Science & Technology Facilities Council

UK Research and Innovation

DL_Software
Enabling Discovery

DL_FIELD/DL_ANALYSER
Force field and analysis tools for DL_POLY

Dr C W Yong
Scientific Computing Department,
STFC,
Daresbury Laboratory, Sci-Tech Daresbury,
Warrington WA4 4AD

(DL_Software training workshop, Chile, May 2019)

CoSeC
Computational Science Centre for Research Communities

EPSRC
Engineering and Physical Sciences Research Council

CCP5

CCPBioSim

HEC MCC

alc
ada lovelace centre

ADDOPT

Overview

- What is DL_FIELD and why?
- DL_FIELD features, what it can do.
- DL_F notation, DANAI.
- Demo

DL_Software
Enabling Discovery

DL_FIELD program

- One of the DL_Software program component – a collective term for computational chemistry software developed at Daresbury Laboratory (http://www.ccp5.ac.uk/software)
- First DL_FIELD version Oct 2010. Since then, registered user > 1400
- Current version 4.4, going to release 4.5, June 2019.
- A computer program package written in C that primarily serves as a support application software tool for DL_POLY molecular dynamics (MD) simulation package.
- Important application tool to enhance the usability of DL_POLY MD simulation package and to facilitate the use of a wide range of advance features included in the DL_POLY program.

DL_FIELD Functions

(1) Force field model convertor: DL_FIELD converts user's atom models, into file formats (CONFIG, FIELD) that are recognisable and ready to run in DL_POLY_2.19, DL_POLY_3 and DL_POLY_4 programs with minimum user's intervention.

(2) Force field editor: DL_FIELD allows user to edit or modify a particular force field (FF) scheme to produce a customised scheme that is specific to a particular simulation model. (Inclusion of new features to a FF scheme will always make available to all, whenever possible)

(3) Force field model repertoire: Easily expand the existing library to include user-defined models.

(4) Full automatic atom typing and identification of chemical nature of atoms.

DL_FIELD Development

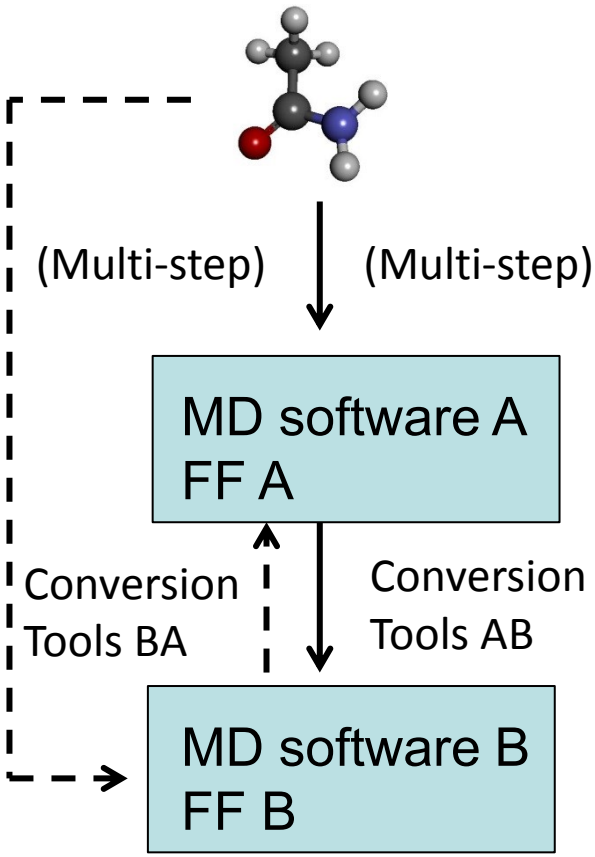
Philosophy: minimise the requirement to understand detailed knowledge and inner workings of force field descriptions. A user-friendly software tool that automatically processes the molecular information with minimum user's intervention.

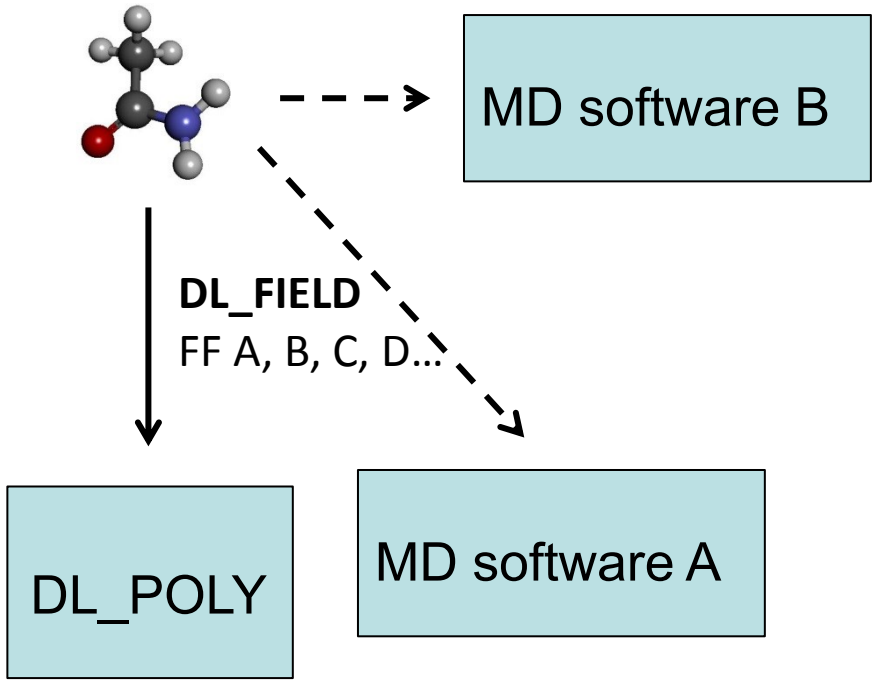
Able to access to different types of potential schemes all in single, easy to read format (all-atom, united-atom, inorganic)

Robust, be able work on typical PC. Flexible, general: from a single molecular system to complex fully-solvated biological systems containing several million atoms.

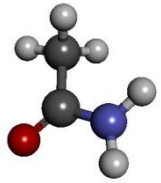
Unique features: (1) Universal atom typing (DL_F Notation), (2) full integration of various force field schemes, from such, (3) multiple-potential capabilities.

Conversion Software tools





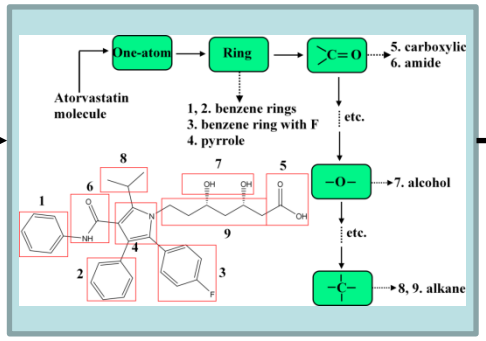
Molecular simulation engine infrastructure



User's structure

```
O  2.129000  0.063000  2.421000
C  2.004000  0.182000  1.026000
H  1.513000 -1.209000 -0.602000
C  1.611000 -1.202000  0.507000
C  0.907000  1.179000  0.657000
H  0.657000  1.067000 -0.426000
C  1.146000  2.672000  0.936000
H -0.012000  0.872000  1.214000
C  2.015000  2.959000  2.178000
H  0.148000  3.161000  1.037000
```

DL_FIELD



DL_F Notation

FF model

DL_POLY

Force field model,
Configuration file

Dynamic computations

Molecular trajectory files

DL_ANALYSER

Filter module **DANAI**

Formatted atomic structure

Structure extraction, analysis

Current, future development

[Template file]

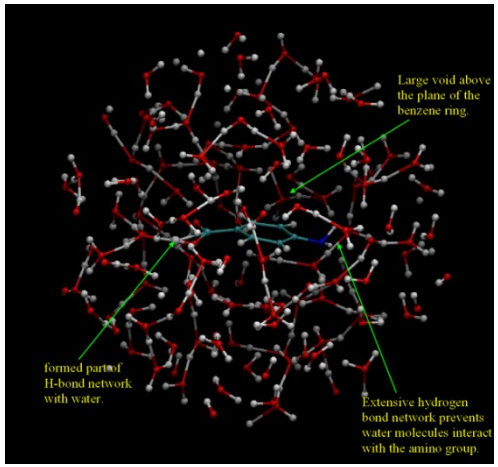
DL_MONTE (monte carlo)

ChemShell (QM/MM)

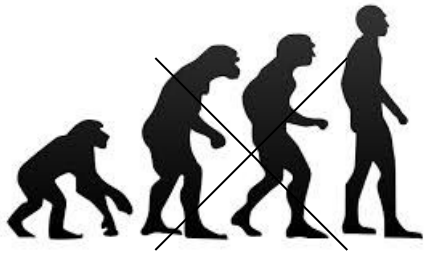
Force field model

Configuration file

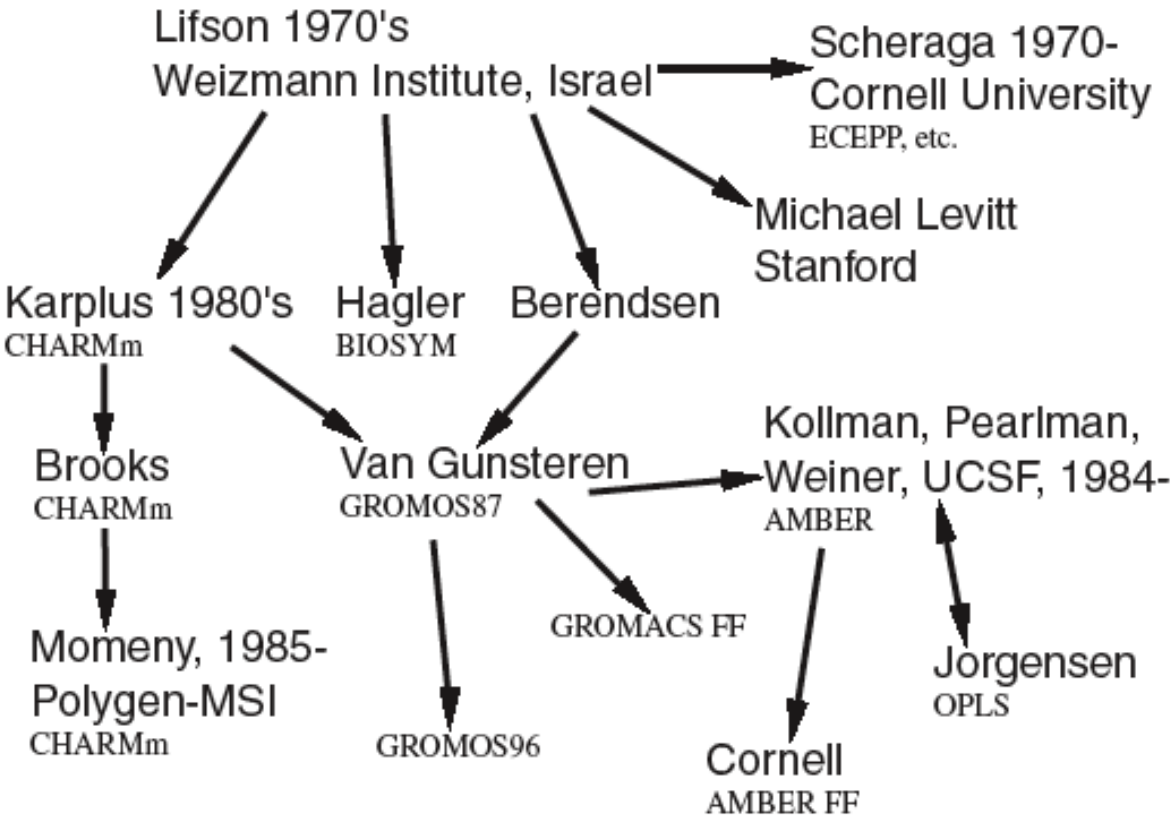
Control file.



Evolution of FF



A set of interactions constitutes a *force field* (FF)



FF model set is not trivial

- **Atom typing**: no general consensus, universal format, naming, guidance
- Different FF schemes have their own way of setting things up
- Different functional forms
- Wide variation of file formats, often not easy to interpret.
- Migration of FF models ?
- How to setup multi-component systems? Bio-inorganic.
- User friendliness? Ad-hoc scripts?

Reluctance of using different FFs. Restriction to new model setup and development.

DL_FIELD is developed to overcome these barriers

Available Force Field schemes

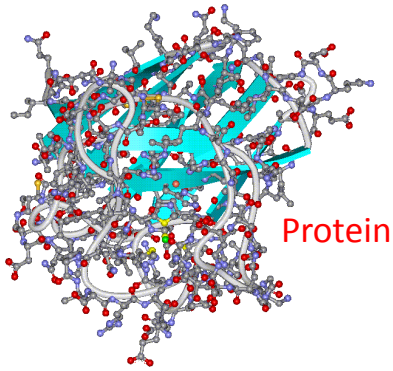
- CHARMM –CHARMM22_prot, CHARMM36_prot, CHARMM36_lipid, CHARMM36_CGenFF, CHARMM36_carb.
- AMBER – proteins, Glycam – sugars, glycans.
- OPLS2005 – proteins, organic molecules
- OPLS-AA/M – proteins
- PCFF – Small organic molecules, organic polymers.
- CVFF – Small organic molecules, proteins.
- DREIDNG – General force fields for organic and other covalent molecules.
- INORGANIC –binary_oxides, glass, clay
- CHARMM19 – united atom with explicit polar H.
- G54A7 – united atom Gromos FF.

