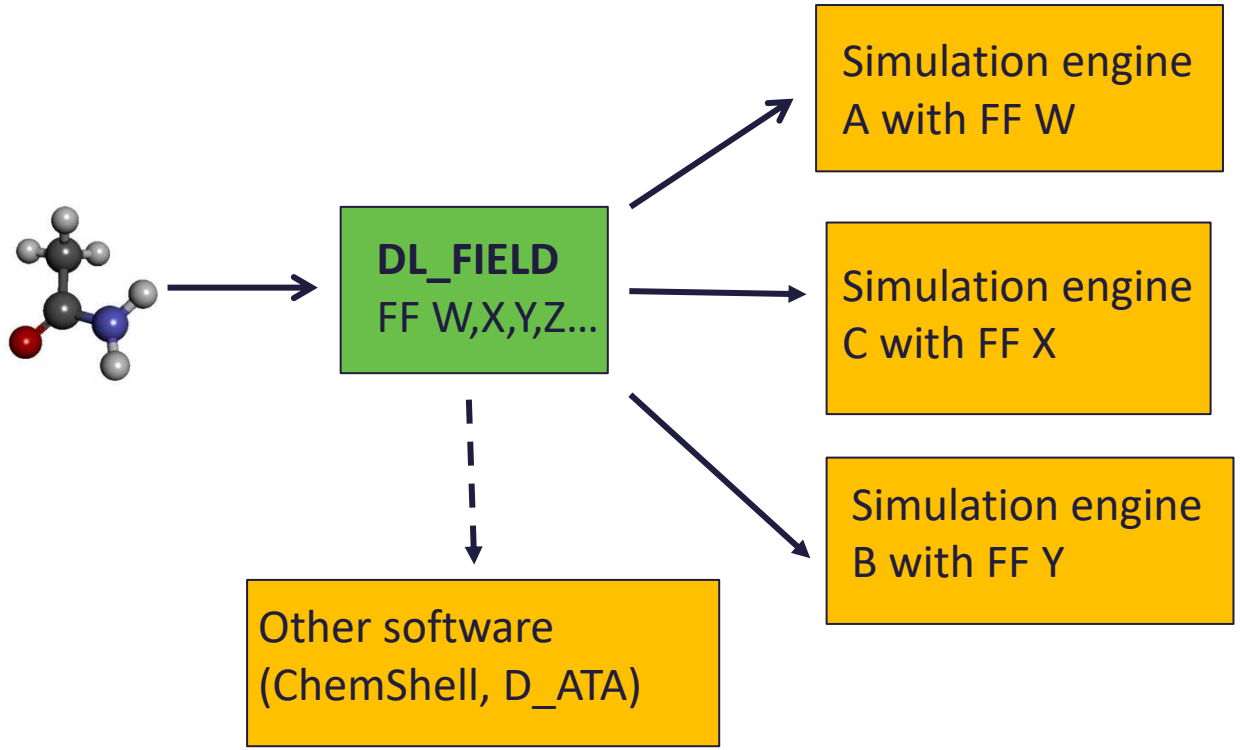
Also 12 dihedrals, and 10 vdw (CHARMM)

DL_FIELD development concept

Challenges for FF model setup – Intertwining factors of technical complexities in terms of data transformation, arbitrary typing rules, molecular system dependencies, variation of file formats, FF-rule specific. **The hardest bit is atom typing – decide what parameters to use**.

DL_FIELD: Striving towards one-step system setup — no patching, no hacking, no scripting: Run simulations and do your science.





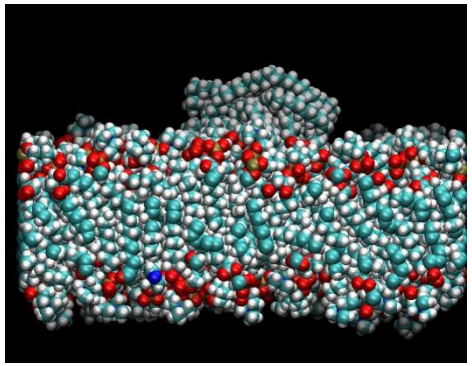
Standardised FF meta-data, typing with full integration of various FF schemes within DL_FIELD framework.

FF Schemes within DL_FIELD

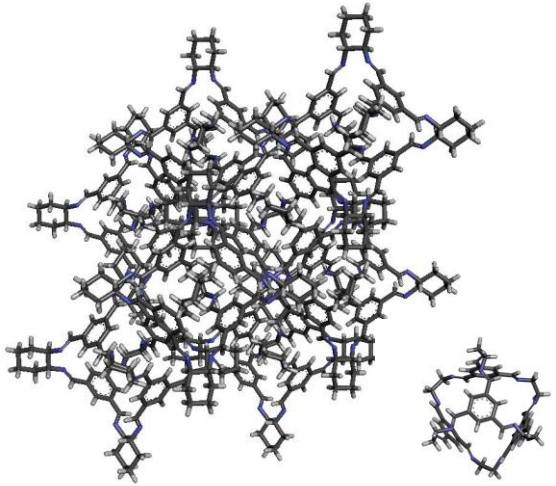
Name of potentials	potential_scheme
CHRMM22 for proteins	CHARMM22_prot
CHARMM36 for amino acids, proteins	CHARMM36_prot
CHARMM36 for nuclei acids	CHARMM36_nucl
CHARMM36 for lipids	CHARMM36_lipid
CHARMM36 for carbohydrates	CHARMM36_carb
CHARMM general force field	CHARMM36_cgenff
United-atom for CHARMM	CHARMM19
All-atom Amber	AMBER
AMBER general force field	AMBER16_gaff
OPLS-AA	OPLSAA
OPLS_2005	OPLS2005
OPLS AA/M for proteins	OPLS_AAM
CL&P for ionic liquids	OPLS_CL_P
OPLS for deep eutectic solvent	OPLS_DES
DREIDING	DREIDING
Polymer consistent force field	PCFF
Consistent Valence Force Field	CVFF
COMPASS force field	COMPASS
Transferable potentials (explicit hydrogen) for phase equilibria	TRAPPE_EH and TRAPPE_UA
Gromos united atom G54A7	G54A7
Inorganic force field**	INORGANIC
Inorganic FF for binary oxides	INORGANIC_binary_oxide
Inorganic FF for ternary oxides	INORGANIC_ternary_oxide
Inorganic FF for binary halides	INORGANIC_binary_halide
Inorganic FF for glass	IORGANIC_glass
Inorganic FF for clay	INORGANIC_clay
Inorganic FF for zeolite	INORGANIC_zeolite
Zeolite Hill-Sauer	INORGANIC_zeolite_HS
Non-specific, miscellaneous FF	MISC_FF
Multiple potential	multiple

- All-atom, united-atom models.
- Customisable, can also create new, user-define FF schemes.
- Each FF schemes are independent components express in a consistent standard and format within DL_FIELD.
- Application of multiple FF schemes for multiple complex component systems - bio-inorganic, organic-mineral, organic-organic, etc. (works for DL_POLY and GROMACS but not LAMMPS)

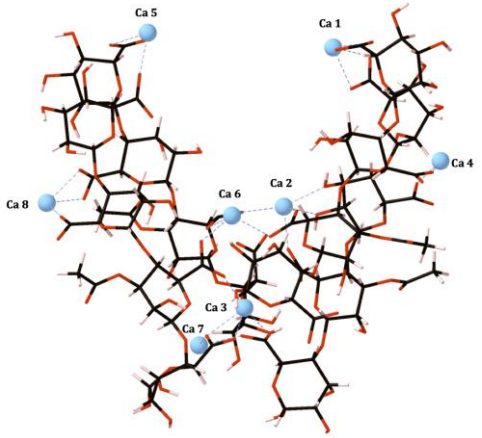
Example structures



Polymer on lipid membrane

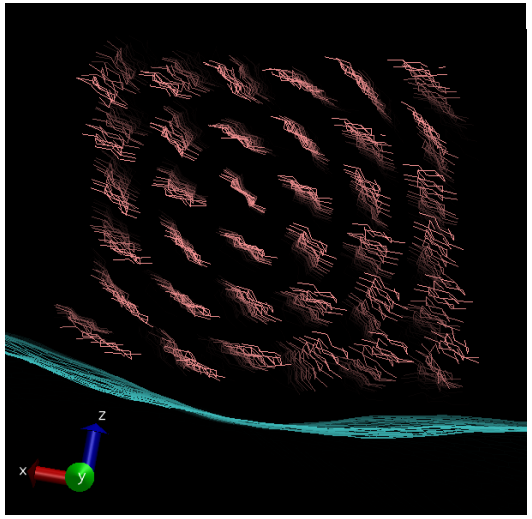


Molecular cages

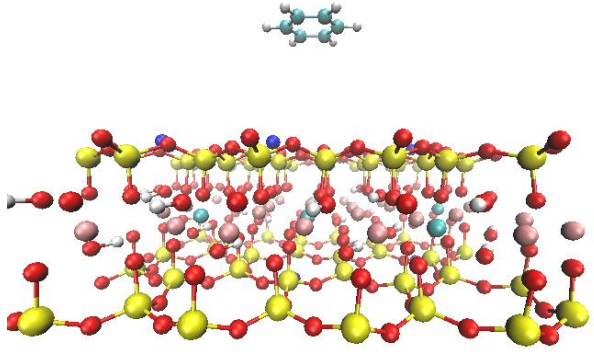


Biomolecules

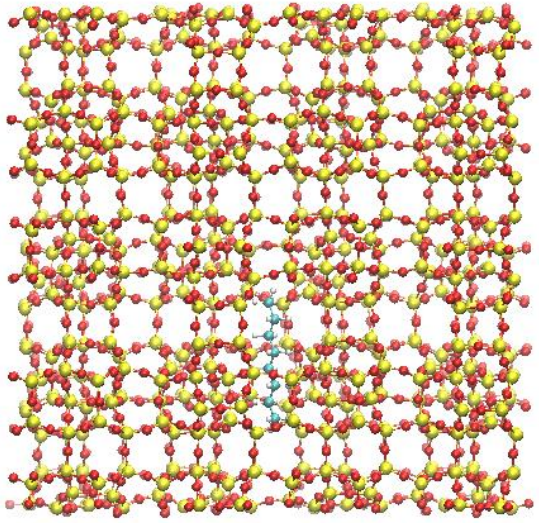
Other materials:
Proteins,
DNA,
Polymeric systems
Glass



Graphene-cellulose contact adhesion



Bio-inorganic systems
(montemorillonite+organic)



Zeolite+ adsorbates (hydrocarbons, organic molecules, gas, etc)

Science and Technology Facilities Council

DL_FIELD main features

- Convert a *PDB*, *xyz, mol2* structures to FF files for DL_POLY, GROMACS, LAMMPS.
- Define rigid bodies, and freeze, tether, constrain atoms.
- Pseudo point assignment.
- Core-shell model assignment.
- User-defined force field file (*udff*)
- Equivalence atom assignment
- Solvation with insertion of counter ions
- Bond, angles and dihedrals selections and exclusions
- **Fully automatic universal atom typing scheme (*xyz, mol2*)**
- **Standard chemistry-based atom annotation (DL_F Notation)**
- Multiple potentials (bio-inorganic)
- Auto refit LJ9-6 and Morse to LJ12-6 functions for vdw mixings (for multiple potential).
- Analytical derivation for LJ12-6 parameters based on atomic data (for multiple potential)
- **Solution Maker – set up disordered system**.
- Running DL_POLY, GROMACS or LAMMPS – single or multi-processors.
- Third party FF file conversions to *udff*.

