



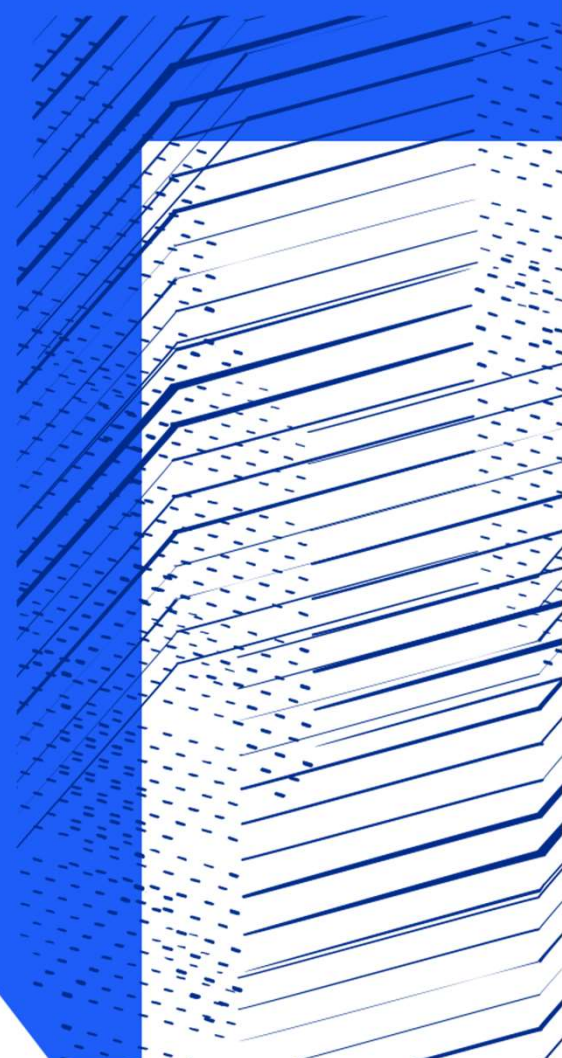
Science and
Technology
Facilities Council

ISIS Neutron and
Muon Source

Introduction to Disordered Materials Data Analysis Methods

Tom Headen (tom.headen@stfc.ac.uk)

DMUGM Cosener's House April 2024



Agenda

1 Comparing Data to simulation

g(r), S(Q), F(Q), isotopic substitution.....

2 Molecular simulation: key concepts

Interatomic potentials, Monte Carlo, Molecular Dynamics, Periodic Boundary Conditions.....

3 Potential Refinement

Linking g(r) and V(r), “EPSR method”

4 What next: Simulation analysis, Common pitfalls, Additional features



Science and
Technology
Facilities Council

ISIS Neutron and
Muon Source

Comparing data to simulations



Total Structure Factor: From raw data....

Theory can link measured scattering intensities to real-space simulation data – but is **idealised**. **So we must:**

- Remove multiple scattering and beam attenuation effects
- Remove self-scattering – Interference scattering only
- Remove inelastic scattering: Important (and difficult) for light H.
- Have a normalised intensity on an absolute scale (barns/atom/sr)
- For X-rays need to also take into account form factors of atoms

Gudrun: .min01 (neutrons) and .int01 (X-rays) files

Total Structure Factor: Link to simulation

$$F(Q) = \sum_{i,j} (2 - \delta_{ij}) c_i c_j b_i b_j S_{ij}(Q)$$

Total Structure Factor
(Experimental Observable)

$$c_i = \frac{n_i}{N} \quad b_i = \text{scattering length}$$

$$S_{ij}(Q) = \rho \int_0^{\infty} 4\pi r^2 g_{ij}(r) \frac{\sin Qr}{Qr} dr$$

Partial Structure Factor

(Accessed via Simulation)