Future inclusion: FF for ionic liquids, general-purpose FF, coarse-grain FF

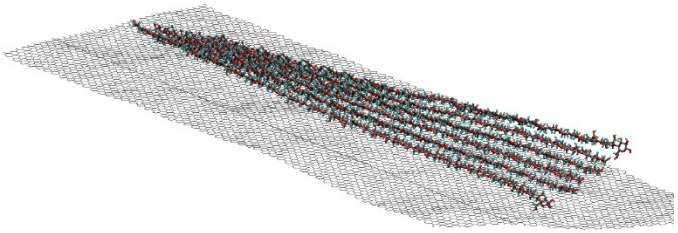
All FFs expressed in the same format.

Model Conversions

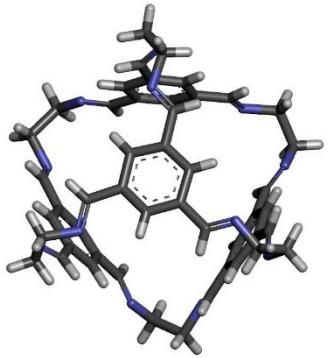
- Organic molecules, from simple to complex: covalent molecules, amino acids, proteins, glycans, polymers.
- Complex 3D networked structures: graphenes, molecular cages.
- Random structures – gels, polymers.
- Inorganic materials – simple ionic oxides, minerals
- Mixed component materials – bio-inorganic



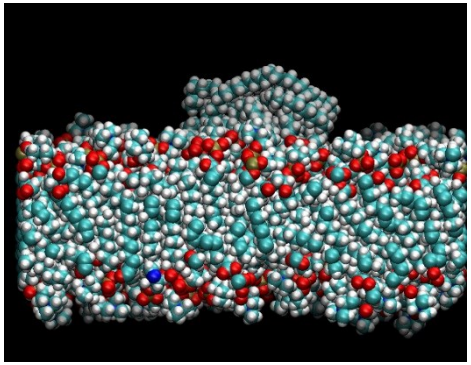
Protein



Carbohydrate on graphene



Organic cage



Polymer on
lipid membrane

DL_FIELD Main Features

- Convert a PDB, *xyz, mol2* structures to CONFIG and FIELD files for DL_POLY.
- Assign freeze and tether atoms, bond constrains.
- Define rigid bodies, and freeze, tether, constrain atoms.
- Pseudo point assignment.
- Core-shell polarisability 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 atom typing using a universal notation** (*xyz*)
- **Multiple potentials** (bio-inorganic)
- **Solution Maker** – set up disordered system.
- Running DL_POLY (*fork*)

DL_FIELD Files

- Each force field scheme consists of

(1) Structure files (*.sf*) – define MOLECULE templates, ATOM_TYPEs
(2) Parameter files (*.par*) – potential parameters
(3) Misc. supporting file (BCI)

- The *udff* file – user-define force field
- The *control* file.
- *DLF_Notation* – Chemical Group list and specific conversion rules
- *dl_field.atom_type*

DL_FIELD Conversion Procedures

Just a summary!

1) Read *control* file.

2) Examine user configuration files.

3) Carry out **atom typing** procedures to obtain ATOM_TYPEs (**Conversion mechanism**).
   (i) Looking for matching templates, in *sf* files (for *PDB*).
   (ii) Molecular topology analysis (for *xyz, mol2*).

4) Assign parameters, obtained from *par* file.

5) Produce *dl_poly.CONFIG*, *dl_poly.FIELD* and *dl_poly.CONTROL* files

(i) Conversion Mechanism – Template Based

- Useful for specific class of molecules – proteins, carbohydrates.
- Largely based on molecular template fit in terms of bond connectivity.
- A molecular template must be explicitly defined, indicating the charges and types of atoms that made up the molecule.
- Possible to use auto-CONNECT feature to simplify the template definition. Useful for complex molecules.
- Flexible, allows users to adjust the model behaviour – introduce constrains, rigid body, core-shell, pseudo points etc.
- User's structures in PDB format.
- Cons: tedious. Need to decide atom types.

MOLECULE Template

MOLECULE_TYPE **Total atom**

DLPOLY_CHARMM.sf

MOLECULE definition
(MOLECULE template)

```
MOLECULE acetamide   9   0.0
C1    C_aliphatic3  -0.27
H11   H_aliphatic    0.09
H12   H_aliphatic    0.09
H13   H_aliphatic    0.09
C     C_amide        0.55
O     O_amide       -0.55
N     N_amide       -0.64
H1    H_hydroxyl     0.32
H2    H_hydroxyl     0.32
CONNECT C1  > H11 H12 H13 C
CONNECT H11 > C1
CONNECT H12 > C1
CONNECT H13 > C1
CONNECT C   > C1 O N
CONNECT O   > C
CONNECT N   > C H1 H2
CONNECT H1  > N
CONNECT H2  > N
IMPROPER C  C1 N  O
IMPROPER C  N  C1 O
IMPROPER N  C  H1 H2
IMPROPER N  C  H2 H1
END MOLECULE acetamide
```

```
     H11 O    H1
     |   ||   /
H12-C1--C--N
     |      \
     H13     H2
```

ATOM_LABELs within this *MOLECULE.*
(do not confuse with* ATOM_KEY*s)

Total charge

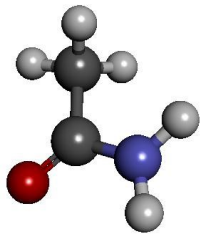
Molecular sketch (optional)

Atomic charges

ATOM_TYPEs

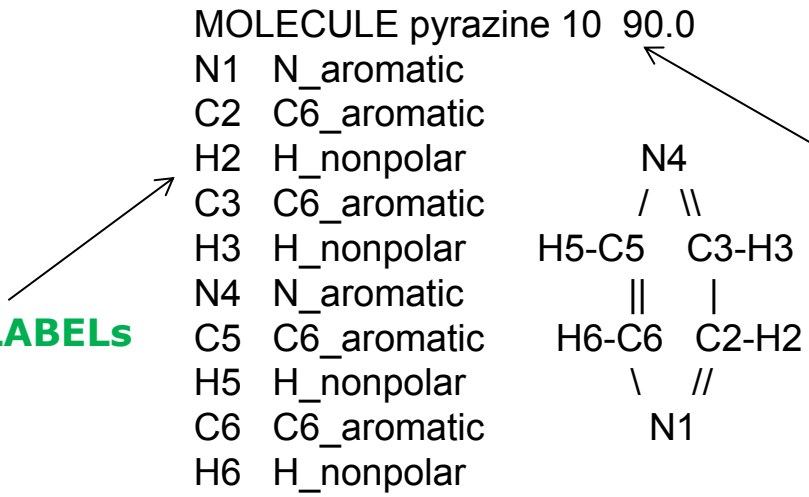
Connectivity information

Optional directives



DLPOLY_PCFF.sf

```
MOLECULE pyrazine 10  90.0
N1   N_aromatic
C2   C6_aromatic
H2   H_nonpolar
C3   C6_aromatic
H3   H_nonpolar
N4   N_aromatic
C5   C6_aromatic
H5   H_nonpolar
C6   C6_aromatic
H6   H_nonpolar
CONNECT  N1  > C6 C2
CONNECT  C2  > H2 C3 N1
CONNECT  H2  > C2
CONNECT  C3  > C2 N4 H3
CONNECT  H3  > C3
CONNECT  N4  > C5 C3
CONNECT  C5  > H5 N4 C6
CONNECT  H5  > C5
CONNECT  C6  > C5 N1 H6
CONNECT  H6  > C6
INVERSION H2 C2 C3 N1
INVERSION H3 C3 C2 N4
INVERSION H6 C6 C5 N1
INVERSION H5 C5 C6 N4
END MOLECULE pyrazine
```



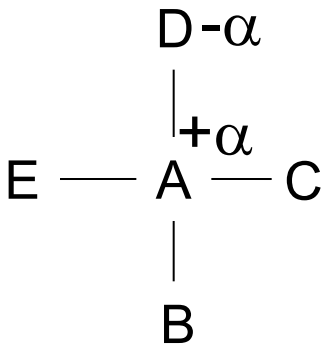
ATOM_LABELs

Charges determined from bond charge increment (BCI)

Total charge (unrealistic, to indicate BCI)

BCI lists partial charges, a, of all ATOM_KEY pairs

Charge on A = B+C+D+E



Use of auto-CONNECT feature to define MOLECULE for a class of molecules

For example: ethanol, propanol, butanol etc ‘all-in-one’

```
MOLECULE aliphatic_alcohol 6 -0.22      Any aliphatic alcohol, except methanol
C3  C_aliphatic3   -0.27
C2  C_aliphatic2    0.05       H1 H1    H1
C1  C_aliphatic1    0.14        |    |       |
H1  H_aliphatic     0.09    H1-C3-C2-...C2-OH-HO
OH  O_hydroxyl     -0.66        |    |       |
HO  H_hydroxyl      0.43       H1   H1    H1
CONNECT C3 > 4 auto
CONNECT C2 > 4 auto
CONNECT C1 > 4 auto
CONNECT H1 > 1 auto
CONNECT OH > 2 auto
CONNECT HO > 1 auto
END MOLECULE
```

List only what ATOM_KEYs are needed.
NOTE: These are no longer ATOM_LABELs

Total charge refers to the sum of the list of ATOM_TYPEs, NOT the actual alcohol molecules (which is zero).

DLPOLY_INORGANIC_ternary_oxides.sf

MOLECULE calcium_carbonate1 5 0.0 Xiao et. al., J. Phys. Chem. vol. 115, pg 20067-20075 (2011)
Ca_1 Calcium_ion_1 1.668 OAR1 A New transferable Forcefield for
CAR1 C_carbonate1 0.999 | Simulating the Mechanics of CaCO3 Crystals
OAR1 O_carbonate1 -0.889 Ca CAR1
OAR1 O_carbonate1 -0.889 / \
OAR1 O_carbonate1 -0.889 OAR1 OAR1
CONNECT Ca_1 > 0 AUTO
CONNECT CAR1 > 3 AUTO
CONNECT OAR1 > 1 AUTO
CONNECT OAR1 > 1 AUTO
CONNECT OAR1 > 1 AUTO
ANGLE ONLY OAR1 CAR1 OAR1
IMPROPER CAR1 OAR1 OAR1 OAR1
END MOLECULE

Optional Directives for MOLECULE Templates

- ***IMPROPER*** ATOM_LABEL1 ATOM_LABEL2 ATOM_LABEL3 ATOM_LABEL4

- ***INVERSION*** ATOM_LABEL1 ATOM_LABEL2 ATOM_LABEL3 ATOM_LABEL4

- ***SHELL*** ATOM_LABEL1 ATOM_LABEL2

- ***DIHEDRAL ONLY*** ATOM_LABEL1 ATOM_LABEL2 ATOM_LABEL3 ATOM_LABEL4
- ***DIHEDRAL REMOVE*** ATOM_LABEL1 ATOM_LABEL2 ATOM_LABEL3 ATOM_LABEL4
- ***DIHEDRAL OFF***

- ***ANGLE ONLY*** ATOM_LABEL1 ATOM_LABEL2 ATOM_LABEL3
- ***ANGLE REMOVE*** ATOM_LABEL1 ATOM_LABEL2 ATOM_LABEL3
- ***ANGLE OFF***

- ***BOND ONLY*** ATOM_LABEL1 ATOM_LABEL2
- ***BOND REMOVE*** ATOM_LABEL1 ATOM_LABEL2
- ***BOND OFF***

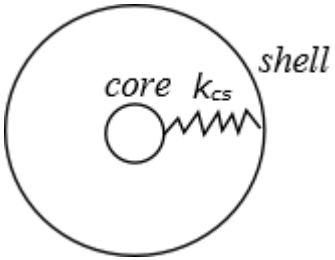
- ***RIGID*** ATOM_LABEL1 ATOM_LABEL2 …

- ***THREE-BODY*** ATOM_LABEL1 ATOM_LABEL2 ATOM_LABEL3

- ***EXCLUSION_14*** ATOM_LABEL1 …

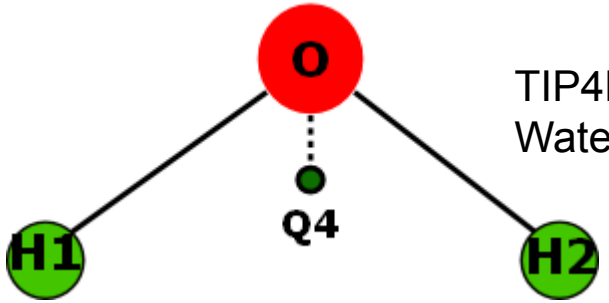
Template Customisation

Core-shell model



The core-shell model is usually employed in inorganic materials to simulate polarisability effect of an atom such as oxygen.

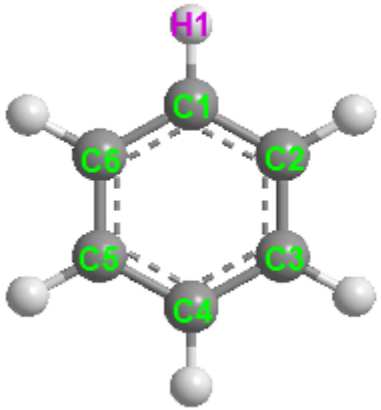
Pseudo points



TIP4P
Water model

An off-center point charge water model.

Rigid body



Note that all the hydrogen atoms are still flexible relative to the rigid ring. The following [***RIGID***] directive defines one of the the hydrogen atom to be part of the rigid body:

RIGID C1 H1 C3 C2 C6 C4 C5

DL_FIELD Limitation:
Each [***RIGID***] statement in DL_FIELD refers to a single rigid unit within a MOLECULE that can only contain up to 15 atoms. This is different from DL_POLY, which allows more than that but confine to 15 atoms in a line statement in the *FIELD* file.

(ii) Conversion Mechanism – Molecular Topology Analysis

- Determine chemical nature of every atom in system. Make use of **DL_F Notation** for atom typing (more about this later)

- Do not need to create MOLECULE template

- Do not need to decide ATOM_TYPEs.

- Everything is done automatically.