UKRI Science and Technology Facilities Council

More details see DL_FIELD Reference Guide
https://dl-sdg.github.io/RESOURCES/TUTORIALS/dlf_0.html

DL_F Notation Typing

Unique atom typing procedure and annotate atoms in chemistry-sensitive notation

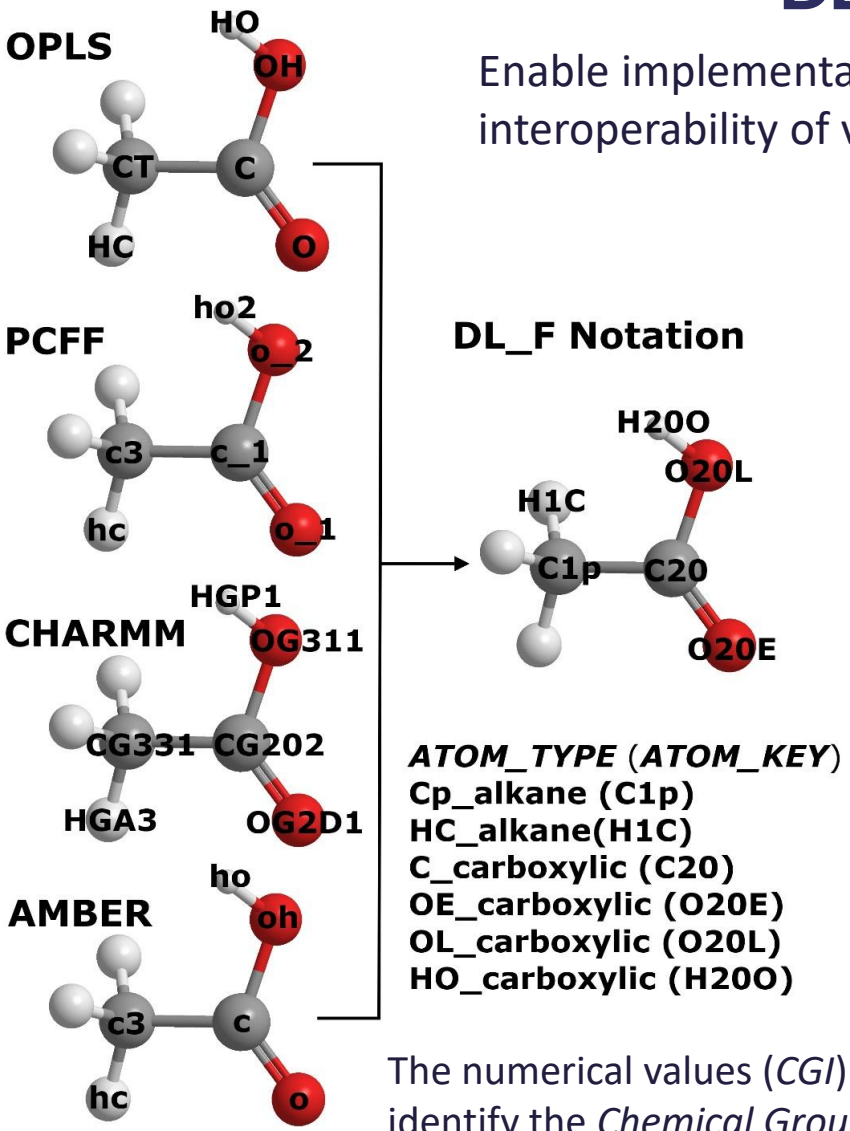
C. W. Yong, *J. Chem. Inf. Model.* 2016, **56**, 1405-1409

SSRN: https://dx.doi.org/10.2139/ssrn.4254942

Science and Technology Facilities Council

DL_F Notation Typing

Enable implementation of the unique integral framework that ensures interoperability of various FF schemes.



DL_F Notation

ATOM_TYPE **(*ATOM_KEY*)**
Cp_alkane (C1p)
HC_alkane(H1C)
C_carboxylic (C20)
OE_carboxylic (O20E)
OL_carboxylic (O20L)
HO_carboxylic (H20O)

The numerical values (*CGI*) uniquely identify the *Chemical Groups* (*CG*).

CGI **1** = alkane
CGI **20** = carboxylic

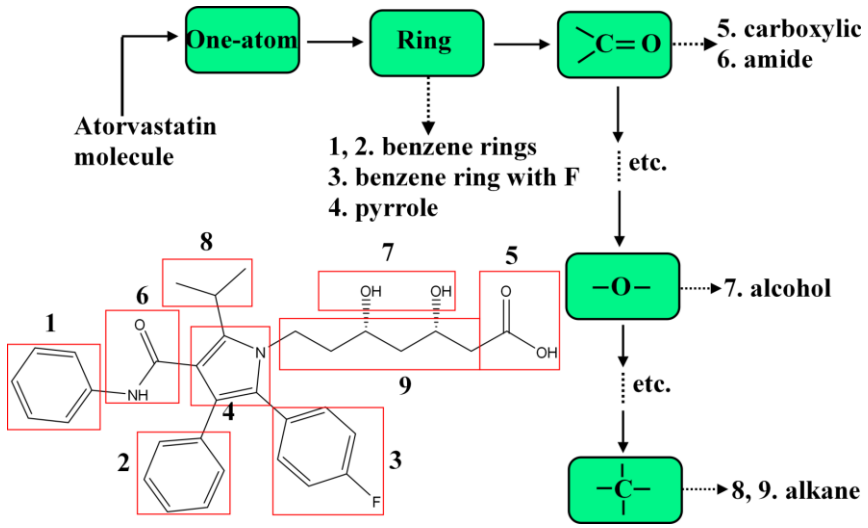
Science and Technology Facilities Council

- Unique to DL_FIELD
- Expression of standard atom typing that is contiguous across a range of FF schemes.
- Universal typing that smoothen data transition when migrating from one force field scheme to another.
- Allows one single conversion kernel applicable for different FF schemes.
- Easy to identify, with sensible format of naming atoms that indicates precisely the chemical nature of every atom in the system.
- Facilitate post analysis base on chemical identity rather than some arbitrary atom labels.
- **You have options to choose either standard FF or DL_F Notation format**.

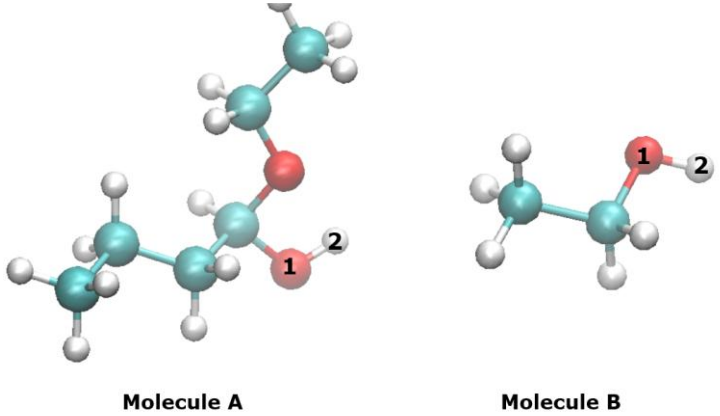
Each atom is assigned to a Chemical Group (CG)

The numerical values (*CGI*) uniquely identify the *Chemical Groups* (*CG*).

CGI **1** = alkane
CGI **20** = carboxylic



Analysis based on chemical identity rather than labels.



Molecule A | Molecule B

Molecule A
1 = O_hemiacetal (O25)
2 = HO_hemiacetal (H25O)

Molecule B
1 = O_alcohol (O15)
2 = HO_alcohol (H15O)

Science and Technology Facilities Council

Running DL_FIELD

Using DL_FIELD (in general)

Registration:
http://www.ccp5.ac.uk/DL_FIELD

Unpack program:
tar -xvf *dl_f_4.12.tar.gz*

Compile program:
cd source
make clean
make

Run program:
./dl_field

Science and Technology Facilities Council

DL_FIELD Directories

/source – source codes, where you do the compilation
/lib – Standard library file (.sf, .par)
/output – dl_poly.CONFIG, dl_poly.FIELD, dlf_notation.output, gromacs, lammps, etc
/Examples – Some PDB and *xyz* structures
/solvent –solvent templates
/utility – Contain some useful scripts.
/control_files – control files that run the example structures (Chapter 16)

Science and Technology Facilities Council

DL_FIELD FF Conversion Procedures

(Just a summary!)

When DL_FIELD is run:

1) Read *dl_f_path* file (master file that indicates where all relevant input and output files are located).

2) Read *control* file (options to setup specific FF models).

3) Examine user configuration files.

4) **Conversion mechanism** - Carry out **atom typing** procedures to obtain ATOM_TYPEs, depending on FF schemes and type of configuration files.

   (a) Looking for matching templates, in *sf* files (for *PDB*).

   (b) Molecular topology analysis (for *xyz, mol2*).

   (c) Combination of both (for *xyz*).

4) From ATOM_TYPEs, assign parameters, obtained from library and/or *udff* files.

5) Produce FF files:

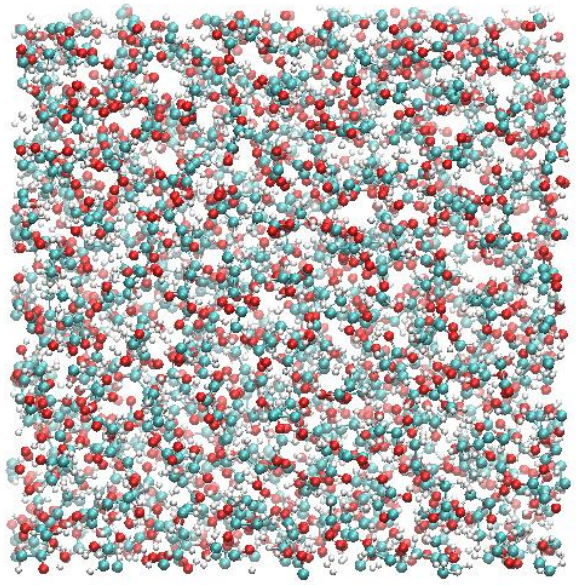
   dl_poly.CONFIG, *dl_poly.FIELD* and *dl_poly.CONTROL* files (DL_POLY) – by default

   lammps.in, *lammps1.data*, … (LAMMPS) - optional

   gromacs.gro, *gromacs.top*, *gromacs1.itp*, … (GROMCAS) - optional

6) Run simulation program (if this option is switched on), from native machine.

Conversion Mechanisms (atom typing and parameter assignment).



User's input structure
(PDB, xyz, mol2)

Atom typing

DL_F Notation
ATOM_TYPEs

DL_F Notation
ATOM_KEYs

Bonds,
Angles,
Dihedral, etc

Potential parameters

Add solvent
Atom states,
etc.

Undetermined structure,
Template unavailable, etc

Parameters
unavailable

Error. Program termination

FF model files

DL_POLY, GROMACS, LAMMPS

Science and
Technology
Facilities Council

Conversion Mechanisms - characteristics.

(a) **Template-based (MOLECULE in sf files)**

1. Matching of user input structures against some pre-defined MOLECULE templates (defined in .sf files in *lib/*). ATOM_TYPEs are directly obtained from the template.
2. If the template is not available, users are required to construct it manually in a *udff* file.
3. Flexible, allows users to adjust the model behaviour - introduce constrains, rigid body, core-shell, pseudo points, etc.
4. Possible use of the auto-CONNECT feature to simplify the MOLECULE template construction. Useful for complex molecules.
5. Suitable for structures in **PDB format**.
6. May be able to use ATOM_TYPEs in DL_F Notation in MOLECULE template. In this way, ATOM_KEYs in DL_F Notation can be produced.

Template-based FF (CHARMM, AMBER, etc)
Also for other FF is MOLECULEs are defined

(b) **Molecular topology analysis**

1. Do not need the MOLECULE template.
2. ATOM_TYPEs are automatically determined by carrying out detailed analysis of atomic structures.
3. Can make use of the DL_F Notation and determine the chemical nature of every atom in the system.
4. Suitable for ***xyz*** and **mol2** file formats.

OPLS2005, CVFF, PCFF, etc

(c) **Combination of both**

Suitable for ***xyz*** formats.
For all FF.

Science and Technology Facilities Council

Example PDB format

DL_FIELD can read cell parameters and convert this is cell vectors.

(For multiple-potential only)

Define FFs – indicate the extent of each FF.