- Reads *xyz, mol2* files.

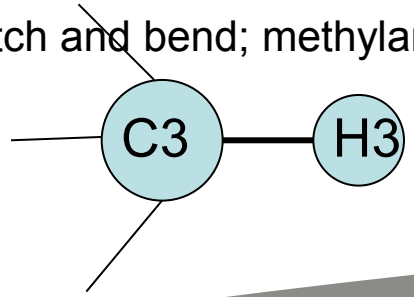
Parameter File (.par)

- Consists of all types of interactions with the exception of *coulombic charges* (defined for every ATOM in the *.sf* file).
- Each type of interaction is enclosed within the appropriate ***DIRECTIVES***.
- Examples: Bond definitions. $U(b) = k(b - b_0)^2$

BOND	k	b_0	remark
…			
C3 H3	330.0	1.0800	PEP Alanine Dipeptide ab initio calc's (LK)
C3 N1	240.0	1.4550	ALLOW POL, methylamine geom/freq, adm jr., 6/2/92
C3 N2	320.0	1.4300	ALI PEP POL ARO, NMA Gas & Liq. Phase IR Spectra
C3 N3	200.0	1.4800	ALI POL new stretch and bend; methylammonium (KK)
…			
END BOND			

ATOM_KEYS

Potential parameters



Other *DIRECTIVES* in *.par* file:

ANGLE	***END ANGLE***
DIHEDRAL	***END DIHEDRAL***
INVERSION	***END INVERSION***
IMPROPER	***END IMPROPER***
SHELL	***END SHELL***
VDW	***END VDW***
VDW_FIX	***END VDW_FIX***
THREE_BODY	***END THREE_BODY***
EQUIVALENCE	***END EQUIVALENCE***

Note: Can only have one set of ***DIRECTIVE***s in a *.par* file.

User-define force field (udff) File

- Define new molecular structures 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*

For example: override the angle parameter

In *DLPOLY_CHARMM.sf*

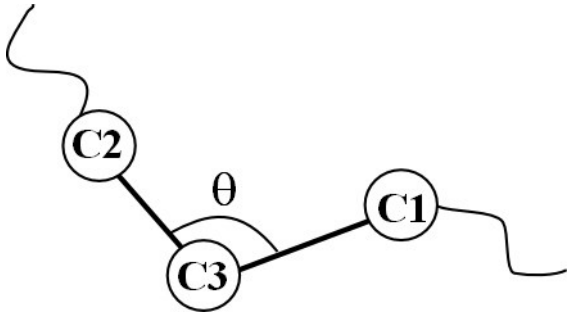
ANGLE

...

C1 C3 C2 58.35 113.50 11.16 2.561 alkane freq (MJF), alkane geom (SF)

...

END ANGLE



In *udff*

ANGLE

C2 C3 C1 **68.35** 113.50 **12.16** 2.561 my new parameter

END ANGLE

DL_FIELD will use these values

For example: override the MOLECULE methenol

In *DLPOLY_CHARMM.sf*

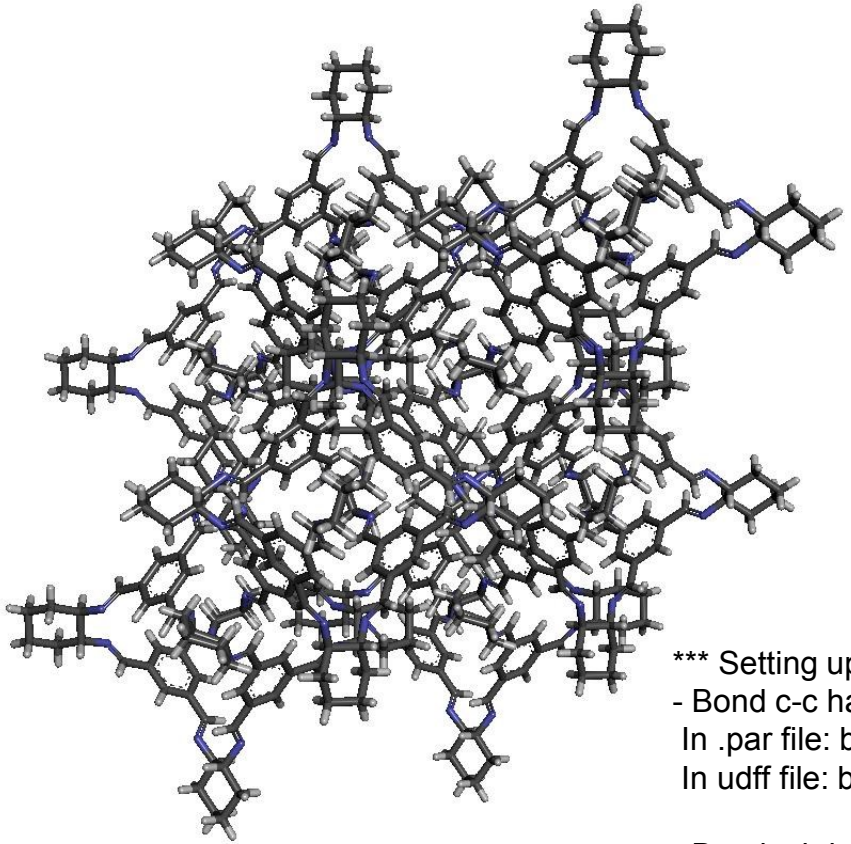
```
MOLECULE_TYPE
…
methanol          MeOH  32.042    override MeOH
…
END MOLECULE_TYPE

MOLECULE methanol 6 0.0    charmm22_protein
C1   C_aliphatic3  -0.04          H11
H11  H_aliphatic    0.09           |
H12  H_aliphatic    0.09     H12-C1-O-H
H13  H_aliphatic    0.09           |
O    O_hydroxyl    -0.66          H13
H    H_hydroxyl     0.43
CONNECT C1  > H11 H12 H13 O
CONNECT H11 > C1
CONNECT H12 > C1
CONNECT H13 > C1
CONNECT O   > C1 H
CONNECT H   > O
END MOLECULE
```

In *udff* file

```
MOLECULE_TYPE
methanol          MeOH  32.042    override MeOH
END MOLECULE_TYPE

MOLECULE methanol 6 0.0    modified MeOH
C1   C_aliphatic3  -0.04          H11
H11  H_aliphatic    0.09           |
H12  H_aliphatic    0.09     H12-C1-O-H
H13  H_aliphatic    0.09           |
O    O_hydroxyl    -0.60          H13
H    H_hydroxyl     0.37
CONNECT C1  > H11 H12 H13 O
CONNECT H11 > C1
CONNECT H12 > C1
CONNECT H13 > C1
CONNECT O   > C1 H
CONNECT H   > O
END MOLECULE
```

Change partial charges For O and H

Organic cage (8 units)



PDB structure produced from Materials Studio.

Force field conversion in DL_FIELD using PCFF force field with user-defined parameters

*** Setting up potential set from standard lib/DLPOLY_PCFF.par file...
- Bond c-c has been overrided by dl_field.udff.
 In .par file: b0=1.530000 K2=299.670000 K3=-501.770000 K4=679.810000
 In udff file: b0=1.536000 K2=252.200000 K3=-513.200000 K4=407.900000

- Bond c-h has been overrided by dl_field.udff.
 In .par file: b0=1.101000 K2=345.000000 K3=-691.890000 K4=844.600000
 In udff file: b0=1.121000 K2=346.800000 K3=-691.890000 K4=544.600000

DL_FIELD Control File

```
This is the title line. Put what you want. Reads only 120 columns
multiple            * 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)
```

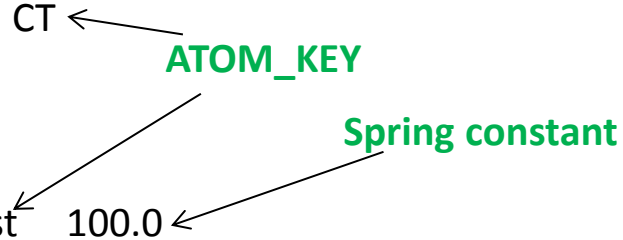
Options

Brief statement, describing all available options

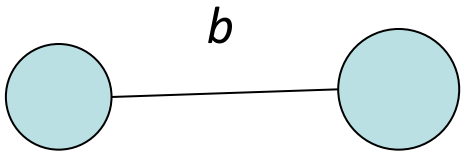
A complete list of control features

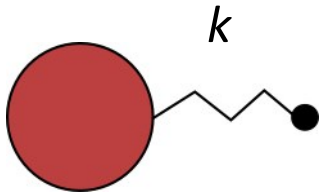
Atom State Selection (in DL_FIELD control file)

```
#############################################################
Atom state specification:  type  Molecular_Group   filter   [value]

FREEZE    ORG1   cp
FREEZE    ORG2   CT

RIGID A

TETHER    CLY    st    100.0

CONSTRAIN  ORG1   h-bond
CONSTRAIN  ORG2   h-bond
CONSTRAIN  not_define h-bond
CONSTRAIN  MOL    h-bond

#############################################################
```



Bond constrains, freeze, tether atoms





Filters for CONSTRAIN

h-bond
all
rigid_water

Filters for FREEZE and TETHER directives

all_backbone
c-alpha
atom_type
all

Usage:
DIRECTIVE Molecular_Group *filters*
CONSTRAIN *A1* *h-bond*

DL_POLY Control Section (in DL_FIELD control file)

```
************** DL_POLY control ******************
1              * Run DL_POLY program
DLPOLY.Z   * DL_POLY executable filename
/home/dl_field_4.5/output * absolute path to DL_POLY program
0              * MM calculation 1=on  0=off
1   3          * Equilibration on(1)/off(0)  level (1,2 or 3)
1000       * Timestep for DL_POLY runs.
9.0          * cutoff (vdw and electrostatic)
10000      * Time limit for DL_POLY run (in seconds)
```

The *dl_f_path* File

Can have multiple versions of *control* file. You name it.

```
# Directory paths for DL_FIELD version 4.4 onwards.
# C W Yong, October 2018
# 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/

# DL_FIELD control files
control = dl_field.control
# control = control_files/example1.control
# control = tutorial/tutorial_1.control
```

Lets walk for your first run.

First Run

Start from a protein PDB file, run DL_FIELD.

(a) Edit *dl_f_path* file to point to a *control* file.

(b) Edit *control* file.
Choose a FF.
Setup FF model.
Solvate TIP4P
Automatic insertion of counter ions.

(c) Run DL_POLY
Rename and move *dl_poly.CONFIG* and *dl_poly.FIELD* to become CONFIG and FIELD file.
Run DL_POLY (MM energy)

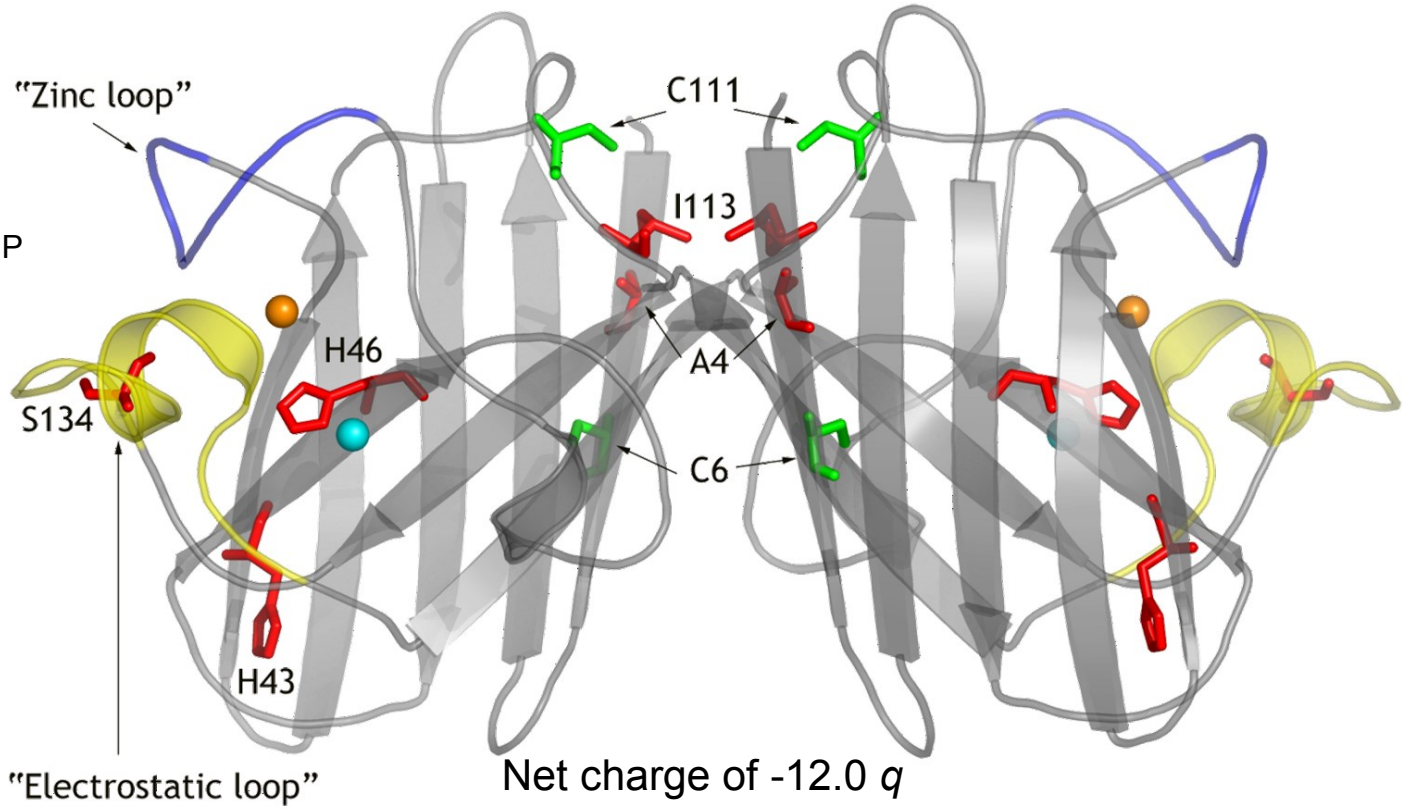
All in one go.

Example: Apo-SOD1 (dimer protein, 153 residues per dimer)

dl_field.control1
charmm22_prot
amber
opls-aam

Solvate with TIP4P
Counter ion

Run DL_POLY



Wild type like mutants

Mutations at the metal binding region results in deficient in metals (*apo*) cause structural disorder of channel loops.

SOD1 dimer showing some ALS mutation sites

R. Strange et. al. *PNAS* **104**, 10040 (2007)

Unique Features

DL_FIELD

DL_F Notation
Full automatic determination of the chemical nature of every atom in molecular systems.

Multiple potentials
Capable to setup complex mixed component systems such as bio-inorganic Systems. Enable setting up of novel simulation models.

DL_ANALYSER

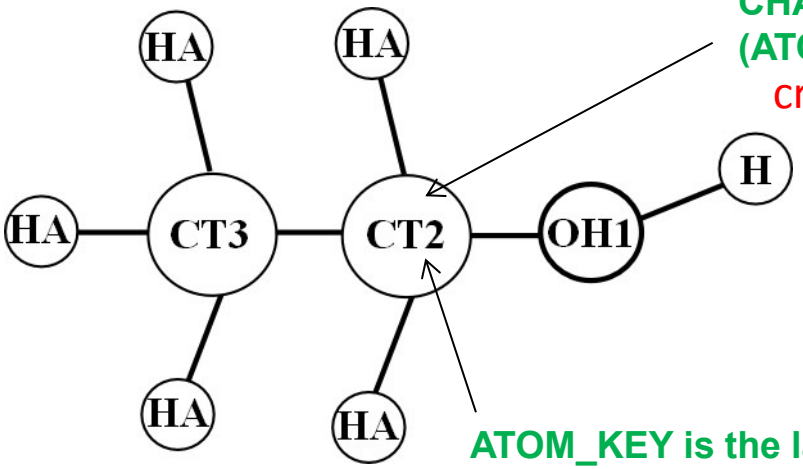
DANAI
Detect, annotate and quantify specific atomic interactions in the systems.

DL_F Notation

A universal atom typing implemented within the DL_FIELD conversion engine.

Atom typing

Procedure to decide the type of atom (ATOM_TYPE) that are referenced to a specific atom label (ATOM_KEY); and from such, to assign the appropriate force field parameters for the atom that involves in a given set of interaction.



CHARMM (ATOM_KEY) cryptic

ATOM_KEY is the label that will be used in DL_POLY's FIELD and CONFIG files

Different ATOM_KEYs within a MOLECULE .

Standard force fields

ATOM_TYPE -> ATOM_KEY

Can be cryptic always cryptic

DL_FIELD without DL_F Notation

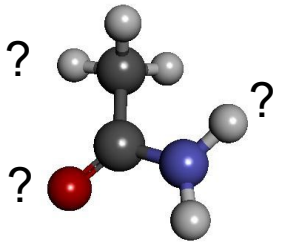
ATOM_TYPE -> ATOM_KEY

Human-readable always cryptic

DL_FIELD with DL_F Notation

ATOM_TYPE -> ATOM_KEY

Human-readable human-readable



Atom typing assignment

Some existing implementations

Templates, Logic syntax, symbolic syntax, file scripts

Have a guess, what are these atoms (CT and C791)?

Typing procedure (cryptic)

(MacroModel)
SMILE-like:
CT = [CX4](CO)(F)(F)(F) (symbolic syntax)

(Vanderbilt U – Concept paper, 2016)
C791 : type = C &
count(bonded_atoms(type = F)) = 3 &
count(bonded_atoms(type = C)) = 1 (logic syntax)

Answer: -**C**F_3 Trifluoro carbon

ATOM_KEYs cryptic

CT – OPLS (in MacroModel)

C791 – OPLS (in Tinker)

Atom typing in DL_F Notation

- Expression of the standard atom typing that is contiguous across a range of force field.

- Universal typing that smoothen data transition with minimum learning curve when migrating from one force field scheme to another.

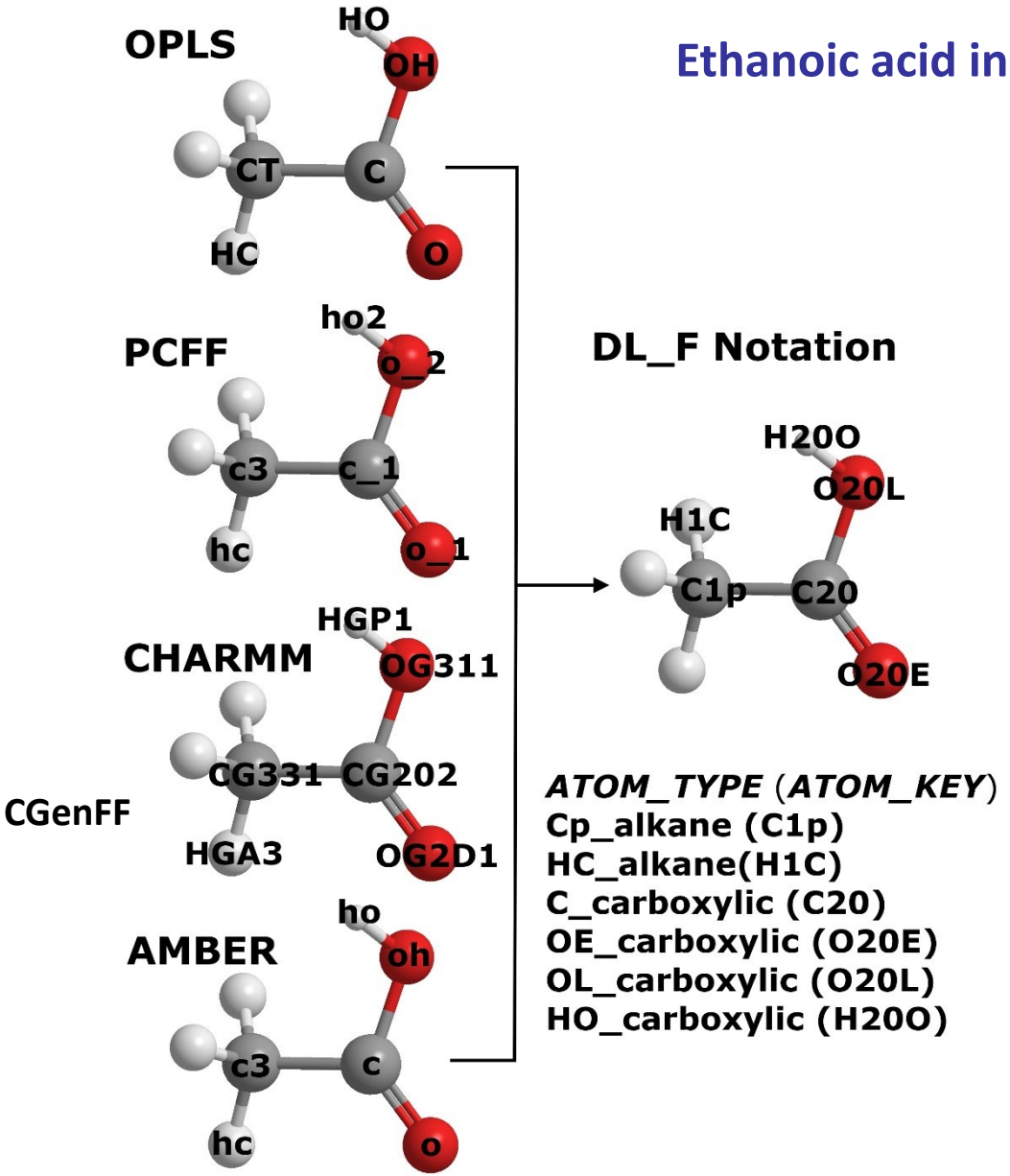
- Allows one single conversion scheme.

- Easy to identify, with sensible format of naming atoms that indicates precisely the chemical nature of every atom in the system.

- Allows full automatic determination of ATOM_TYPEs without the need of any pre-defined MOLECULE template.

- Reads configuration files in simple *xyz and mol2* formats.

Ethanoic acid in DL_F Notation



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

CGI **1** = alkane

CGI **20** = carboxylic

DLF_Notation File in *lib/* Directory

```
# carbon
1       alkane
2       alkene
3       alkyne
4       cyclopropyl
5       cyclobutyl
6       benzene
7       benzyl
8       phenol
9       biphenyl
10      allene
11      aromatic
12      cyclopropene
13      aliphatic
# oxygen
15      alcohol
16      ketone
17      aldehyde
18      acylhalide
19      ester
20      carboxylic
21      carboxylate
…
…
```

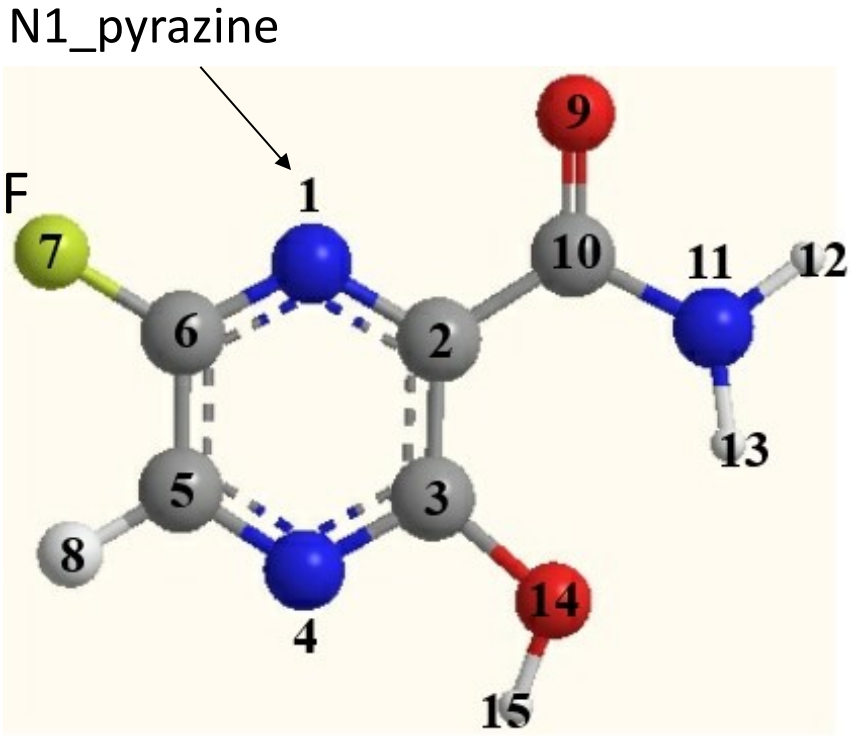
CGI

CG

```
# nitrogen
40      carbamate
41      carbamic_acid
42      urea
43      carbamoyl_chloride
44      amide
45      amine
46      aniline
47      ammonium
48      ketimine
49      aldimine
50      hydrazone
51      amidine
52      aldoxime
53      ketoxime
54      imide
55      cyanate
56      isocyanate
57      nitrate
58      nitrile
59      isonitrile
60      nitrite
…
…
```

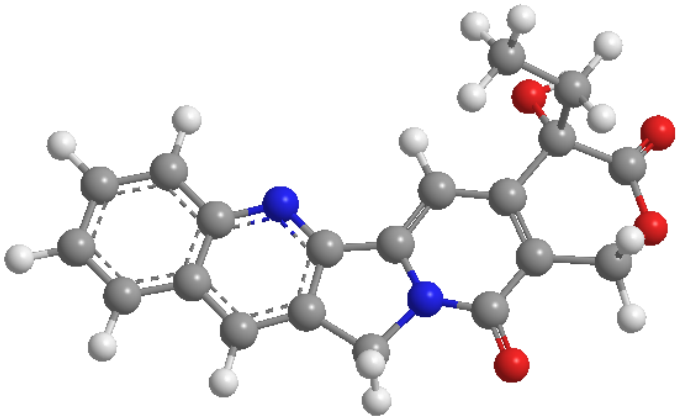
```
# sulphur
100     thiol
101     sulphide
102     disulphide
103     thione
104     thial
105     sulphoxide
106     sulphone
107     sulphate
108     thiocyanate
109     isothiocyanate
110     sulphinic_acid
111     sulphonic_acid
112     thionoester
113     sulphonamide
114     thioacyl_chloride
115     sulfinamide
116     sulphamide
117     sulphite_ester
118     thioamide
…
…
```

Favipiravir – antiviral drug



		PCFF	DL_F
2	- C2_pyrazine	cp	C502-2
3	- C3_pyrazine	cp	C502-3
4	- N4_pyrazine	np	N502-4
5	- C5_pyrazine	cp	C502-5
6	- C6_pyrazine	cp	C502-6
1	- N1_pyrazine	np	N502-1
8	- HC_aromatic	hc	H11C
7	- FC_aromatic	f	F11C
14	- O_phenol	oh	O8
15	- HO_phenol	ho	H8O
10	- C_amide	c_1	C43
9	- O_amide	o_1	O43
11	- N*p*_amide	n_2	N43p
12	- HN_amide	hn2	H43N
13	- HN_amide	hn2	H43N

Camptothecin – a quinoline alkaloid



PCFF

Camptotheca acuminata

```
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
...
```