PDB residue label, or MOLECULE_KEYs. Must have matching MOLECULE template

Residue id number

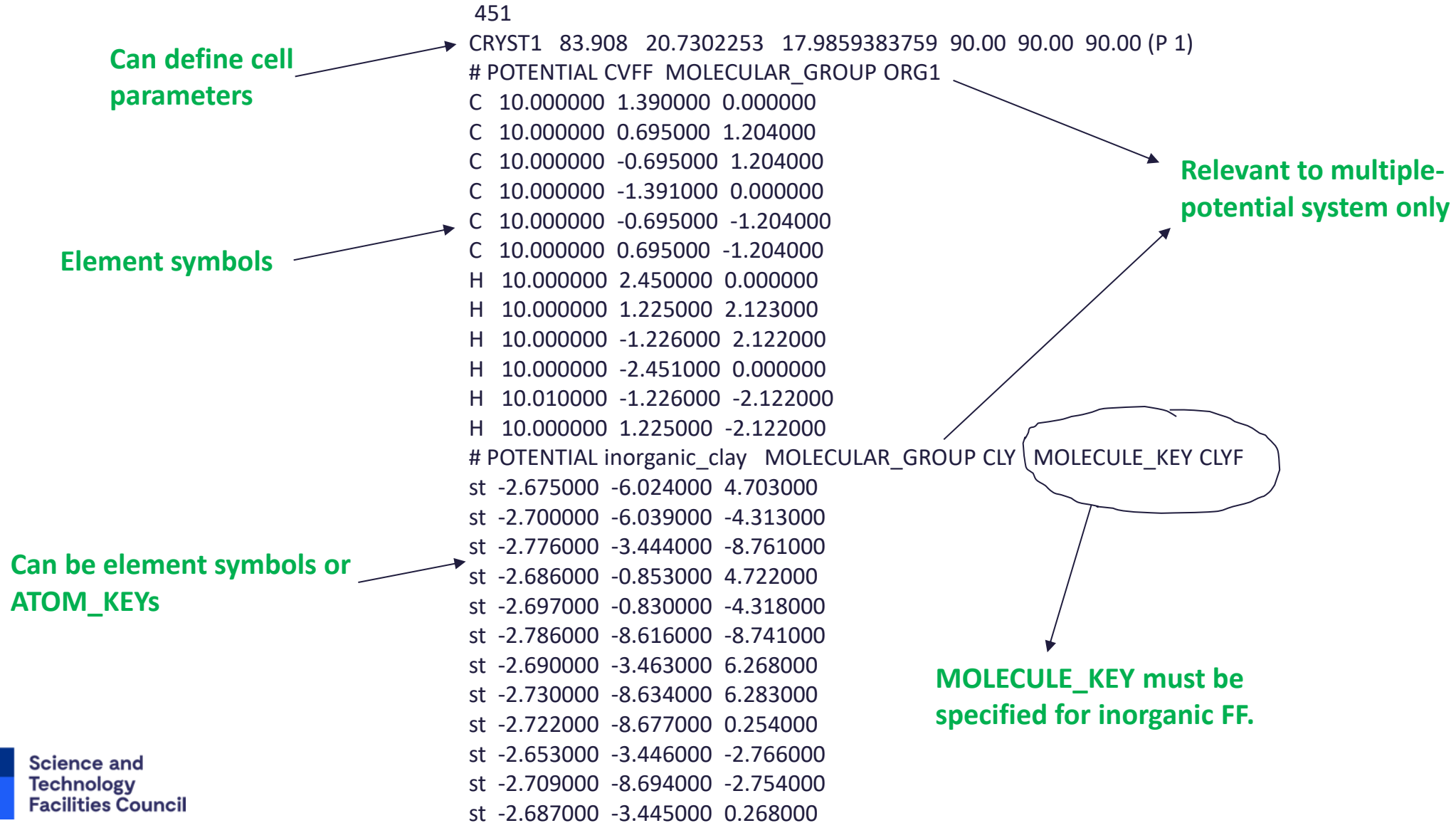
Molecular Group

Nothing is read beyond this

```
REMARK POTENTIAL charmm22_prot
CRYST1   83.908   20.7302253   17.9859383759  90.00  90.00  90.00 (P 1)
ATOM      1  C2  ETOH    1      0.995   0.329   5.000             GRP1
ATOM      2  H21 ETOH    1      1.844  -0.392   5.000             GRP1
ATOM      3  H22 ETOH    1      1.096   0.975   4.098             GRP1
ATOM      4  H23 ETOH    1      1.096   0.976   5.901             GRP1
ATOM      5  C1  ETOH    1     -0.340  -0.404   5.000             GRP1
ATOM      6  H11 ETOH    1     -0.456  -1.038   5.907             GRP1
ATOM      7  H12 ETOH    1     -0.457  -1.039   4.093             GRP1
ATOM      8  HO1 ETOH    1     -2.235   0.073   5.000             GRP1
ATOM      9  O1  ETOH    1     -1.394   0.538   5.000             GRP1
...
...
REMARK POTENTIAL inorganic_binary_oxide
HETATM    4 O10  MO3     1      0.000   6.312   0.000             GRP2
HETATM    1 Mg3  MO3     1      0.000   0.000   0.000             GRP2
HETATM    2 O10  MO3     1      0.000   2.104   0.000             GRP2
HETATM    3 Mg3  MO3     1      0.000   4.208   0.000             GRP2
HETATM    5 O10  MO3     1      2.104   0.000   0.0               GRP2
HETATM    6 Mg3  MO3     1      2.104   2.104   0.0               GRP2
HETATM    7 O10  MO3     1      2.104   4.208   0.0               GRP2
HETATM    8 Mg3  MO3     1      2.104   6.312   0.0               GRP2
HETATM    9 Mg3  MO3     1      4.208   0.0     0.0               GRP2
HETATM   10 O10  MO3     1      4.208   2.104   0.0               GRP2
HETATM   11 Mg3  MO3     1      4.208   4.208   0.0               GRP2
HETATM   12 O10  MO3     1      4.208   6.312   0.0               GRP2
HETATM    1 Mg3  MO3     2     10.000   0.000   0.00              GRP2
...
....
END
```

Science and Technology Facilities Council

Example *xyz* file contains an organic molecule on clay surface.



Science and Technology Facilities Council

The *dl_f_path* file

Master file-locator that indicates where all relevant input and output files are found. Must always placed in the folder where *dl_field* executable file is located (DL_FIELD home).

```
library = lib/
solvent = solvent/
output  = output/

# Control file
# To run example structures mentioned in Chaprter 16,
# comment out the default control file, dl_field.control, and then
# remove the comment for running the example structures.

control = control_files/example_1a.control
# control = dl_field.control

############################
# optional files
# Look for CHARMM rtf, and or prm files
# See Chapters 13.3 and 13.4
# charmm_rtf = pentanoic_acid.rtf  OPLS2005
# charmm_prm = pentanoic_acid.prm  OPLS2005
# charmm_psf = example.psf charmm36_prot
# charmm_pdb = example.pdb charmm36_prot
```

DL_FIELD Control File

Must always switch on (1),
To produce DL_POLY FF files

Option to produce other
FF files: gromacs, lammps, chemshell

Options

```
This is the title line. Put what you want. Reads only 120 columns
1            * Construct DL_POLY output files
none         * Secondary output file (gromacs, lammps, chemshell or none)
charmm36_prot     * Type of force field require.
kcal/mol          * Energy unit: kcal/mol, kJ/mol, eV, K or default
normal            * Conversion criteria (strict, normal, loose)
1                 * Bond type (0=default, 1=harmonic , 2=Morse)
2                 * Angle type (0=default, 1=harmonic, 2=harmonic cos)
none              * Include user-defined information. Put 'none' or a .udff filename
0                 * Verbosity mode: 1 = on, 0 = off
example.pdb   * Filename for the user's atomic configuration.
None              * Output file in PDB. Put 'none' if not needed.
0                  * Optimise FIELD output size, if possible? 1=yes  2=no
2                  * Atom display: 1 = DL_FIELD format. 2 = Standard format
1                 * Vdw display format: 1 = 12-6 format   2 = LJ format
Default            * Epsilon mixing rule (organic FF only) : 1 = geometric   2 = arithmatic or default
Default            * Sigma mixing rule (organic FF only) : 1 = geometric   2 = arithmatic or default
1                  * Epsilon mixing rule (inorganic FF only) : 1 = geometric   2 = arithmatic
2                  * Sigma mixing rule (inorganic FF only) : 1 = geometric   2 = arithmatic
1                  * Epsilon mixing rule (between different FF) : 1 = geometric   2 = arithmatic or default
1                  * Sigma mixing rule (between different FF) : 1 = geometric   2 = arithmatic or default
1                   * Display additional info. for protein 1=Yes  0=No
1                    * Freeze atoms? 1 = Yes (see below)  0 = No
0                  * Tether atoms? 1 = Yes (see below)  0 = No
0                 * Constrain bonds? 1 = Yes (see below) 0 = No
0                 * Apply rigid body? 1 = Yes (see below) 0 = No
auto               * Periodic condition ? 0=no, other number = type of box (see below)
50.00  0.0  0.0   * Cell vector a (x, y, z)
0.00  50.0  0.0   * Cell vector b (x, y, z)
0.00  0.0  50.0   * Cell vector c (x, y, z)
default           * 1-4 scaling for coulombic (put default or x for scaling=x)
```

Brief statement, describing all available options

A complete list of control features

Science and Technology Facilities Council

User-define force field (udff) File

- Define new molecular structures (such as MOLECULE templates) without tampering with the library *.sf* and *.par* files.
- Define new force field parameters, ATOM_TYPEs, etc specific to user's model.
- Redefine data by overriding existing force field information from the standard library (without doing any physical change to the data in the library file).
- Filename: *any_name.udff*

Simulation run Control Section (in DL_FIELD control file)

This part will only execute after FF files are successfully produced by DL_FIELD.

Optional. Only works if simulation engines are pre-installed in local computer.

```
************* DL_POLY control ******************
1        * Run DL_POLY program
DLPOLY.Z   * DL_POLY executable filename
/home/usr/ * absolute path to DL_POLY program
4        * Number of processors (mpirun)
1        * MM calculation 1=on  0=off
0   1000   * Structural Relaxation level (0 - off, 1,2 or 3). Total timestep
8.0       * cutoff (vdw and electrostatic)
100000      * Time limit for DL_POLY run (in seconds)

************* Gromacs control ******************
0         * Run Gromacs
gmx         * Gromacs executable filename
/usr/bin/gmx * absolute path to Gromacs
0         * MM single-point calc.

************* Lammps control ******************
0         * Run Lammps
lmp         * Lammps executable filename
/usr/bin/ * absolute path to Lammps
1         * MM single-point calc.
```

Not a proper equilibration, but to tame highly energetic starting structure

All simulation software must locate in the same machine.

Science and Technology Facilities Council

Use of DL_FIELD for GROMACS and LAMMPS

Version 4.11 – DL_POLY and GROMACS
Version 4.12 – DLPOLY, GROMACS, LAMMPS

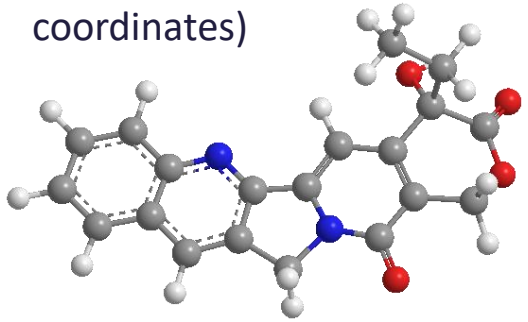
All features implemented in DL_FIELD are available for DL_POLY (except CMAP for proteins)

Only a subset of features implemented in DL_FIELD are available for GROMACS and LAMMPS.

UKRI Science and Technology Facilities Council

One-step conversions in DL_FIELD (to DL_POLY and GROMACS)

Can be *xyz* files
(elemental symbols with three coordinates)



(organic drug molecules)

1 - C6_quinoline
2 - C7_quinoline
3 - C8_quinoline
4 - C8a_quinoline
…
…
11 - CE_alkene
12 - Nt_amide
13 - Cs_alkane
14 - CL_alkene
15 - CL_alkene
16 - CE_alkene
17 - C_amide
…
…
25 - Cs_alkane
26 - Cp_alkane
27 - HC_aromatic
28 - HC_aromatic
29 - HC_aromatic
30 - HC_aromatic
31 - HC_aromatic
32 - HC_alkane
33 - HC_alkane
34 - HC_alkene
35 - HC_alkane
36 - HC_alkane

(DL_POLY FIELD file)

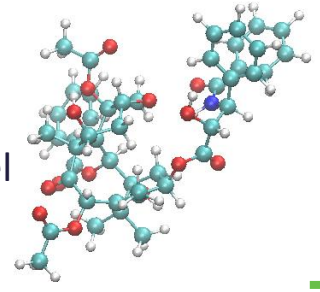
```
Generated by DL_FIELD v3.50
Units kcal/mol
Molecular types 1
Molecule name not_define
nummols 1
atoms 42
cp      12.01150   -0.12680   1   0
cp      12.01150   -0.12680   1   0
cp      12.01150   -0.12680   1   0
cp      12.01150    0.24050   1   0
cp      12.01150    0.00000   1   0
cp      12.01150   -0.12680   1   0
np      14.00670   -0.48100   1   0
…
```

(GROMACS itp file(s))

```
…
…
[ moleculetype ]
; name    nrexcl
XYZ    3

[ atoms ]
; nr   type   resnr   residu    atom    cgnr
charge    mass
    1       CA    1  MOL      C1     1  -0.1150000
    2       CA    1  MOL      C2     2  -0.1150000
    3       CA    1  MOL      C3     3  -0.1150000
    4       CA    1  MOL      C4     4  -0.1150000
    5       CA    1  MOL      C5     5  -0.1150000
…
```

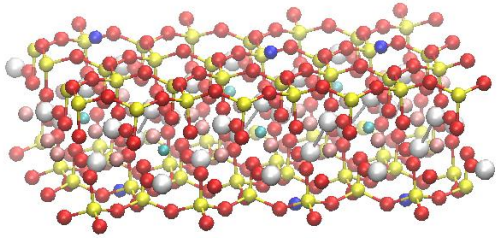
OPLS FF

paclitaxel



clay

CLAYFF



CHARMM FF

benzene

(inorganic clay + organic molecules)

DL_FIELD Universal atom typing (DL_F Notation)

Multiple potential
Inorganic FF + organic FF

A_5010
B_5010
etc…

Where A, B… = E[X][Y]#[y]

(DL_POLY FIELD file)