1 - C6_quinoline
2 - C7_quinoline
3 - C8_quinoline
4 - C8a_quinoline
5 - C4a_quinoline
6 - C5_quinoline
7 - N1_quinoline
8 - C2_quinoline
9 - C3_quinoline
10 - C4_quinoline
11 - CE_alkene
12 - Nt_amide
13 - Cs_alkane
14 - CL_alkene
15 - CL_alkene
16 - CE_alkene
17 - C_amide
18 - Cq_alkane
19 - C_ester
20 - OL_ester
21 - Cs_alkane
22 - O_amide
23 - OE_ester
24 - O_alcohol
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
37 - HO_alcohol
38 - HC_alkane
39 - HC_alkane
40 - HC_alkane
41 - HC_alkane
42 - HC_alkane

Colistin (polypeptide antibiotic) effective against gram-negative bacteria

1 - Cs_alkane
2 - Cp_alkane
3 - Ct_alkane
4 - Cs_alkane
5 - Cs_alkane
6 - Cs_alkane
7 - C_amide
8 - Ns_amide
9 - O_amide
10 - Cp_alkane
11 - CA_amino_acid
12 - C_amide
13 - Cs_alkane
14 - Cs_alkane
15 - Np_amine
16 - Ns_amide
17 - O_amide
18 - CA_amino_acid
19 - CB_threonine
20 - CG_threonine
21 - C_amide
22 - Ns_amide
23 - O_threonine
24 - O_amide
25 - CA_amino_acid
26 - C_amide
27 - Cs_alkane
28 - Cs_alkane
29 - Np_amine
30 - Ns_amide
31 - O_amide
32 - CA_amino_acid
33 - C_amide
34 - Ns_amide
35 - O_amide
36 - Cs_alkane
37 - Cs_alkane
38 - Ns_amide
39 - C_amide
40 - CA_amino_acid
41 - O_amide
42 - CA_amino_acid
43 - C_amide
44 - Cs_alkane
45 - Cs_alkane
46 - Np_amine
47 - Ns_amide
48 - O_amide
49 - CA_amino_acid
50 - CB_leucine
51 - CG_leucine
52 - CD1_leucine
53 - CD2_leucine
54 - C_amide
55 - Ns_amide
56 - O_amide
57 - CA_amino_acid
58 - CB_leucine
59 - CG_leucine
60 - CD1_leucine
61 - CD2_leucine
62 - C_amide
63 - Ns_amide
64 - O_amide
65 - CA_amino_acid
66 - Cs_alkane
67 - Cs_alkane
68 - Np_amine
69 - C_amide
70 - Ns_amide
71 - O_amide
72 - CA_amino_acid
73 - Ns_amide
74 - CB_threonine
75 - CG_threonine
76 - O_threonine
77 - C_amide
78 - O_amide
79 - Cs_alkane
80 - Cs_alkane
81 - Np_amine
82 - HC_alkane
83 - HC_alkane
84 - HC_alkane
85 - HC_alkane
86 - HC_alkane
87 - HC_alkane
88 - HC_alkane
89 - HC_alkane
90 - HC_alkane
91 - HC_alkane
92 - HC_alkane
93 - HC_alkane
94 - HN_amide
95 - HC_alkane
96 - HC_alkane
97 - HC_alkane
98 – HCA_amino_acid
99 - HC_alkane
100 - HC_alkane
…
…

(DL_FIELD version 4.1)

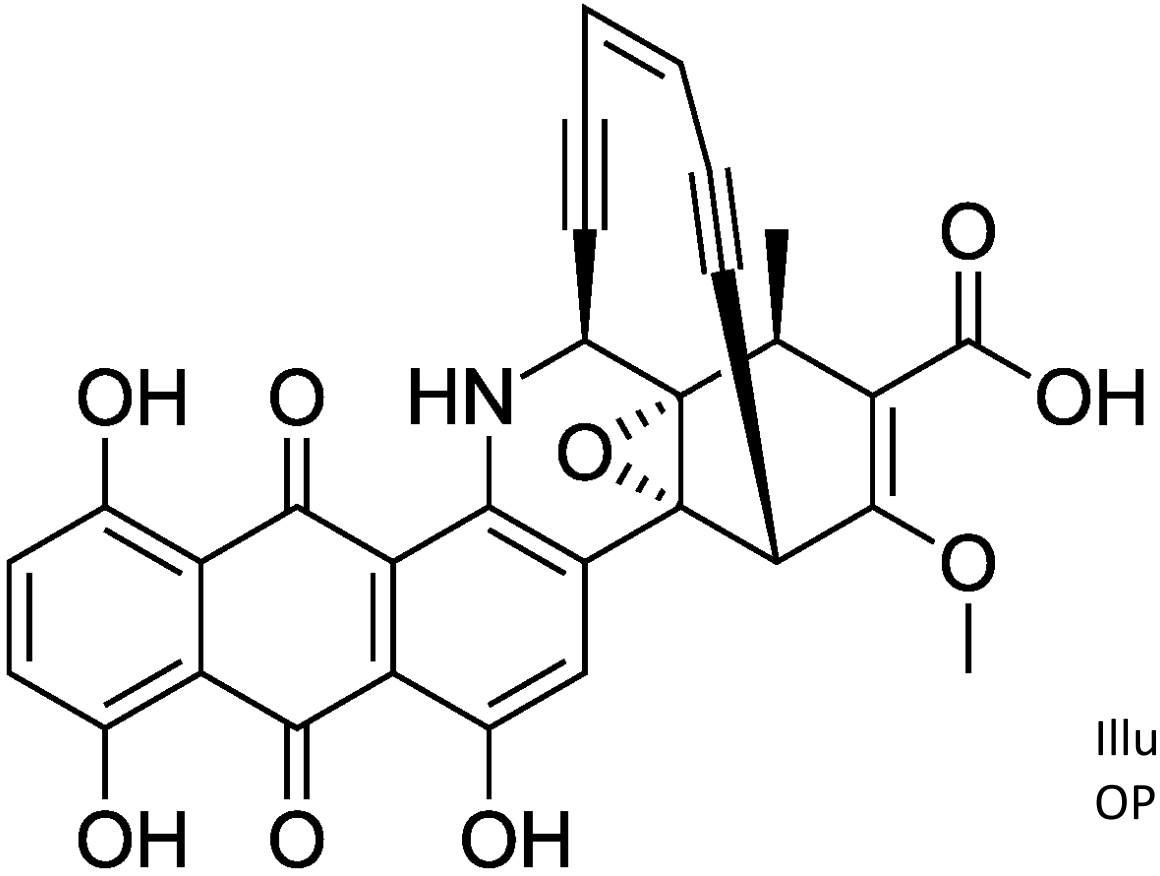
Science & Technology Facilities Council

UK Research and Innovation

3.xyz/3xyz.control

Demo: Dynemicin A

An antitumor antibiotic derived from microbial fermentation



Illustrate DL_F Notation
OPLS, PCFF, CVFF

Multiple Potential in DL_FIELD

Implementation of more than one type of FF schemes in a model: be it two or more organic FF schemes or inorganic FF schemes, or even the mixtures of the organic/inorganic FF schemes.

Auto-mixing different FF schemes possible.
Otherwise, leave blank. (Can use *VDW_FIX* directive)

Example PDB file contains A number of methanol molecules

```
REMARK  POTENTIAL AMBER
HETATM    1  C   MEOH    1      0.185   0.739  -0.000              GRP1 C
HETATM    2  H   MEOH    1     -0.770   1.307   0.000              GRP1 H
HETATM    3  H   MEOH    1      0.721   0.934  -0.953              GRP1 H
HETATM    4  H   MEOH    1      0.721   0.934   0.953              GRP1 H
HETATM    5  O   MEOH    1     -0.158  -0.634  -0.000              GRP1 O
HETATM    6  H   MEOH    1      0.666  -1.128  -0.000              GRP1 H
REMARK POTENTIAL OPLS2005
HETATM    8  H   MEOH    1     -0.770   6.307   0.000              GRP2 H
HETATM   10  H   MEOH    1      0.721   5.934   0.953              GRP2 H
HETATM    7  C   MEOH    1      0.185   5.739  -0.000              GRP2 C
HETATM   11  O   MEOH    1     -0.158   4.366  -0.000              GRP2 O
...
...
HETATM    7  C   MEOH    2      0.185   9.739  -0.000              GRP2 C
HETATM   11  O   MEOH    2     -0.158   8.366  -0.000              GRP2 O
HETATM   12  H   MEOH    2      0.666   7.872  -0.000              GRP2 H
HETATM    9  H   MEOH    2      0.721   9.934  -0.953              GRP2 H
HETATM    1  C   MEOH    3      0.185  -3.261  -0.000              GRP3 C
HETATM    2  H   MEOH    3     -0.770  -2.693   0.000              GRP3 H
HETATM    3  H   MEOH    3      0.721  -3.066  -0.953              GRP3 H
HETATM    4  H   MEOH    3      0.721  -3.066   0.953              GRP3 H
HETATM    5  O   MEOH    3     -0.158  -4.634  -0.000              GRP3 O
HETATM    6  H   MEOH    3      0.666  -5.128  -0.000              GRP3 H
REMARK  POTENTIAL CHARMM22_prot
HETATM    8  H   MEOH    1      3.230   6.307   0.000              GRP4 H
HETATM   10  H   MEOH    1      4.721   5.934   0.953              GRP4 H
HETATM    7  C   MEOH    1      4.185   5.739  -0.000              GRP4 C
HETATM   11  O   MEOH    1      3.842   4.366  -0.000              GRP4 O
HETATM   12  H   MEOH    1      4.666   3.872  -0.000              GRP4 H
HETATM    9  H   MEOH    1      4.721   5.934  -0.953              GRP4 H
END
```

Define FFs – indicate the extent of each FF.

This is just an example. Do not try this.

Bio-inorganic model in PDB format

Define FFs – indicate the extent of each FF.

PDB format

```
REMARK POTENTIAL charmm22_prot
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
…
….
```

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

```
 451
CRYST1   83.908   20.7302253   17.9859383759  90.00  90.00  90.00 (P 1)
# POTENTIAL CVFF  MOLECULAR_GROUP ORG1
C   10.000000  1.390000  0.000000
C   10.000000  0.695000  1.204000
C   10.000000  -0.695000  1.204000
C   10.000000  -1.391000  0.000000
C   10.000000  -0.695000  -1.204000
C   10.000000  0.695000  -1.204000
H   10.000000  2.450000  0.000000
H   10.000000  1.225000  2.123000
H   10.000000  -1.226000  2.122000
H   10.000000  -2.451000  0.000000
H   10.010000  -1.226000  -2.122000
H   10.000000  1.225000  -2.122000
# POTENTIAL inorganic_clay   MOLECULAR_GROUP CLY   MOLECULE_KEY
CLYF
st  -2.675000  -6.024000  4.703000
st  -2.700000  -6.039000  -4.313000
st  -2.776000  -3.444000  -8.761000
st  -2.686000  -0.853000  4.722000
st  -2.697000  -0.830000  -4.318000
st  -2.786000  -8.616000  -8.741000
st  -2.690000  -3.463000  6.268000
st  -2.730000  -8.634000  6.283000
st  -2.722000  -8.677000  0.254000
st  -2.653000  -3.446000  -2.766000
st  -2.709000  -8.694000  -2.754000
st  -2.687000  -3.445000  0.268000
```

Element symbols

Can be element symbols or ATOM_KEYs

Must define this for inorganic FF.

Van-der-Waal mixing schemes

(between two atoms from two different force fields)

Mixing rule no	Scheme name	Energy, ε_{ij}	Steric, σ_{ij}
1	Standard geometric	$\sqrt{\varepsilon_{ii}\varepsilon_{jj}}$	$\sqrt{\sigma_{ii}\sigma_{jj}}$
2	Standard arithmatic	$\frac{\varepsilon_{ii}+\varepsilon_{jj}}{2}$	$\frac{\sigma_{ii}+\sigma_{jj}}{2}$
3	Fender-Halsey	$\frac{2\epsilon_{ii}\epsilon_{jj}}{\epsilon_{ii}+\epsilon_{jj}}$	$\frac{\sigma_{ii}+\sigma_{jj}}{2}$
4	Halgren HHG	$4\frac{\varepsilon_{ii}\varepsilon_{jj}}{(\sqrt{\varepsilon_{ii}}+\sqrt{\varepsilon_{jj}})^2}$	$\frac{\sigma_{ii}^3+\sigma_{jj}^3}{\sigma_{ii}^2+\sigma_{jj}^2}$
5	Waldman-Hagler	$2\sqrt{\varepsilon_{ii}\varepsilon_{jj}}\frac{(\sigma_{ii}\sigma_{jj})^3}{\sigma_{ii}^6+\sigma_{jj}^6}$	$\left(\frac{\sigma_{ii}^6+\sigma_{jj}^6}{2}\right)^{\frac{1}{6}}$

Lorentz-Berthelot – specify 1 for energy, 2 for steric

Hogervorst (good hope) – specify 1 for energy, 1 for steric

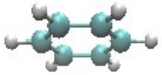
1.pdb/1pdb.control

Change to opls2005
Change mixing rules

Demo: multiple potential, organic-inorganic model

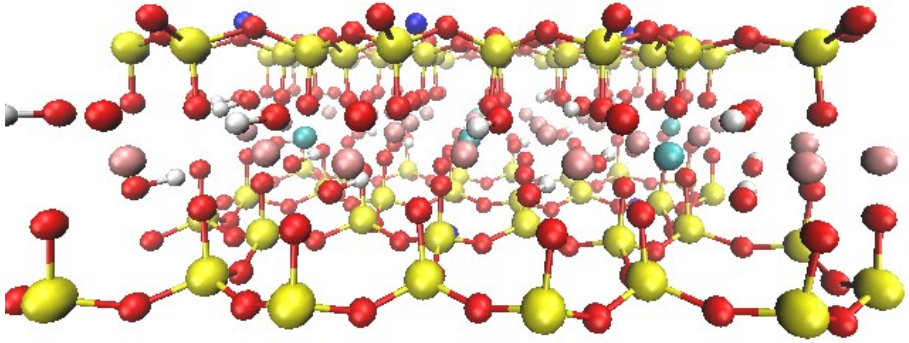
FF schemes

CHARMM22



benzene

CLAYFF



Montemorillonite (phyllosilicate clay mineral)
Two tetrahedral Si sandwiched a central octahedral sheet of alumina.