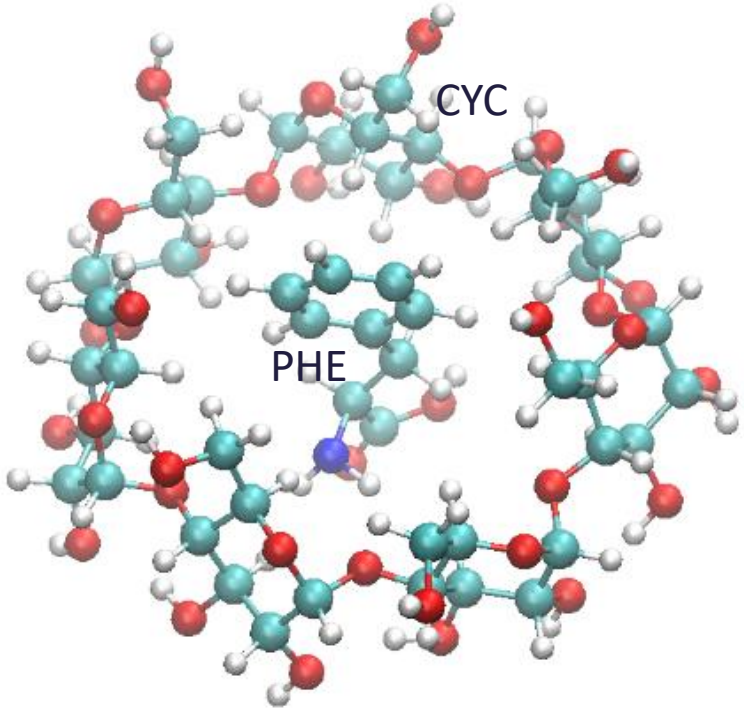
```
Generated by DL_FIELD v3.50
Units kcal/mol
Molecular types 1
Molecule name not_define
nummols 1
atoms 42
cp      12.01150   -0.12680   1   0
cp      12.01150   -0.12680   1   0
cp      12.01150   -0.12680   1   0
cp      12.01150    0.24050   1   0
cp      12.01150    0.00000   1   0
cp      12.01150   -0.12680   1   0
np      14.00670   -0.48100   1   0
…
```

(GROMACS itp file(s))

```
…
…
[ moleculetype ]
; name    nrexcl
XYZ    3

[ atoms ]
; nr   type   resnr   residu    atom    cgnr
    1       ao    1 CLYF     A1     1  1.5750000
    2       ao    1 CLYF     A2     2  1.5750000
    3       ao    1 CLYF     A3     3  1.5750000
    4       ao    1 CLYF     A4     4  1.5750000
    5       ao    1 CLYF     A5     5  1.5750000
…
```

One-step conversions in DL_FIELD (to DL_POLY and LAMMPS)

Model: phenylalanine enclosed in a β-cyclodextrin
Instruct DL_FIELD to solvate the system with ethyl acetate (ETAC) solvent in 60x60x60 Angstrom cubic box.



Input structure in xyz

```
170
Cyclodextrin with PHE.
# MOLECULAR_GROUP CYC
C  -6.049000  0.945000  0.569000
C  -6.341000  2.134000  -0.313000
C  -5.113000  3.000000  -0.453000
C  -4.574000  3.388000  0.915000
C  -4.429000  2.153000  1.811000
C  -4.114000  2.550000  3.250000
C  -4.381000  -3.897000  -0.680000
C  -5.277000  -3.152000  -1.645000
C  -5.119000  -1.651000  -1.519000
C  -5.285000  -1.237000  -0.064000
...
...
H  -2.830000  6.347000  -2.095000
H  0.359000  5.982000  -1.647000
H  0.509000  6.438000  4.356000
# MOLECULAR_GROUP PHE
C  -0.830000  0.683000  2.322000
C  -0.114000  -0.185000  3.052000
C  0.873000  -0.882000  2.472000
...
...
```

Science and Technology Facilities Council

DL_FIELD creates four files:
(1) lammps.in
(2) lammps1.data (CYC)
(3) lammps2.data (PHE)
(4) lammps3.data (ETAC)

Atom labels are included in remarks (#)

Prefix #1 means atom keys from first Molecular Group (CYC)

Prefix #solv means atom keys from solvent.

Science and Technology Facilities Council

lammps1.data

```
# LAMMPS data file. Produced from DL_FIELD 4.11
# Molecular Group 1: CYC
# Force field scheme: opls2005

   147 atoms
   154 bonds
   287 angles
   462 dihedrals
    0 impropers

    7 atom types
    6 bond types
   13 angle types
   14 dihedral types
    0 improper types

   -20.000000    20.000000 xlo xhi
   -20.000000    20.000000 ylo yhi
   -20.000000    20.000000 zlo zhi

Masses

  1   12.011500   # CAO
  2   12.011500   # CT
  3   15.999400   # OAL
  4   15.999400   # OAS
  5   15.999400   # OS
  6    1.007970   # HC
  7    1.007970   # HO
...
...
 13 opls     1.13500    -0.15100     0.40000     0.00000
 14 opls     2.23800    -2.32700    -0.68300     0.00000

Atoms

      1       1       1  0.300000  -5.976818   0.946629   0.523112 # CAO
      2       1       2  0.205000  -6.268818   2.135629  -0.358888 # CT
      3       1       2  0.205000  -5.040818   3.001629  -0.498888 # CT
      4       1       2  0.170000  -4.501818   3.389629   0.869112 # CT
      5       1       2  0.170000  -4.356818   2.154629   1.765112 # CT
      6       1       2  0.145000  -4.041818   2.551629   3.204112 # CT
```

Atom keys

lammps.in

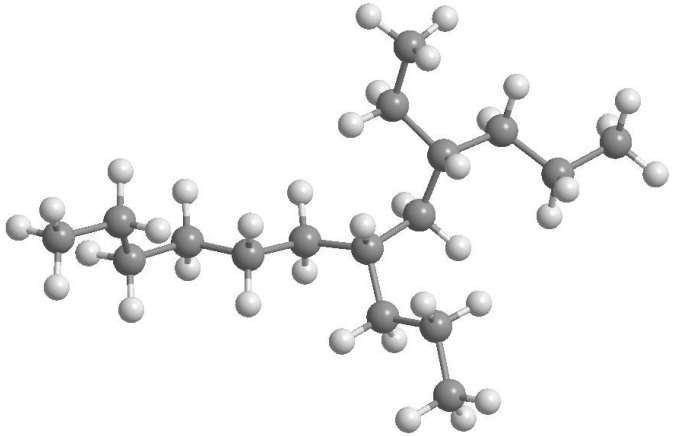
```
# LAMMPS input script, produced by DL_FIELD 4.11
# Title: Control file title. For DL_FIELD 4.2.
# Input structure: test_structures/cyclodextrin_phe.xyz

units             real  # in kcal/mol
dimension         3     # 3D simulation
atom_style        full  # molecular system with charges
boundary          p p p # Normal periodic boundary
timestep          0.5   # in fs

bond_style        hybrid harmonic
angle_style       hybrid harmonic
dihedral_style    hybrid opls
improper_style    hybrid cvff
pair_style        hybrid lj/cut/coul/long 12.000000
kspace_style      pppm 1.0e-4
special_bonds     lj 0.0 0.0 0.500000 coul 0.0 0.0 0.500000

read_data lammps1.data &
  extra/atom/types       17  &
  extra/bond/types       17  &
  extra/angle/types      25  &
  extra/dihedral/types   22  &
  extra/improper/types    4  &
  extra/improper/per/atom 4 &
  extra/bond/per/atom    1
read_data lammps2.data add append offset 7 6 13 14 0
read_data lammps3.data add append offset 18 17 30 29 3

group CYC type 1 2 3 4 5 6 7
group PHE type 8 9 10 11 12 13 14 15 16 17 18
group ETAC type 19 20 21 22 23 24

pair_coeff    1    1  lj/cut/coul/long    0.066000    3.500000 # CAO#1 CAO#1
pair_coeff    1    2  lj/cut/coul/long    0.066000    3.500000 # CAO#1 CT#1
pair_coeff    1    3  lj/cut/coul/long    0.105925    3.304542 # CAO#1 OAL#1
pair_coeff    1    4  lj/cut/coul/long    0.096125    3.185906 # CAO#1 OAS#1
pair_coeff    1    5  lj/cut/coul/long    0.096125    3.185906 # CAO#1 OS#1
pair_coeff    1    6  lj/cut/coul/long    0.044497    2.958040 # CAO#1 HC#1
pair_coeff    1    7  lj/cut/coul/long    0.044497    1.322876 # CAO#1 HO#1
...
...
```

Auto-group atoms into respective groups.

Atom keys append with group labels

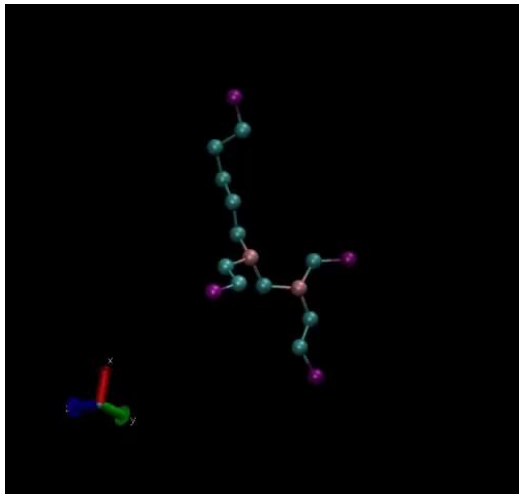
4-ethyl-6-propyl dodecane

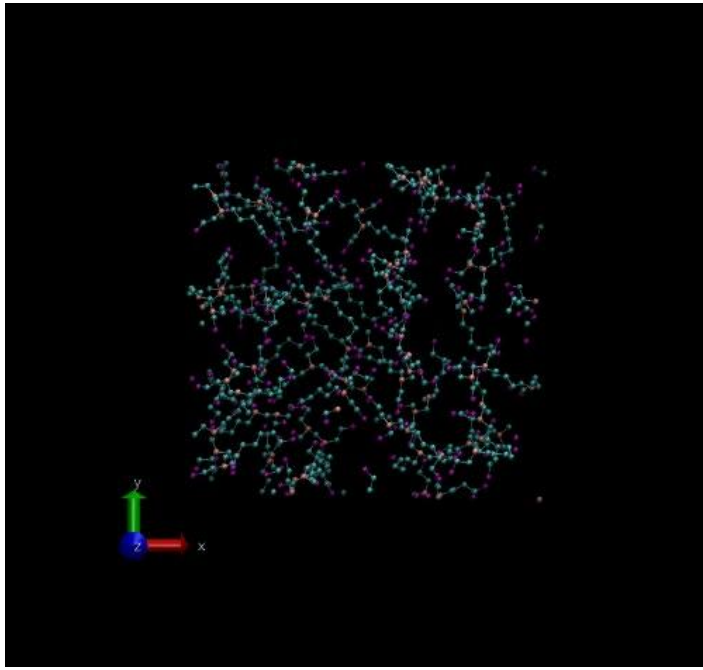
Use a single *MOLECULE alkanes* template defined in *TRAPPE_UA.sf* to setup alkanes of any length and branches (use of auto-CONNECT feature)



```
REMARK   This PDB file was created by CS Chem3D.
HETATM    1  CH1 ALK    1      2.620  -1.142   0.764                C
HETATM    4  CH2 ALK    1      1.567  -0.015   0.869                C
HETATM   10  CH2 ALK    1      0.103  -0.479   0.772                C
HETATM   11  CH3 ALK    1     -0.865   0.708   0.656                C
HETATM   13  CH2 ALK    1      2.488  -2.151   1.929                C
HETATM   16  CH1 ALK    1      3.362  -3.424   1.792                C
HETATM   18  CH2 ALK    1      4.527  -3.390   2.808                C
HETATM   21  CH2 ALK    1      5.673  -4.366   2.490                C
HETATM   24  CH2 ALK    1      6.852  -4.201   3.464                C
HETATM   27  CH2 ALK    1      7.974  -5.234   3.260                C
HETATM   30  CH2 ALK    1      8.694  -5.108   1.908                C
HETATM   33  CH3 ALK    1      9.880  -6.077   1.802                C
HETATM   37  CH2 ALK    1      4.052  -0.561   0.703                C
HETATM   40  CH3 ALK    1      4.336   0.330  -0.515                C
HETATM   46  CH2 ALK    1      2.519  -4.710   1.943                C
HETATM   50  CH3 ALK    1      0.759  -6.253   0.944                C
HETATM   53  CH2 ALK    1      1.566  -4.960   0.760                C
END
```

```
Control file title. For DL_FIELD 4.2.
1       * Construct DL_POLY output files
lammps   * Construct DL_MONTE output files
trappe_ua * Type of force field require (see list below for choices).
kcal/mol  * Energy unit: kcal/mol, kJ/mol, eV, or K.
normal  * Conversion criteria (strict, normal, loose)
1       * Bond type (0=default, 1=harmonic , 2=Morse)
1       * Angle type (0=default, 1=harmonic, 2=harmonic cos)
none  * Include user-defined information. Put 'none' or a .udff filename
1      * Verbosity mode: 1 = on, 0 = off
test_structures/UA_models/4-ethyl-5-propyl-dodecane_ua.pdb * Configuration file.
none   * Output file in PDB. Put 'none' if not needed.
1 50 molecules 4.0 * Solution Maker: on/off, density, unit, cutoff)
1       * Optimise FIELD output size, if possible? 1=yes  0=no
2       * Atom display: 1 = DL_FIELD format. 2 = Standard format
2       * Vdw display format: 1 = 12-6 format   2 = LJ format
default * Epsilon mixing rule (organic FF only) : default, or 1 = geometric, 2 = arithmatic
default   * Sigma mixing rule (organic FF only) : default, or 1 = geometric, 2 = arithmatic
2       * Epsilon mixing rule (inorganic FF only) : 1 = geometric   2 = arithmatic
1       * Sigma mixing rule (inorganic FF only) : 1 = geometric   2 = arithmatic
1    * Epsilon mixing rule (BETWEEN different FF) : 1 = geometric   2 = arithmatic
2    * Sigma mixing rule (BETWEEN different FF): 1 = geometric 2 = arithmatic
0       * Display additional info. for protein 1=Yes  0=No
0       * Freeze atoms? 1 = Yes (see below)  0 = No
0       * Tether atoms? 1 = Yes (see below)  0 = No
0       * Constrain bonds? 1 = Yes (see below) 0 = No
0       * Apply rigid body? 1 = Yes (see below) 0 = No
1          * Periodic condition ? 0=No, other number = type of box
40.0    0.0         0.0          * Cell vector a (x, y, z)
0.0     40.0        0.0          * Cell vector b (x, y, z)
0.0     0.0     40.0     * Cell vector c (x, y, z)
```





Duplicate 50 molecules in 40x40x40 cubic box, NVE, 50 ps
(Lammps simulation. Setup as non-charge system)

DL_FIELD

Example structures

Tutorial Exercises

Science and Technology Facilities Council

DL_FIELD Tutorial Pre-requisite

Compile your own

Unpack program:
tar -xvf *dl_f_4.12.tar.gz*

Compile program:
cd source
make clean
make

Run program:
./dl_field

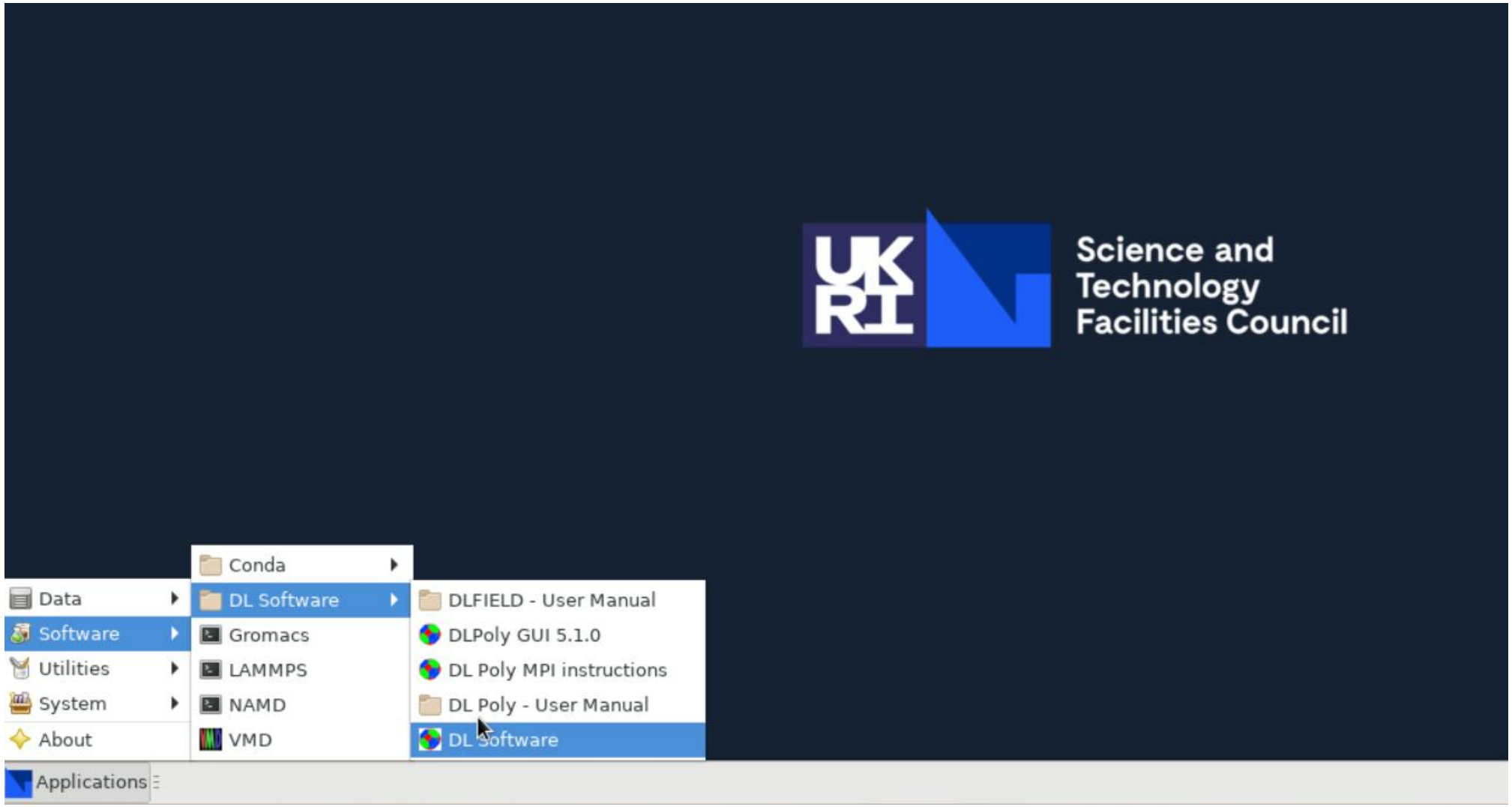
OR

Use virtual training

Access: https://training.analysis.stfc.ac.uk .

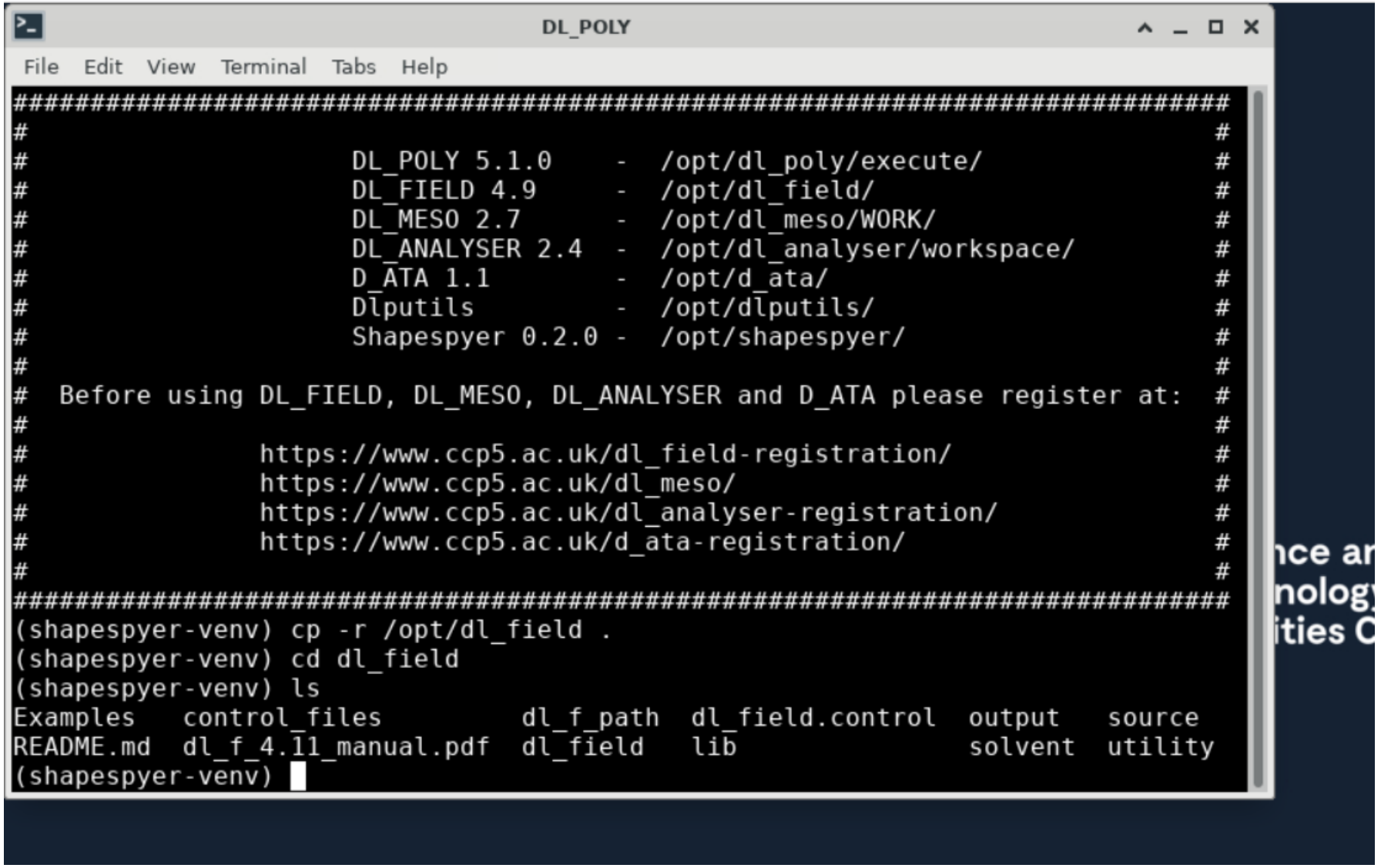
Sign up:
https://training.analysis.stfc.ac.uk//add_user/?token=IkLH6zFganC97P9A5rSR3E9g

Science and Technology Facilities Council



Science and
Technology
Facilities Council

Virtual training: Call up DL_Software window



Science and Technology Facilities Council

Virtual training: Copy DL_FIELD over to your DL_Software window

DL_FIELD Tutorial Exercises

Running example structures

```
# Directory paths for DL_FIELD version 4.10.
# C W Yong, January 2024
# This file must be located where DL_FIELD executable is located.
# The directory must be changed correspondly if you move the file components.
# All directory paths are RELATIVE to DL_FIELD home directory.
# Do not use absolute paths.

# paths
library = lib/
solvent = solvent/
output  = output/

# Control file
# To run example structures mentioned in Chaprter 16,
# comment out the default control file, dl_field.control, and then
# remove the comment for running the example structures.

control = control_files/example_1a.control
# control = dl_field.control

##########################
# optional files
# Look for CHARMM rtf, and or prm files
# See Chapters 13.3 and 13.4
# charmm_rtf = pentanoic_acid.rtf  OPLS2005
# charmm_prm = pentanoic_acid.prm  OPLS2005
```

Look into *dl_f_path*

DL_FIELD is pointing to the *example_1a.control* file, which will read Example 1(a) structure as described in DL_FIELD manual, Chapter 16

Change the control file to *example_2a.control* will read and convert structure 2(a), etc.

Tutorial exercises

https://dl-sdg.github.io/RESOURCES/EXERCISES/exercises_dlfield.html

Then click on Setup solutions

Science and Technology Facilities Council

DL Atom Typer and Analyser (D_ATA)

Combinations of DL_F Notation and DANAI – two unique features from DL_FIELD and DL_ANALYSER combine into one.