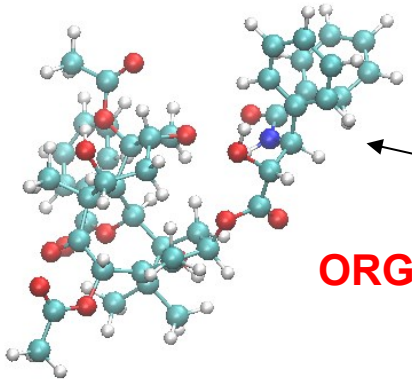
2.xyz/2xyz.control

(see 2.xyz)
(see dlf_notation.output)

Demo 3: multiple potential, organic-inorganic model

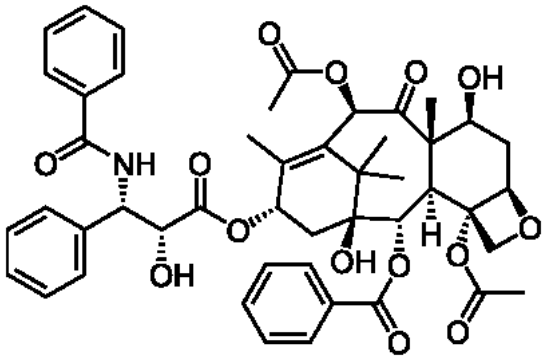
FF schemes

OPLS

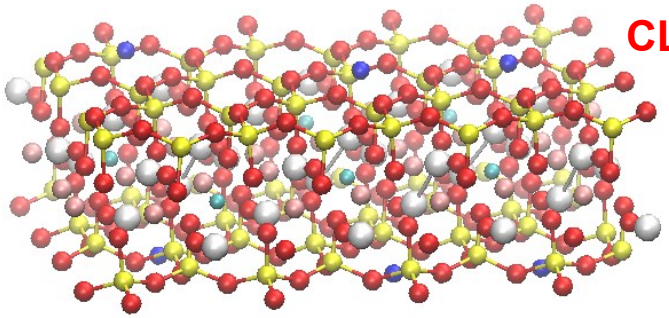


ORG2

Paclitaxel (a chemotherapy drug)



CLAYFF



CLY

CVFF

ORG1

benzene

FREEZE ORG1 cp
FREEZE ORG2 CT
TETHER CLY st
CONSTRAIN H-bond on ORG1, ORG2

Using DL_FIELD

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

Uncompress program:
gunzip *dl_field_4.5.tar.gz*

Compile program:
cd source
make clean
make

Run program:
./dl_field

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
/Examples – Some PDB and *xyz* structures
/solvent –solvent templates
/tutorial – tutorial directory
/utility – Contain some useful scripts.
/control_files – control files that run the example structures

DL_FIELD Tutorial

***dl_f_path* file**

```
# Directory paths for DL_FIELD version 4.4 onwards.
# C W Yong, October 2018
# 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/

# DL_FIELD control files
control = dl_field.control
# control = control_files/example1.control
# control = tutorial/tutorial_1.control
```

run examples

run tutorial

Tutorial: see *tutorial/dl_f_tutorial.pdf*

Example structures: See *dl_f_manual.pdf*, Chapter 13

DL_FIELD registrations

The program is supplied to individuals under an academic licence, which is free to academics pursuing scientific research of a non-commercial nature. Daresbury Laboratory is the sole centre for distribution of the package.

For more information, please visit the web site:
http://www.ccp5.ac.uk/DL_FIELD

Comments, suggestions: chin.yong@stfc.ac.uk

Science & Technology Facilities Council

UK Research and Innovation

DL_Software
Enabling Discovery

DL_ANALYSER
A general analysis tool for DL_POLY

Dr C W Yong
Scientific Computing Department,
STFC,
Daresbury Laboratory, Sci-Tech Daresbury,
Warrington WA4 4AD

(DL_Software training workshop, Chile, May 2019)

EPSRC
Engineering and Physical Sciences Research Council

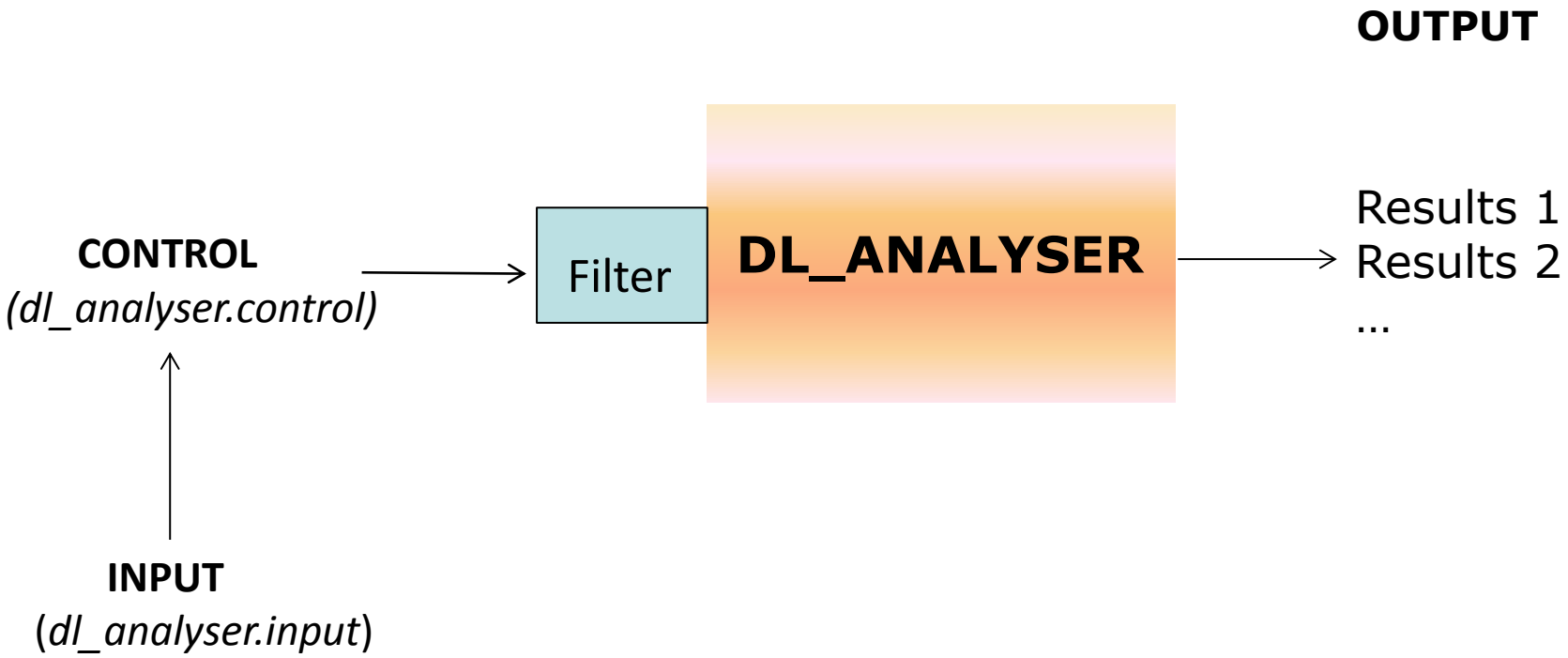
CoSeC
Computational Science Centre for Research Communities

CCP5

ada lovelace centre

ADDOPT

DL_ANALYSER Scheme



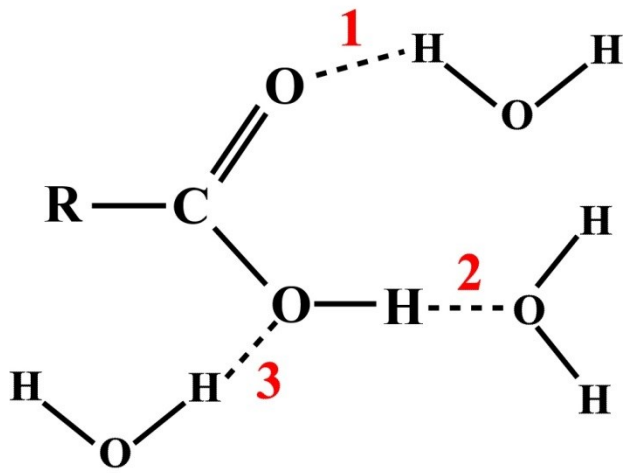
DANAI

DL_ANALYSER Notation for Atomic Interaction

A standard expression system to annotate specific atomic interactions.

Useful for almost every aspect of atomistic simulations including solute-solvent interactions.

(C. W. Yong & I. T. Todorov, *Molecules* (2018), **23**, 36)



Typical pictorial representation of H-bond (HB) between carboxylic acid and water molecules.

How many types of HB interactions between two carboxylic (-COOH) groups?

DANAI – standard annotation system to describe localised atomistic interactions.

- Detailed microscopic descriptions of atomic interactions between 2 or more functional groups.

- Universal expressions that contain the actual chemical details and hides (and bypass) complex data structures that are dependent on FF models.

- Consistent notation for any kind of molecular systems – crystalline, polymeric, condensed-phase.

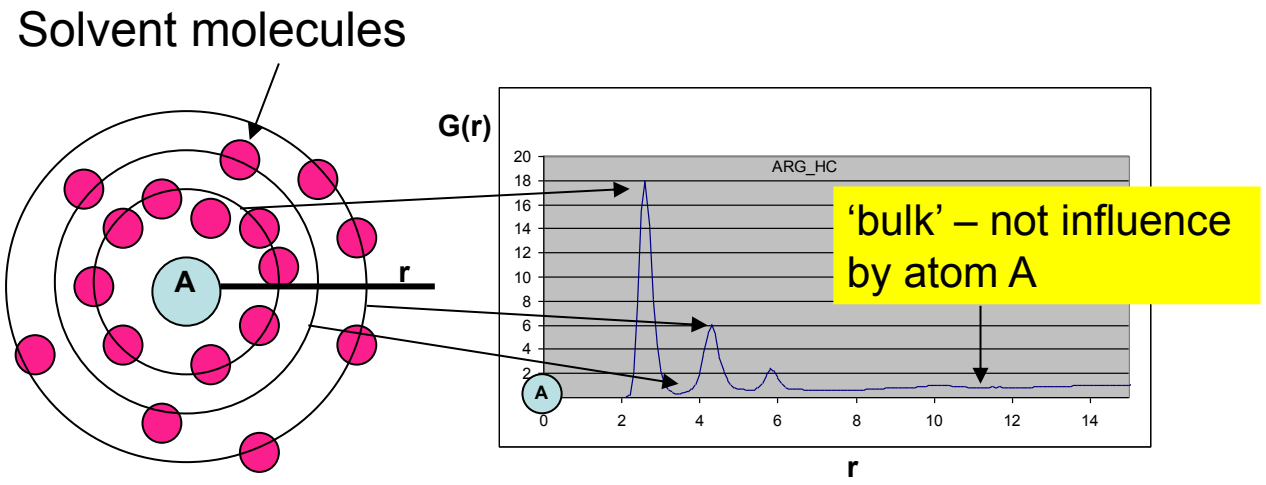
- Easy to construct and interpret – by human and computation.

- Information can be catalogued and subsequently retrieved for data analysis (cheminformatics).

- MD trajectories expressed in the universal notation. Run once and archived. Enable future data analysis.

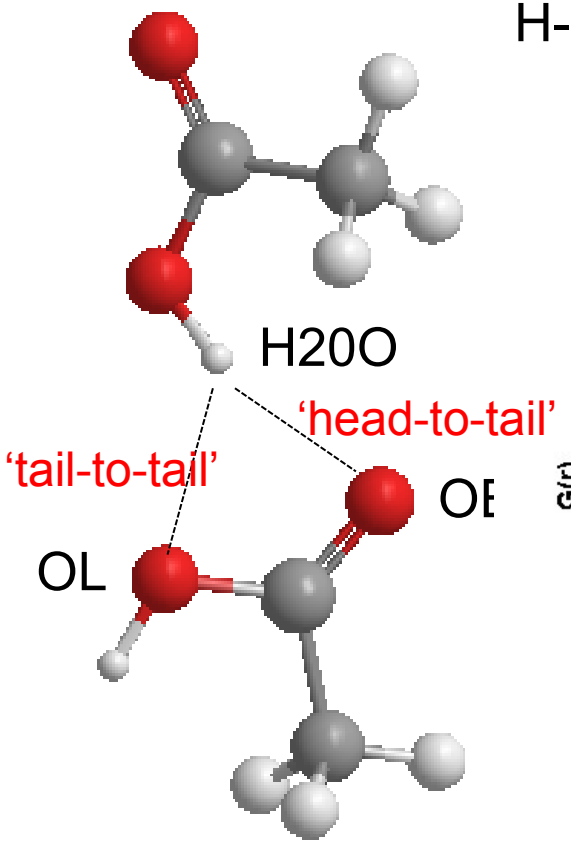
G(r) – radial distribution function. Common analysis method for looking at the molecular structures in solutions.

Radial distribution function (RDF)

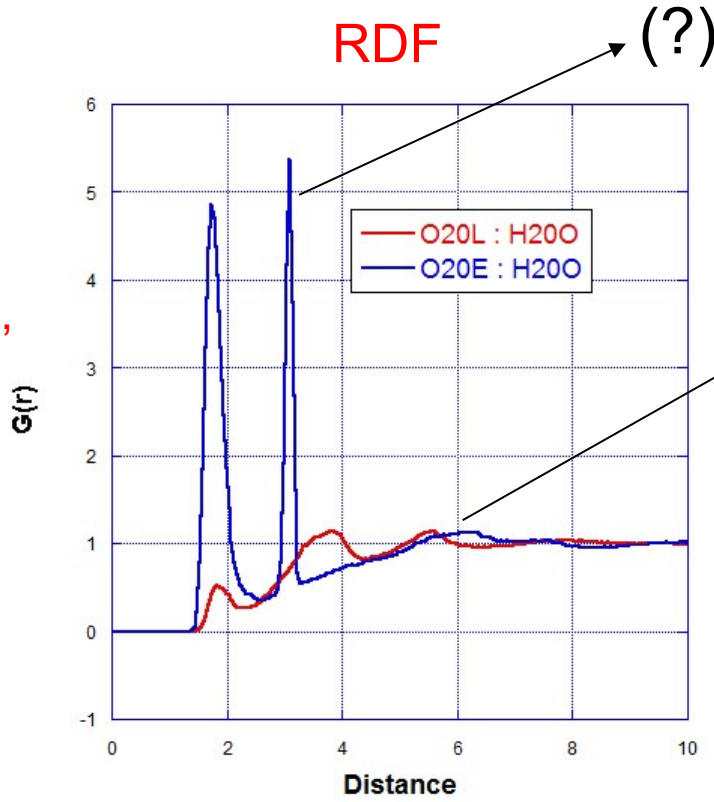


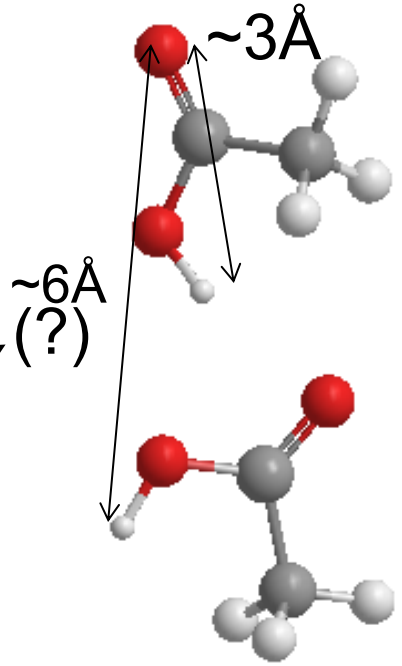
Good for 'general feel' how solvent molecules packed around the solute.

These analysis gives solvent-solute interactions in a broader sense.



H-bond interactions in ethanoic acid

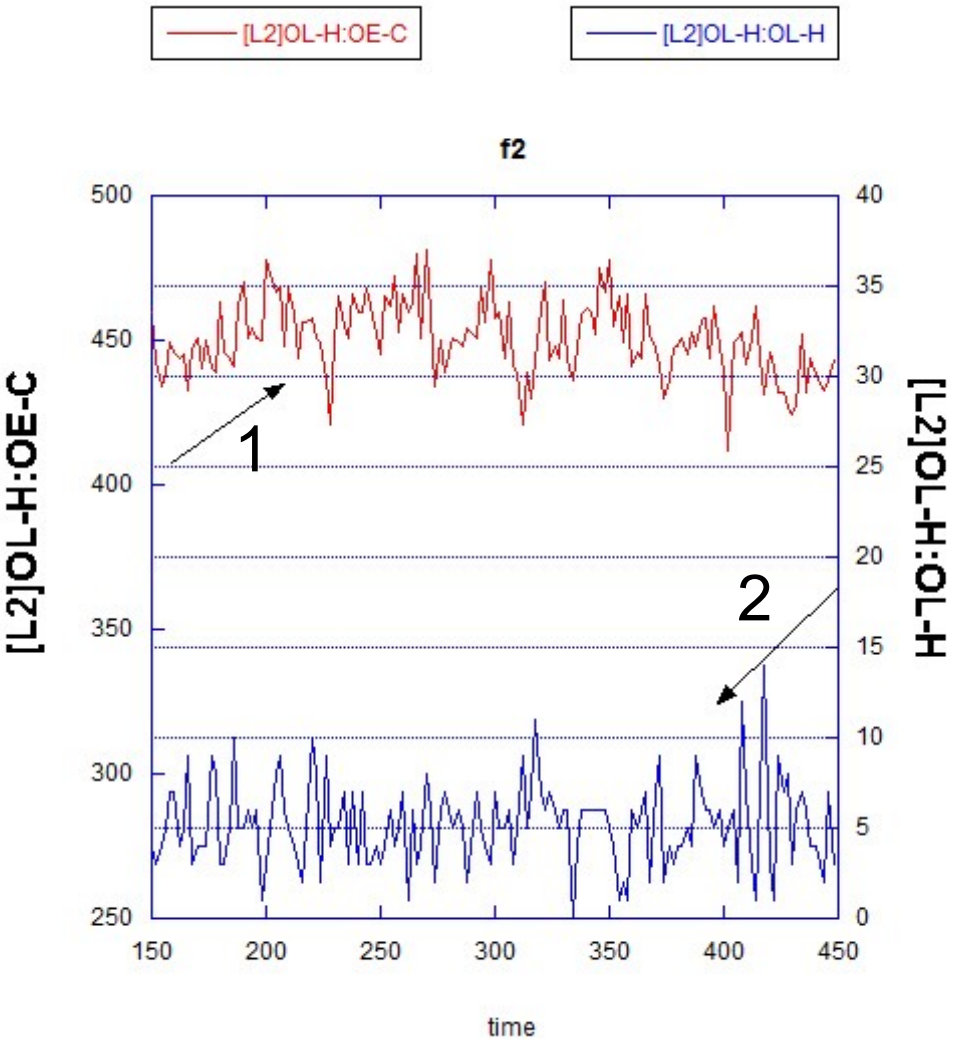


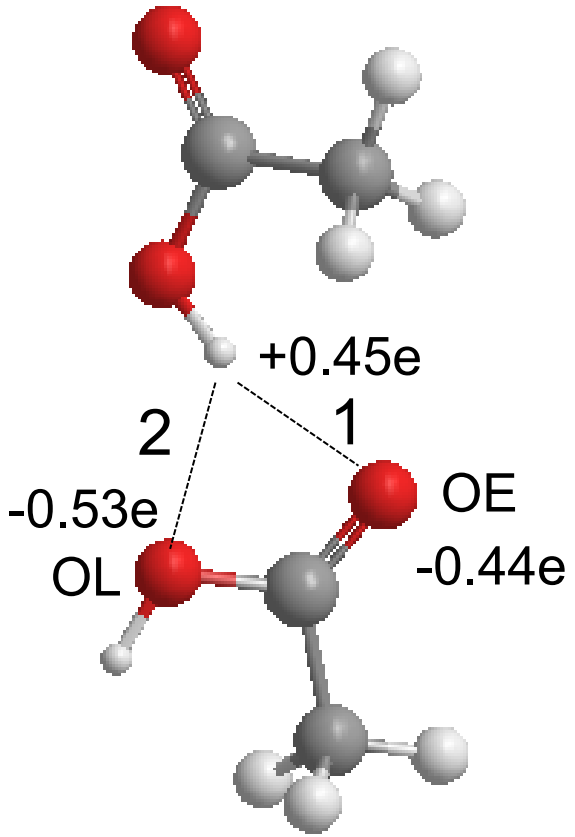


Movie: 3cooh.xyz

Molecular simulations Of pure ethanoic acid

Use of DL_ANALYSER

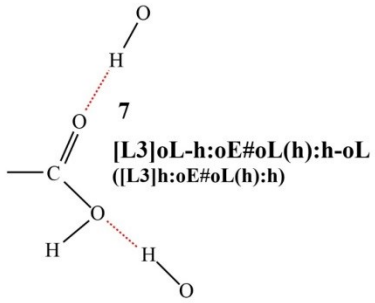




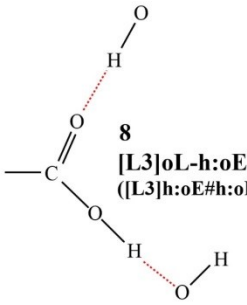
Science & Technology Facilities Council

UK Research and Innovation

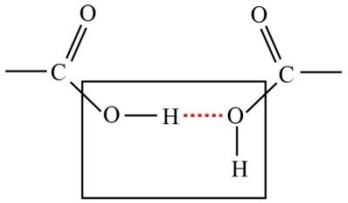
Carboxylic-carboxylic interactions (HB_20_20)



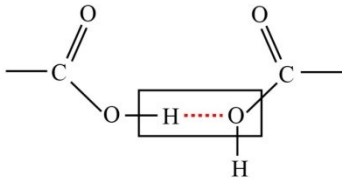
7
[L3]oL-h:oE#oL(h):h-oL
([L3]h:oE#oL(h):h)



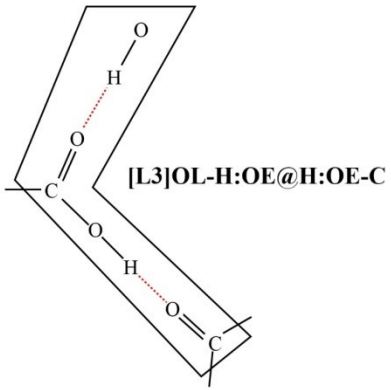
8
[L3]oL-h:oE#h:oL-h
([L3]h:oE#h:oL)



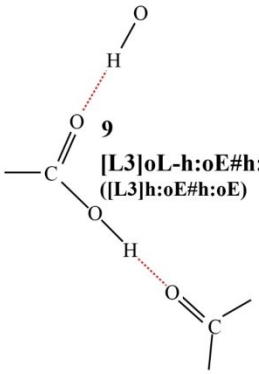
1 [L2]OL-H:OL-H



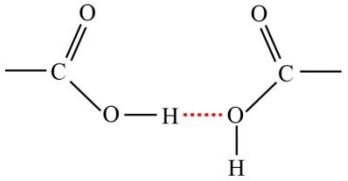
2 [L2]oL-H:OL-h
([L2]H:OL)



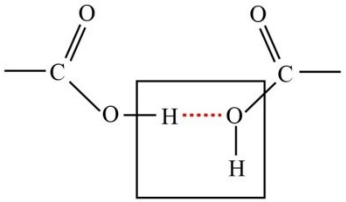
[L3]OL-H:OE@H:OE-C



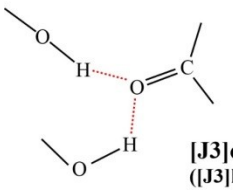
9
[L3]oL-h:oE#h:oE-c
([L3]h:oE#h:oE)



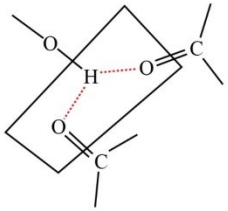
3 [L2]oL-h:oL-h
([L2]h:oL)



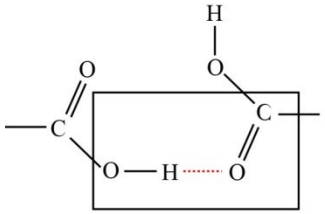
[L2]oL-H:OL-H
([L2]H:OL-H)



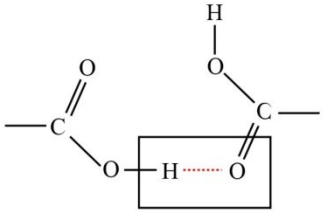
10
[J3]oL-h:oE(c):h-oL
([J3]h:oE(c):h)



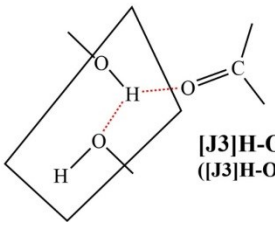
[J3]c-OE:H(oL):OE-c
([J3]OE:H(oL):OE)



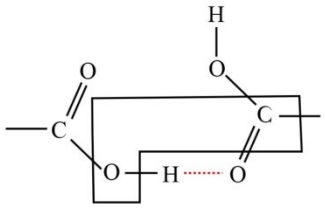
4 [L2]OL-H:OE-C



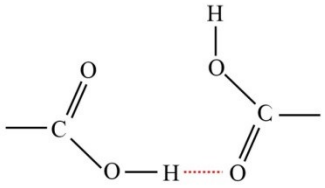
5 [L2]oL-H:OE-c
([L2]H:OE)



[J3]H-OL:H(OL):oE-c
([J3]H-OL:H(OL):oE)



[L2]OL-h:oE-C



6 [L2]oL-h:oE-c
([L2]h:oE)

UK Research and Innovation

Molecular simulations Of pure ethanoic acid

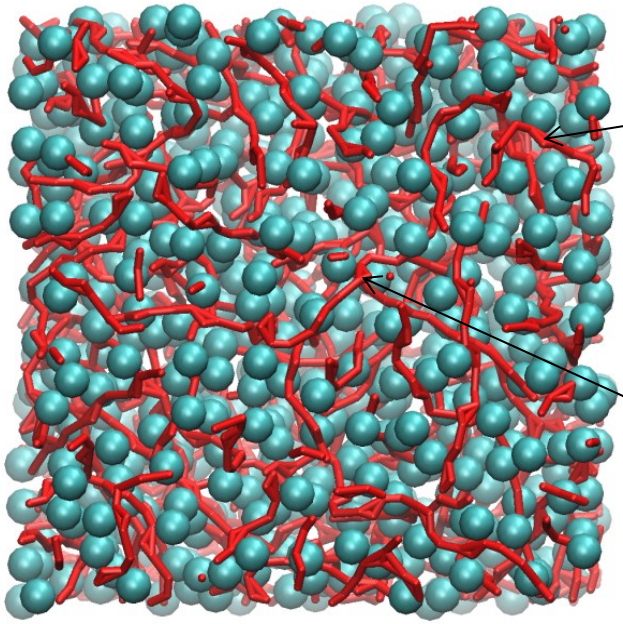
DANAI expression, i	number of interactions, m_i
1. [L2]OL-H:OL-H	1.69 ± 1.27
2. [L2]oL-H:OL-h	79.34 ± 7.58
3. [L2]oL-h:oL-h	113.50 ± 8.75
4. [L2]OL-H:OE-C	418.90 ± 14.65
5. [L2]oL-H:OE-c	504.61 ± 10.76
6. [L2]oL-h:oE-c	581.28 ± 8.50
7. [L3]oL-h:oE#oL(h):h-oL	59.64 ± 6.17
8. [L3]oL-h:oE#h:oL-h	92.95 ±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

Correlation analysis (Pearson)

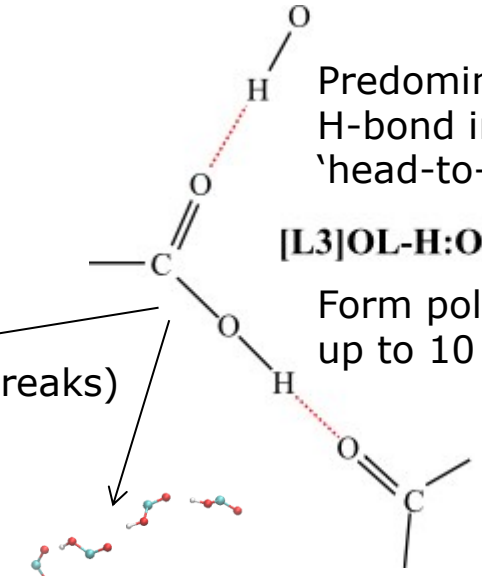
	1	2	3	4	5	6	7	8	9	10
1	1.000	0.056	-0.001	-0.004	-0.080	-0.182	0.005	-0.009	-0.163	-0.063
2		1.000	0.764	-0.610	-0.444	-0.811	0.627	0.644	-0.794	-0.253
3			1.000	-0.834	-0.653	-0.628	0.617	0.853	-0.651	-0.067
4				1.000	0.906	0.526	-0.445	-0.696	0.676	-0.312
5					1.000	0.405	-0.286	-0.506	0.616	-0.546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

C. W. Yong & I. T. Todorov, *Molecules* (2018), **23**, 36

Ethanoic acid liquid interactions.