Chin Yong
Classical Materials & Molecular Modelling Group.
Daresbury Laboratory.

https://www.ccp5.ac.uk/D_ATA/

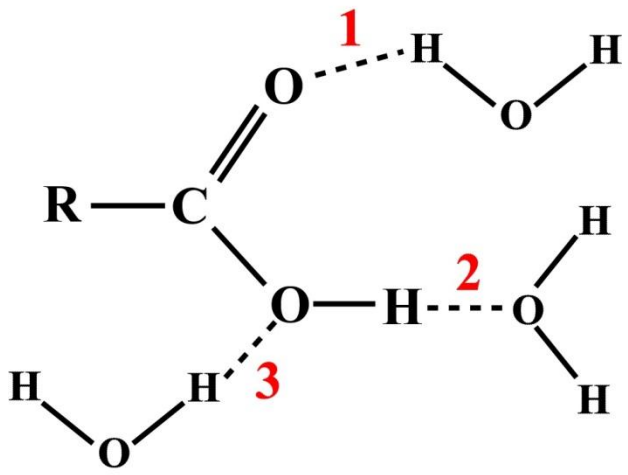
Engineering and Physical Sciences Research Council

DL SOFTWARE Enabling Discovery

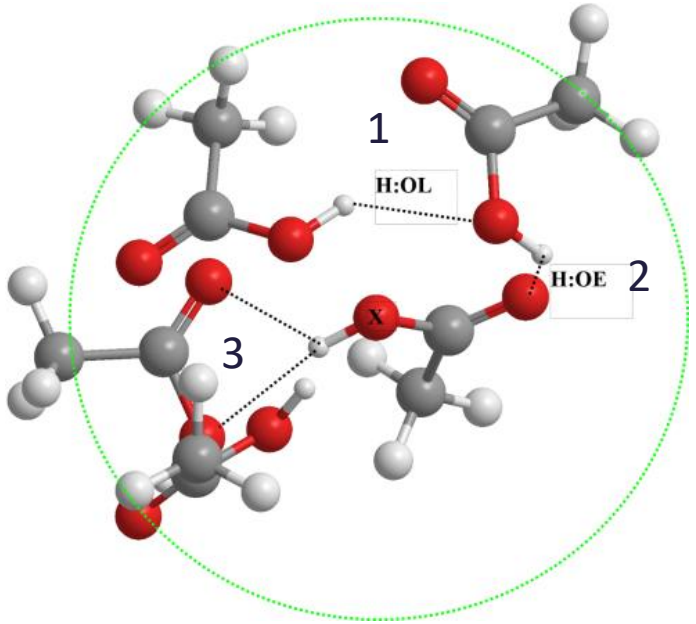
CoSeC Computational Science Centre for Research Communities

Spin offs: Use of DL_F Notation to create a chemistry-based language construct (DANAI) to annotate atomic interactions.

C. W. Yong & I. T. Todorov, *Molecules* (2018), **23**, 36
C. W. Yong *et. al. Data in Brief* (2023) 50:109485



Traditional textual and diagrammatic Illustrations.



How do you describe this mode of localised H-bond interaction centred about the atom X within a defined region of space?

HB_20_20

[C4]H:OL-H:OE@<[R2]H:OE@OL:H>
1 2 3

Science and Technology Facilities Council

Atom Typer and Analyser (D_ATA)

Functions

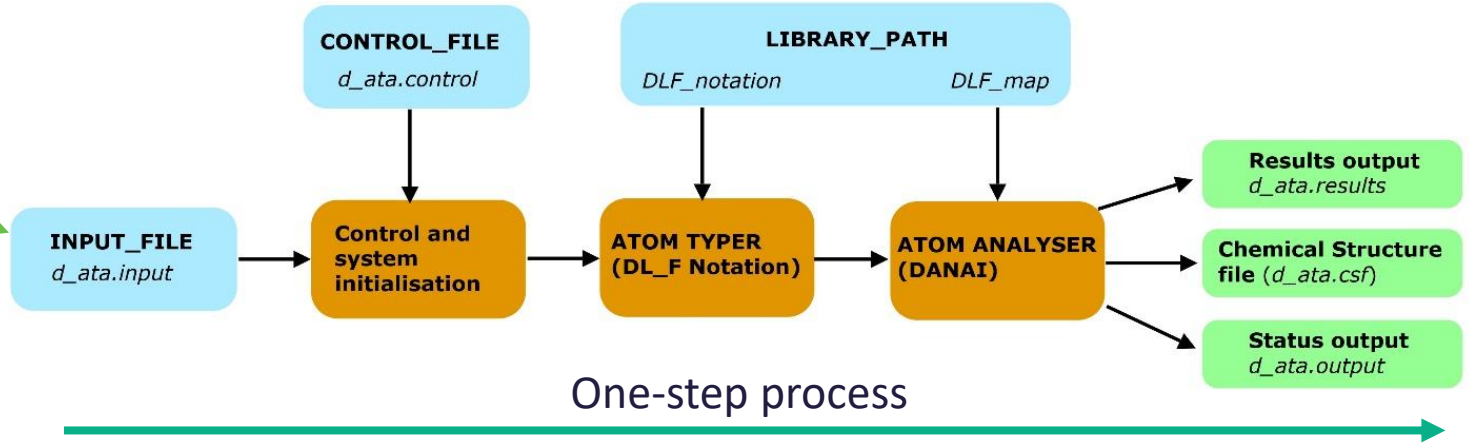
Identification, annotation, quantification and characterisation of atomic interactions: universal, searchable and discoverable.

Applicable to any atomistics systems (molecular simulations, experimental) – crystalline, biological, condensed phase.

Analysis – statistical, correlations, geometrical.

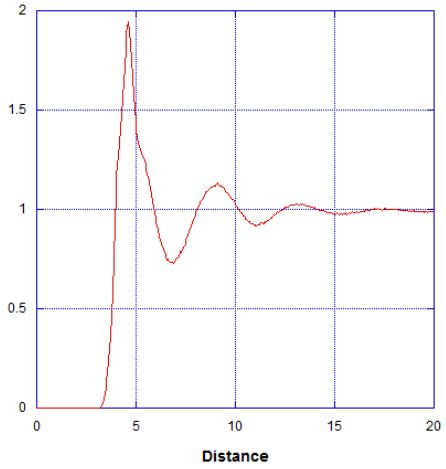
Identification of signature atomic interactions for purposes of structural formulation.

MD trajectories
Single configurations
(PDB, xyz, HISTORY)

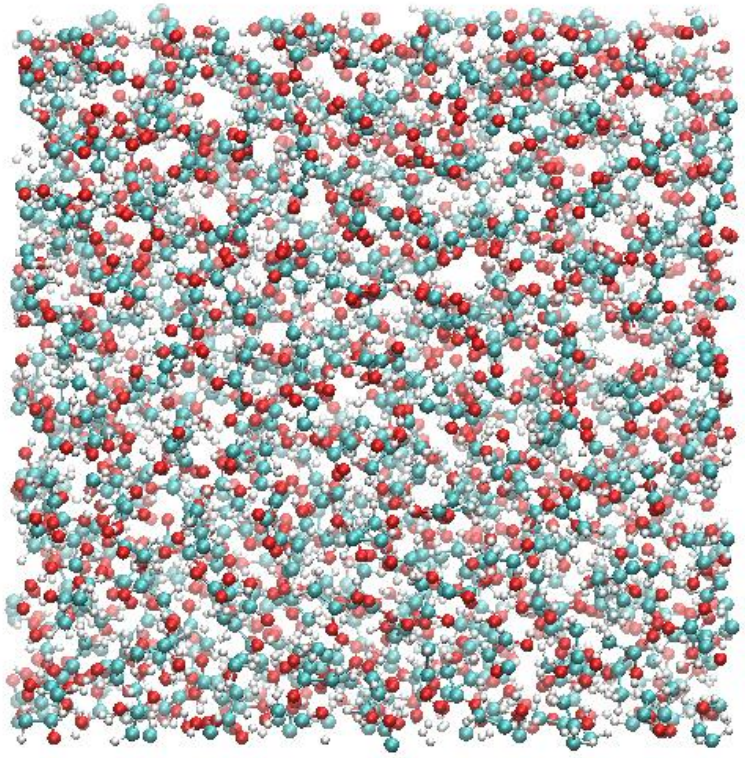


Science and Technology Facilities Council

Example 1: ethanoic acid liquid



Radial distribution functions (average sense packing information)



(MD simulations showing only carboxylic groups)

Total HB interactions: 694

D_ATA outputs

HB_20_20

	DANAI	counts
1.	[L2]O20E:H20O	= 510
2.	[L2]O20L:H20O	= 77
3.	[L2]o20E:h20O	= 14
4.	[L2]o20E:H20O	= 36
5.	[L2]O20L:h20O	= 34
6.	[L2]h20O:O20E	= 26
7.	[L2]o20L:h20O	= 2
8.	[L3]h20O:O20E:H20O	= 14
9.	[L3]O20E:H20O:O20L	= 5
10.	[L3]H20O:O20E:H20O	= 11
11.	[L3]h20O:O20L:h20O	= 1
12.	[L3]O20E:H20O:O20E	= 2
13.	[L3]o20E:H20O:O20E	= 2
14.	[L3]O20L:H20O:O20L	= 1
15.	[L3]o20L:H20O:O20L	= 1
16.	[L3]o20E:H20O:O20L	= 1
17.	[L4]H20O:O20E:H20O:O20E	= 2
18.	[L4]H20O:O20E:H20O:O20L	= 1

HP_1_1

Total HP interactions: 951

	DANAI	counts
1.	[L2]c1p:c1p	= 872
2.	[L2]c1p:C1p	= 78
3.	[L2]C1p:C1p	= 1
4.	[L3]c1p:c1p:c1p	= 1482
5.	[R3]c1p:c1p:c1p:c1p	= 93
6.	[L3]c1p:C1p:c1p	= 119
7.	[L3]c1p:c1p:C1p	= 144
8.	[L3]c1p:C1p:C1p	= 18
9.	[R3]c1p:c1p:C1p:c1p	= 32
10.	[R3]C1p:c1p:C1p:C1p	= 3
11.	[L3]C1p:c1p:C1p	= 2
12.	[L3]C1p:C1p:C1p	= 2
13.	[L4]c1p:c1p:C1p:c1p	= 413
14.	[L4]c1p:c1p:c1p:c1p	= 2395
15.	[L4]c1p:c1p:c1p:C1p	= 238
16.	[L4]c1p:C1p:C1p:c1p	= 28
17.	[L4]C1p:c1p:C1p:c1p	= 20

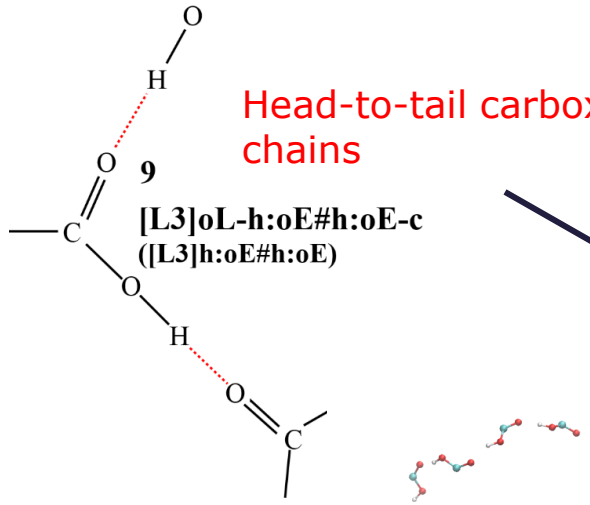
…

…

Science and Technology Facilities Council

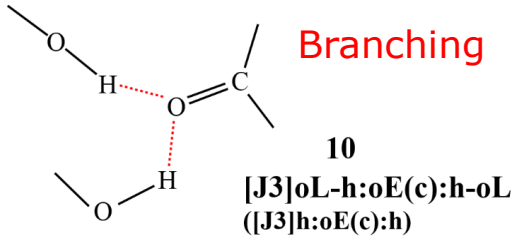
Rationalise complex condensed phase models.

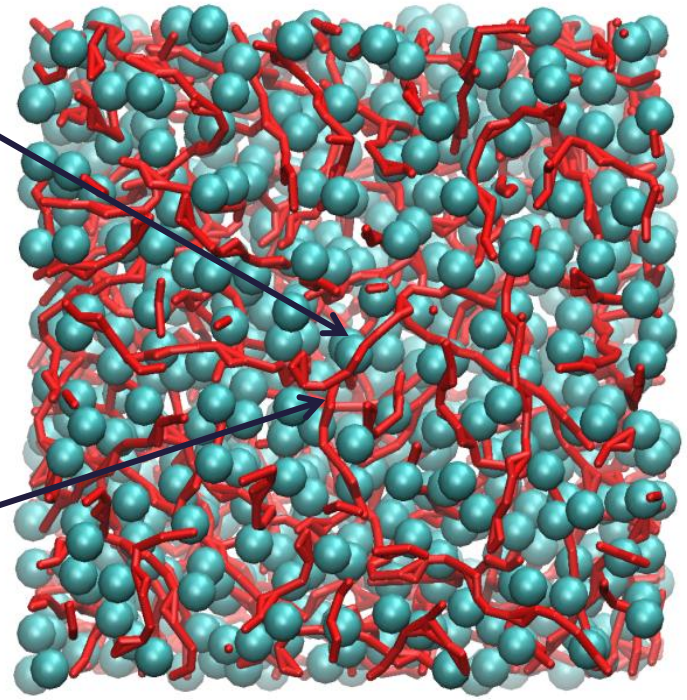
'Structured interactions' in liquid ethanoic acid. HB interactions forming 'head-to-tail' chains (red streaks) with branching. Methyl groups (cyan spheres) formed a sponge-like hydrophobic (HP) super-structure that permeates across the polymeric-like HB networks.