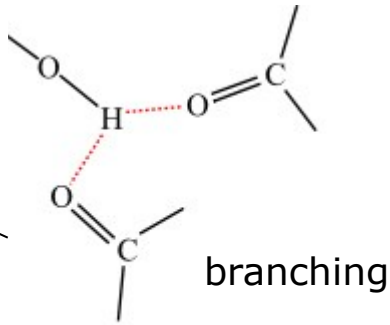
(Red streaks)



Predominant form of H-bond interactions: 'head-to-tail'

[L3]OL-H:OE@H:OE-C

Form polymeric chains, up to 10 carboxylic groups



branching

[J3]C-OE:H(OL):OE-C

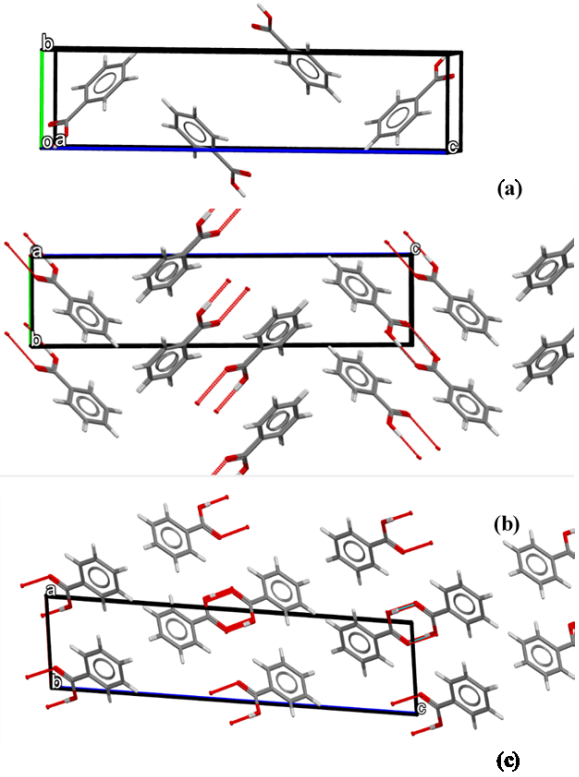
Methyl groups (cyan spheres) hydrophobic interactions, forming a sponge-like super-structure that permeates across the whole system.

Can also carry out cross-correlation analysis between HB and HP interactions.

Benzoic acid case

https://doi.org/10.1080/08927022.2018.1560441

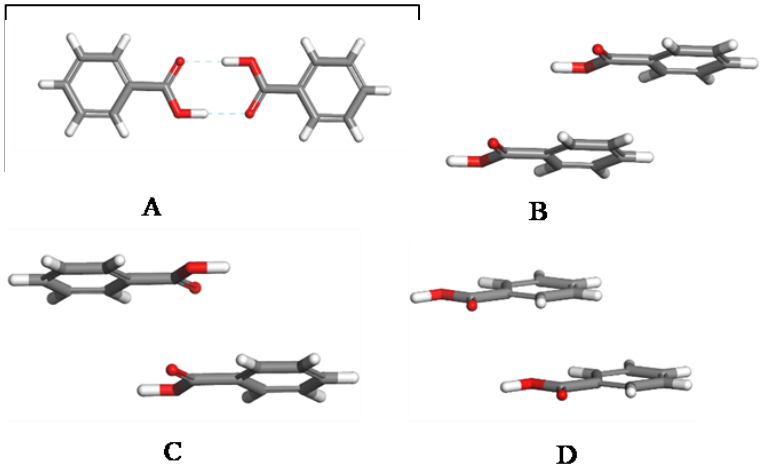
- Benzoic Acid crystallises in a single polymorphic form, containing H:O20E H-bonding dimers
- Solution chemistry experimental data indicates that H-bonding occurs in solution prior to nucleation



It is hypothesised that the pre-aggregation of the H:O20E H-bonding dimers in solution drives the crystallisation of the singular polymorphic form of benzoic acid

Crystal structure analysis

The strongest intermolecular group interactions in benzoic acid solid-state structure were identified

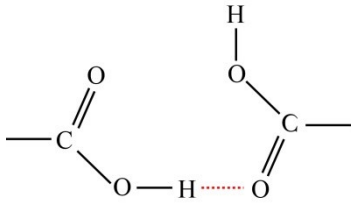


Synthon	Multiplicity	Intermolecular Energy	Lattice Energy Contribution (%)	Interaction Type
A	1	-5.3	26.8	H-bonding dimers
B	2	-2.0	19.8	Dispersive head-head pi-pi stack
C	1	-1.5	7.6	Dispersive head-tail polar stack
D	2	-1.4	14.2	Dispersive offset head-head stack

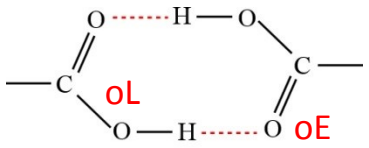
- Strongest intermolecular interaction in the crystal structure found to be O:H bonding dimers
- Also some strong pi-pi stacking interactions found in the crystal structure

MD Simulation of benzoic acid in hexane

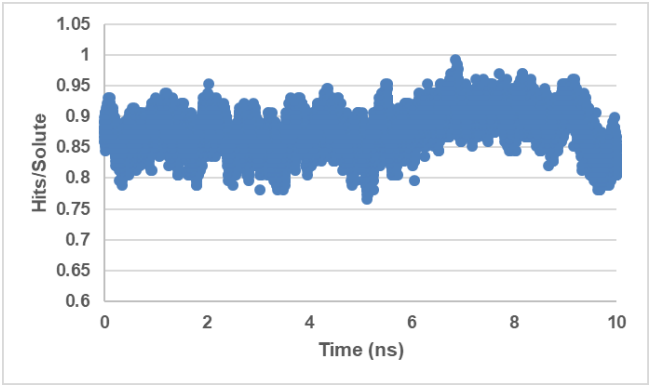
DL_FIELD ⟶ DL_POLY ⟶ DL_ANALYSER



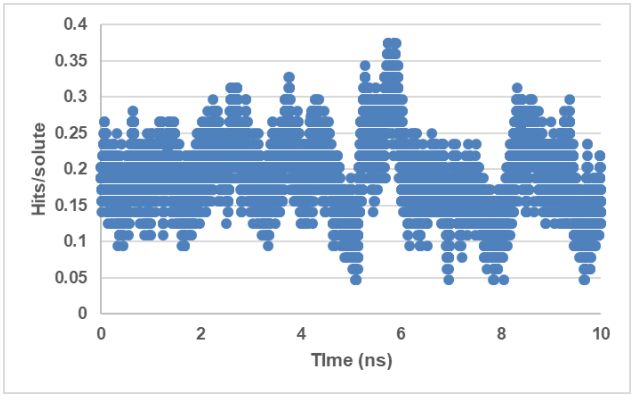
HB_20_20



[L2]oL-h:oE-c



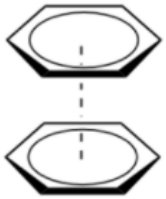
[R2]c-oL-h:oE#h:oE-c



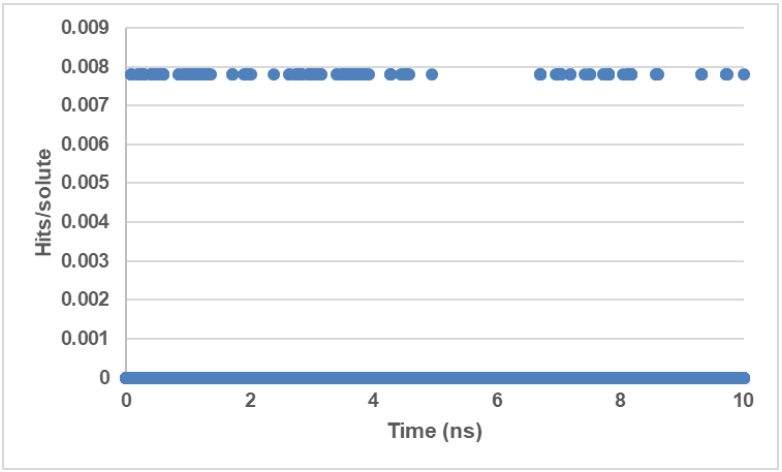
- The lower case letters in the DANAI expression indicate that these atoms can form other interactions with other atoms
- It was found that the vast majority of carboxyl groups were forming singular h:oE interactions
- The carboxylic acid dimer interactions were also found to be present in approximately 20% of the molecules

π-π stacking interactions were also examined from the benzoic acid/hexane systems

PS_6_6



[L2]<c6>:<c6>

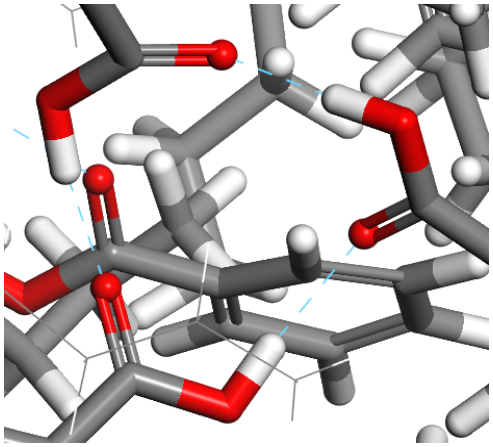


Numbers of pi-pi stacking interactions much lower than the h:oE or h:oL interactions

Though some stacking interactions appear in the trajectory files, they are much less frequent than the h:oE interactions and are therefore thought to influence the crystallisation much less.

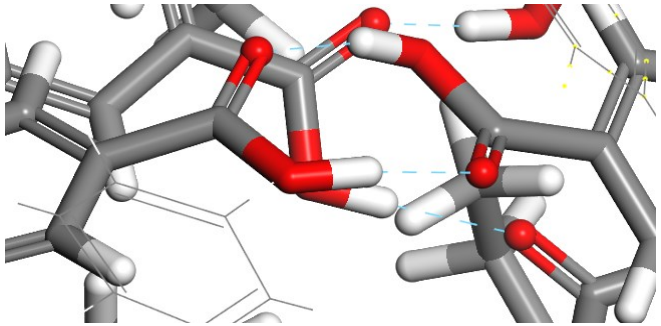
Structural examinations

[R3]c-oL-h:oE#h:oE#h:oE-c

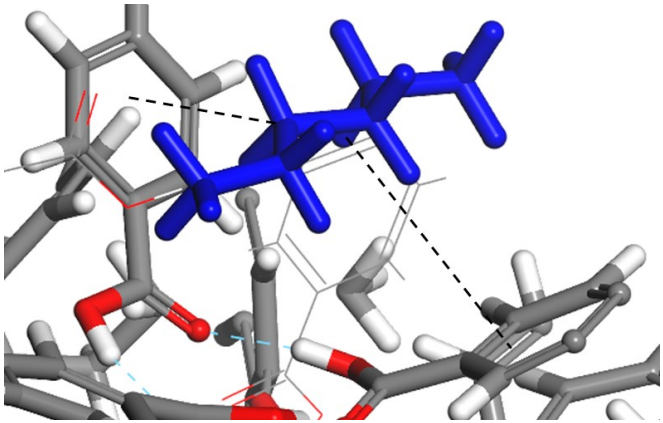


Three membered OH…O H-bonding ring

[R2]c-oL-h:oE-c-oL-h:oE-c



OH…O H-bonding ‘classic’ dimer



Hexane molecule sitting between ring structures

Hexane molecules poorly solvating the COOH group results in the aggregation of these groups, however they can form stronger dispersive interactions with the less polar groups

DL_ANALYSER registrations

The program is supplied to individuals under an academic licence, which is free to academics pursuing scientific research of a non-commercial nature. Daresbury Laboratory is the sole centre for distribution of the package.

For more information, please visit the web site:
http://www.ccp5.ac.uk/DL_ANALYSER

Comments, suggestions: chin.yong@stfc.ac.uk