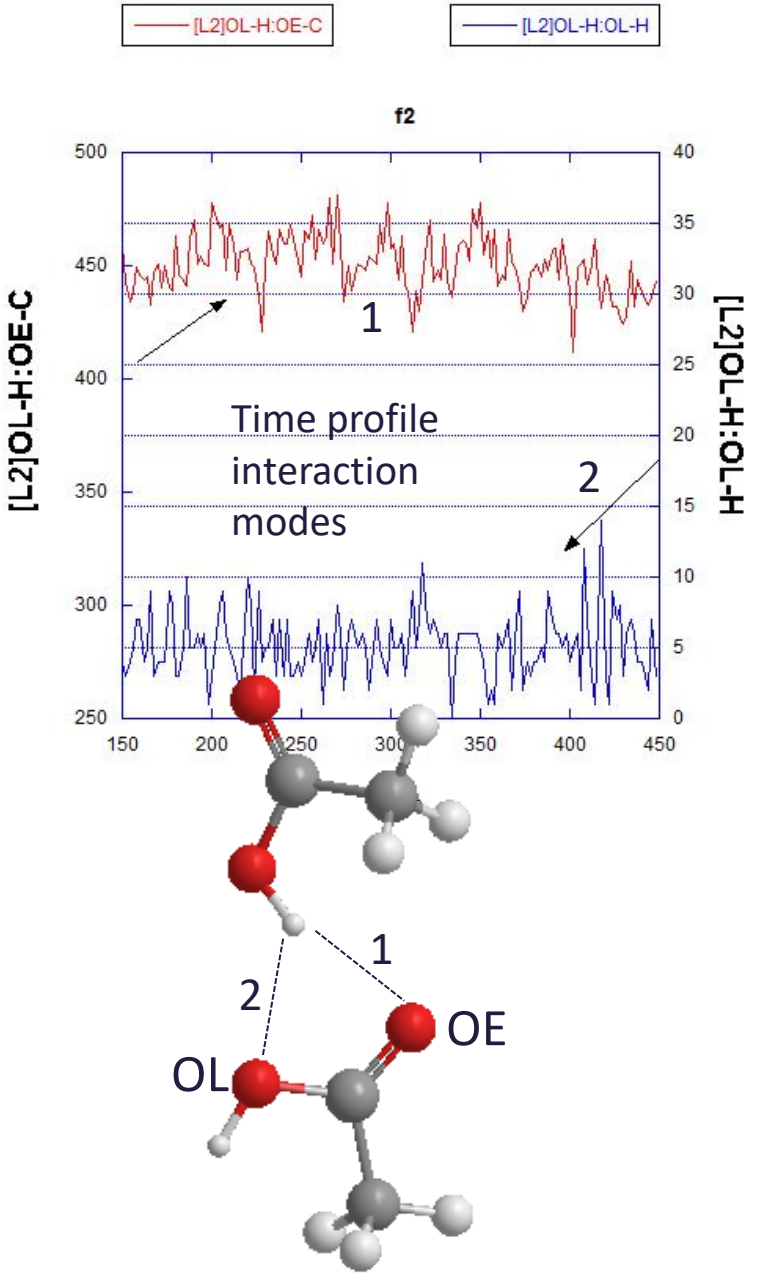


Head-to-tail carboxylic chains

Branching







Science and Technology Facilities Council

Characterisation of atomic interactions 1: correlations

DANAI expression, i	number of interactions
1. [L2]OL-H:OL-H	1.69 ± 1.27
2. [L2]oL-H:OL-h	79.34 ± 7.58
3. [L2]oL-h:oL-h	113.50 ± 8.75
4. [L2]OL-H:OE-C	418.90 ± 14.65
5. [L2]oL-H:OE-c	504.61 ± 10.76
6. [L2]oL-h:oE-c	581.28 ± 8.50
7. [L3]oL-h:oE#oL(h):h-oL	59.64 ± 6.17
8. [L3]oL-h:oE#h:oL-h	92.95 ± 7.38
9. [L3]oL-h:oE#h:oE-c	477.39 ± 13.60
10.[J3]oL-h:oE(c):h-oL	22.32 ± 4.35

Various modes of HB interactions in ethanoic acid liquid

Unpicking of various modes of H-bond interactions and their correlations.
Provide more interpretable details than radial distribution functions.

Correlation analysis (Pearson)

	1	2	3	4	5	6	7	8	9	10
1	1.000	0.056	-0.001	-0.004	-0.080	-0.182	0.005	-0.009	-0.163	-0.063
2		1.000	0.764	-0.610	-0.444	-0.811	0.627	0.644	-0.794	-0.253
3			1.000	-0.834	-0.653	-0.628	0.617	0.853	-0.651	-0.067
4				1.000	0.906	0.526	-0.445	-0.696	0.676	-0.312
5					1.000	0.405	-0.286	-0.506	0.616	-0.546
6						1.000	-0.482	-0.461	0.915	0.364
7							1.000	0.521	-0.457	-0.243
8								1.000	-0.562	-0.094
9									1.000	0.077
10										1.000

Science and Technology Facilities Council

Characterisation of atomic interactions 2: order parameters

Linearity parameter (tortuosity), *l*

perfectly linear $1 \leq l < (x-1)$ highly clustered

HB_20_20

[L3]O20E:H20O:O20L	1.462013
[L3]h20O:O20E:H20O	1.263810
[L3]o20E:H20O:O20E	1.378861
[L3]H20O:O20E:H20O	1.353136
[L4]H20O:O20E:H20O:O20E	1.635499
[L3]O20E:H20O:O20E	1.388601
[L4]o20E:H20O:O20E:H20O	1.706096

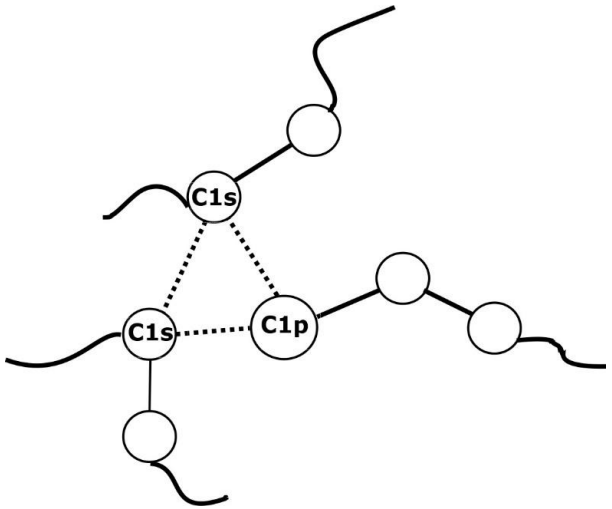
HP_1_1

[L3]c1p:c1p:c1p	1.315507
[L3]c1p:C1p:c1p	1.321935
[L3]c1p:c1p:C1p	1.346630
[L3]c1p:C1p:C1p	1.326914
[L4]c1p:c1p:c1p:c1p	1.536298
[L4]c1p:C1p:C1p:c1p	1.615732
[L4]c1p:C1p:c1p:c1p	1.564829
[L4]c1p:c1p:C1p:C1p	1.631189
[L4]c1p:c1p:c1p:C1p	1.585892
[L4]C1p:c1p:C1p:c1p	1.639232
[L4]C1p:c1p:c1p:C1p	1.649167

HB_20_20
[L4]O20E:H20O:O20E:h20O

carboxylic CG (20)
O20E = carbonyl O
H20O = hydroxyl hydrogen
h20O = hydroxyl hydrogen (non-exclusive)

Triangular order parameter, t_r

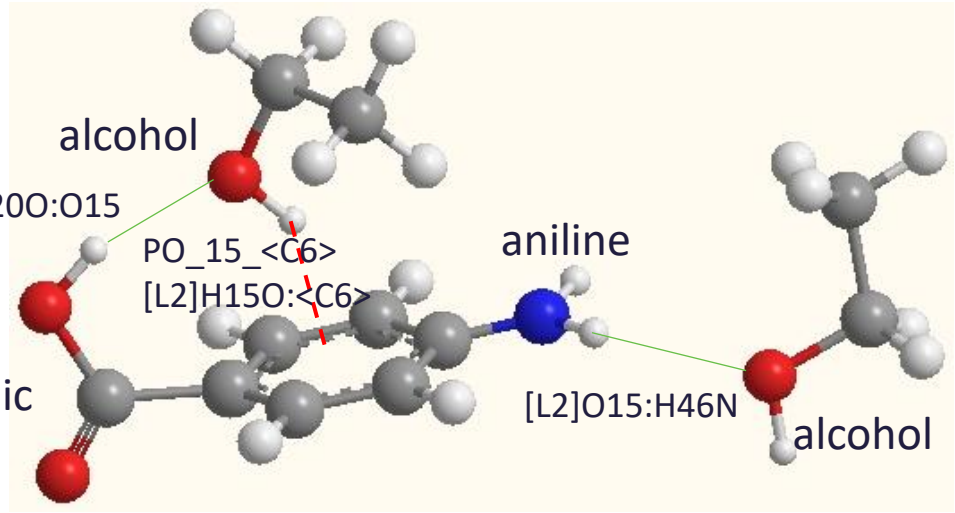


HP_1_1

[R3]c1p:c1p:c1p:c1p	0.988638
[R3]c1p:c1p:C1p:c1p	0.987897

Science and Technology Facilities Council

Example 2: *p*-aminobenzoic acid (PABA) in ethanol solution (1 M)



The following Chemical Groups (CGIs) identified:
benzene (6)
aniline (46)
carboxylic (20)
alkane (1)
alcohol (15)

*** Expected interaction types in the system ***

1 - Hydrogen bond (HB)
Atom pairs involve:
O...H (CGI: 15)
O...H (CGI: 20)
N...H (CGI: 46)

HB_15_15

DANAI	mean	deviation
[L2]O15:H15O	= 3430.419355	27.733495
[L2]o15:H15O	= 795.548387	26.960950
[L2]h15O:o15	= 14.225806	4.285833
[L2]h15O:O15	= 3.774194	1.452688

HB_15_20

DANAI	mean	deviation
[L2]O15:H20O	= 78.838710	0.766087
[L2]H15O:O20E	= 16.645161	2.335537
[L2]H15O:O20L	= 2.354839	1.426667
[L2]H15O:o20E	= 0.774194	1.210430
[L2]O15:h20O	= 0.193548	0.591300

HB_15_46

DANAI	mean	deviation
[L2]H15O:N46p	= 37.096774	3.030542
[L2]O15:H46N	= 64.064516	5.211250
[L2]H15O:n46p	= 1.032258	1.121171
[L2]O15:h46N	= 0.451613	0.836225
[L2]o15:H46N	= 0.322581	0.735597

HB_20_20
Not detected.

HB_20_46

DANAI	mean	deviation
[L2]O20E:H46N	= 0.741935	0.801923
[L2]O20L:H46N	= 0.193548	0.395079
[L2]H20O:N46p	= 0.838710	0.367799

HB_46_46

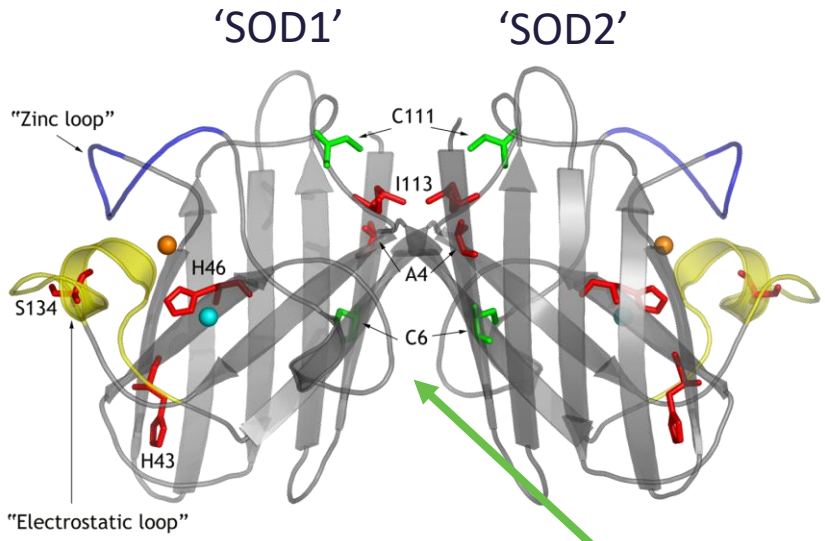
DANAI	mean	deviation
[L2]H46N:N46p	= 0.193548	0.395079

UKRI Science and Technology Facilities Council

Example 3: interfacial interaction of sod1 (superoxide dismutase)

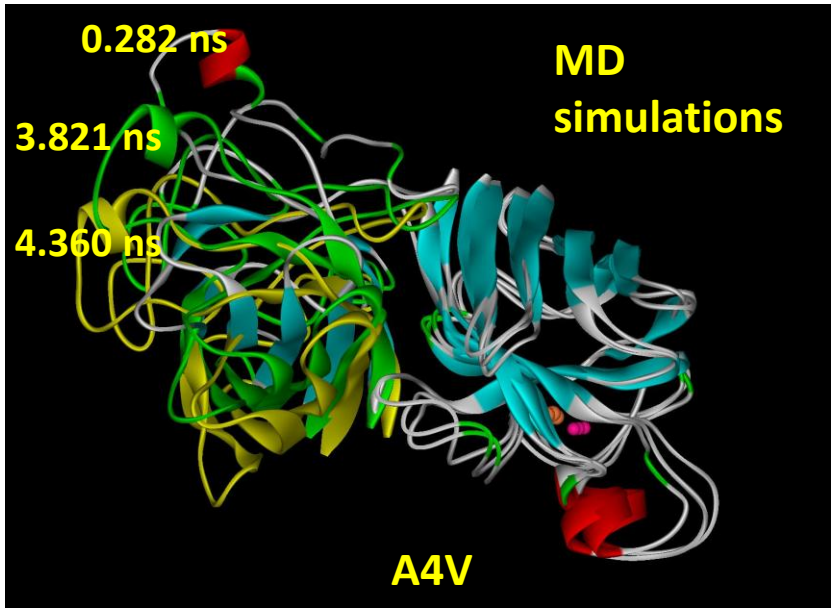
SOD1 mutations are the leading cause in about 20% of familial amyotropic lateral sclerosis (FALS)

SOD1 dimer showing some ALS mutation sites (over a hundred have been identified)



Mutant SOD1 - A4V

- Mutation at interfacial region (far from active sites)
- Experiment and MD indicate rotation of monomers



A possible therapeutic target for disease has been suggested – S. S. Ray and P. T. Lansbury, Jr., *PNAS* **101**, 5701 (2004)

What are the interactions at interface?

Science and Technology Facilities Council

Structure in xyz format (sod1)

```
 4374   time= 0.00000 ps  step= 0.00200
CRYST1   78.909   78.909   78.909  90.00  90.00  90.00
N287  10.895831  -14.604078  -9.445671
H290  11.206366  -14.499845  -8.490263
H290  11.583491  -14.922866  -10.113203
H290  10.239200  -15.369737  -9.497660
C293  10.347367  -13.260753  -9.846646
H140  11.232535  -12.641085  -9.990136
C135  9.459012  -13.389207  -11.087187
H140  8.722326  -14.155294  -10.845311
H140  10.000762  -13.817117  -11.930691
H140  8.887282  -12.489126  -11.313194
C235  9.563745  -12.725381  -8.652987
O236  8.949120  -13.545245  -7.997802
N238  9.640267  -11.399305  -8.429131
H241  9.961442  -10.826210  -9.196273
C224  8.869496  -10.738385  -7.429825
H140  8.443417  -11.538384  -6.824385
C158  9.810496  -10.000010  -6.498428
H140  10.598834  -10.651421  -6.121221
O154  9.154706  -9.501664  -5.299888
H155  8.790055  -10.254603  -4.860416
C135  10.413409  -8.703877  -7.003774
H140  9.605127  -7.973797  -7.045756
H140  10.942601  -8.806027  -7.951202
H140  11.102703  -8.368065  -6.229046
C235  7.705983  -9.969516  -8.159476
O236  7.827362  -9.507176  -9.251123
N238  6.542083  -9.877554  -7.493411
H241  6.484353  -10.172797  -6.529254
C224  5.311272  -9.442195  -8.074798
H140  5.553852  -8.726498  -8.860312
C136  4.425295  -10.620456  -8.505563
H140  3.401466  -10.264192  -8.619365
...
...
```

D_ATA

Select option to analyse interactions between molecules only

CGs identified

```
...
...
Results output file.
Created: 2024-06-20 13:46:10

The following Chemical Groups (CGIs) identified:
NH3-terminus (954)
amino_acid (900)
alanine (901)
amide (44)
threonine (919)
lysine (914)
valine (922)
cysteine (905)
leucine (913)
glycine (908)
aspartic (904)
alkane (1)
glutamine (906)
isoleucine (912)
asparagine (903)
phenylalanine (916)
glutamic (907)
serine (918)
tryptophan (920)
histidine3 (910)
histidine1 (909)
cysteine_S (926)
arginine (902)
COO-terminus (950)
 **** Setup molecules...
 Identified types of molecules as follows:

 Molecule: A
 Total molecules: 2
 Total atoms per molecule: 2187
 Molecular weight: 15796.881890
 Composition: C679 H1077 O224 S203 S4

 **** Completed.
```

Output from D_ATA

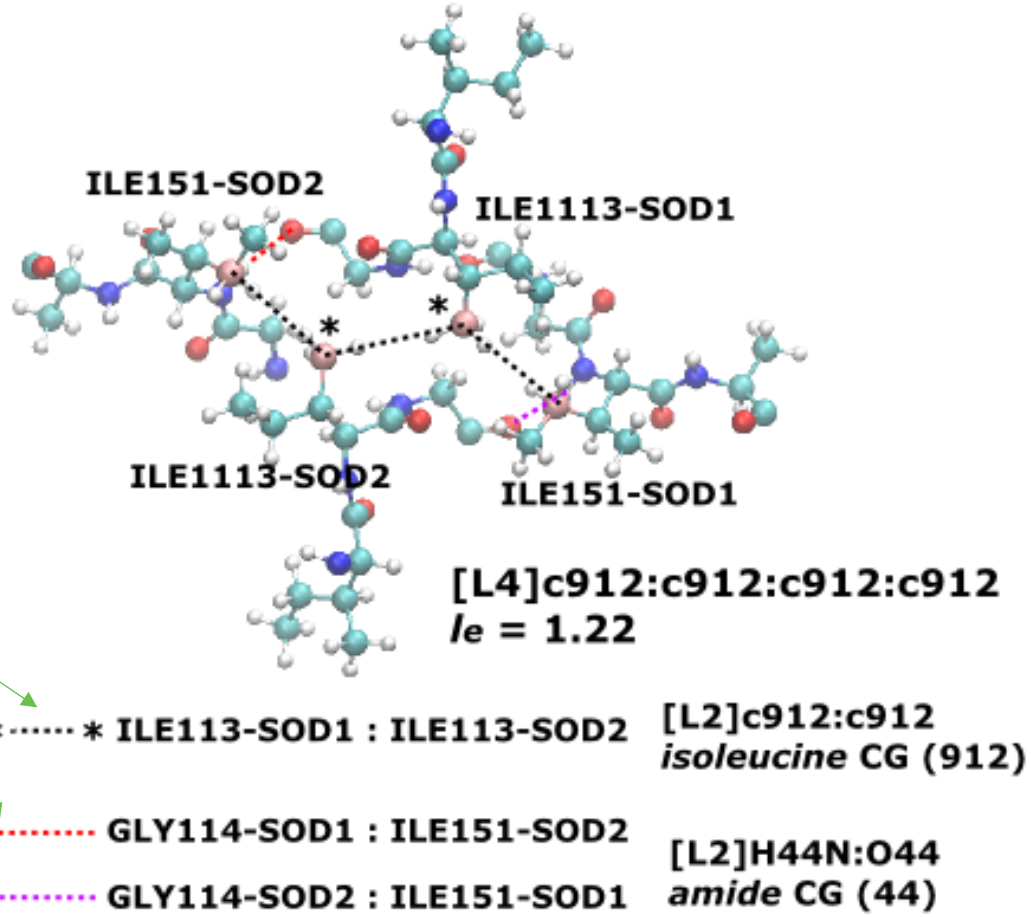
Chemical structure file (DL_F Notation)

```
# Chemical structure file.
# Created from D_ATA version 1.00
# Input file:
# d_ata.input
# Atom_index   ATOM_TYPE  imp_flag   neighbour_number  neighbour1  neighbour2 ...
1     N_NH3-terminus          0 4    2    3    4    5
2     HN_NH3-terminus         0 1    1
3     HN_NH3-terminus         0 1    1
4     HN_NH3-terminus         0 1    1
5     CA_amino_acid           0 4    1    6    7    11
6     HCA_amino_acid          0 1    5
7     CB_alanine              0 4    5    8    9    10
8     HCB_alanine             0 1    7
9     HCB_alanine             0 1    7
10    HCB_alanine             0 1    7
11    C_amide                 1 3    5    12   13
12    O_amide                 0 1    11
13    Ns_amide                1 3    11   14   15
14    HN_amide                0 1    13
15    CA_amino_acid           0 4    13   16   17   25
16    HCA_amino_acid          0 1    15
17    CB_threonine            0 4    15   18   19   21
18    HCB_threonine           0 1    17
19    O_threonine             0 2    17   20
20    HO_threonine            0 1    19
21    CG_threonine            0 4    17   22   23   24
22    HCG_threonine           0 1    21
23    HCG_threonine           0 1    21
24    HCG_threonine           0 1    21
25    C_amide                 1 3    15   26   27
26    O_amide                 0 1    25
27    Ns_amide                1 3    25   28   29
28    HN_amide                0 1    27
29    CA_amino_acid           0 4    27   30   31   47
30    HCA_amino_acid          0 1    29
31    CB_lysine               0 4    29   32   33   34
32    HCB_lysine              0 1    31
33    HCB_lysine              0 1    31
34    CG_lysine               0 4    31   35   36   37
...
...
```

Science and Technology Facilities Council

D_ATA output: residues that involve in cross-monomer interactions

[L2]H44N:O44
GLY51-SOD1 : ILE151-SOD2
GLY114-SOD1 : ILE151-SOD2
ILE151-SOD1 : GLY114-SOD2
ILE151-SOD1 : GLY51-SOD2
[L2]C922:C922
VAL148-SOD1 : VAL148-SOD2
[L2]C922:C908A
VAL5-SOD2 : GLY51-SOD1
VAL148-SOD2 : GLY150-SOD1
VAL148-SOD1 : GLY150-SOD2
[L2]C912:C912
ILE113-SOD1 : ILE113-SOD2
[L2]C912:C908A
ILE113-SOD1 : GLY114-SOD2
ILE151-SOD1 : GLY114-SOD2
ILE113-SOD2 : GLY114-SOD1
[L2]C912:C902Z
ILE151-SOD1 : ARG115-SOD2
ILE151-SOD2 : ARG115-SOD1
[L2]C912:C902
ILE151-SOD1 : ARG115-SOD2
…
…



Science and Technology Facilities Council

Running D_ATA

Using D_ATA

Registration:
http://www.ccp5.ac.uk/D_ATA

Unpack program:
tar -xvf *d_ata_1.1.tar.gz*

Compile program:
cd source
make clean
make

Run program:
./d_ata

Science and Technology Facilities Council

D_ATA Directories and files

/source – source codes, where you do the compilation
/lib – Standard library file
Interaction reference data, *DLF_map*
Chemical Group reference file, *DLF_notation*
/output – including results
d_ata.results, count_XX.results, d_ata.output, d_ata.csf
/examples – Some example *xyz* structures

d_ata.control – control file
d_ata.input – Input configuration/trajectory files.
d_ata.csf – Chemical structure file.

Science and Technology Facilities Council

The *d_ata_path* file

Master file-locator that indicates where all relevant input and output files are found. Must always placed in the folder where *d_ata* executable file is located (D_ATA home).

```
# Directory paths for DL_ATA

CONTROL_FILE = d_ata.control
INPUT_FILE = d_ata.input
OUTPUT_PATH = output/
LIBRARY_PATH = lib/
```

d_ata.control

System size for PBC from input files

structural criteria

Interation criteria

```
Example control file with recommended settings good for most cases.
2           * Program mode: 1=typer, 2=typer and analyser.
1           * All-inclusive interaction mode (1=yes, 0=no).
none         * PDB, gro template, or 'none'
auto         * Periodic boundary condition type
40.0 0.0 0.0   * Cell vector a (x, y, z)
0.0 40.0 0.0   * Cell vector b (x, y, z)
0.0 0.0 40.0   * Cell vector c (x, y, z)
all         * Group A: Atom range. Min Max or 'all'
1           * Molecule-based analysis (0 = off, 1=between molecule, 2=within molecule).
0           * No of configuration to skip
1           * DANAI tier level (must always set to 1).
0.001        * Average count fraction threshold
2.5          * Distance threshold for H...D in HB interactions
120.0        * Angle threshold for HB interactions
4.5          * Distance threshold for C...C in HP interactions
1           * HB analysis
1           * HP analysis
```

d_ata.input

```
1
examples/example1.xyz
examples/example2.xyz.gz
```

Run example trajectory files: see Chapter 10 of D_ATA user manual.

Tutorial exercises:
https://dl-sdg.github.io/RESOURCES/EXERCISES/exercises_data.html

Reference Guide:
https://dl-sdg.github.io/RESOURCES/TUTORIALS/data_0.html

THE END

Science and Technology Facilities Council