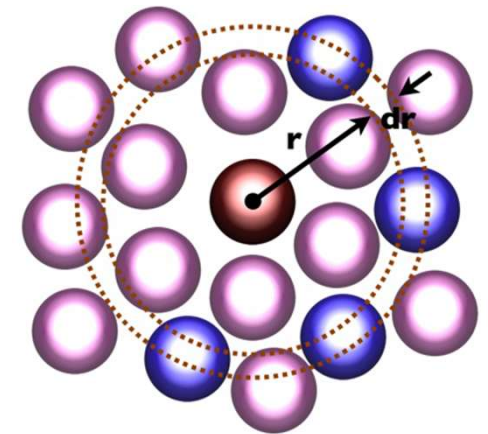
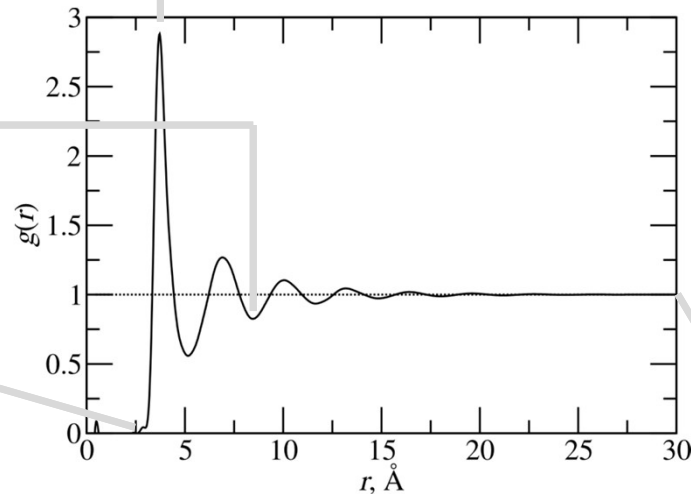
The Radial Distribution Function

- A Radial Distribution Function (RDF) describes the average density of atoms of a given type j at a distance r from an atom of type i , compared to the bulk
- ‘Sit’ on an atom of type i , and count the number of atoms of type j that fall within a spherical shell of width Δr , centred at r

Maxima ($g_{ij}(r) > 1$) reflect higher than average density of j , resulting from packing, favourable interactions etc.

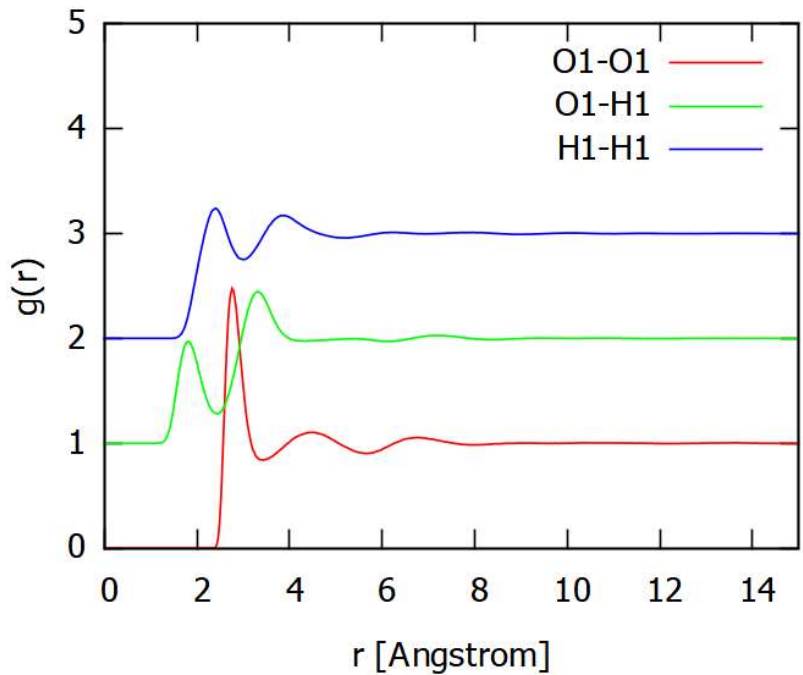
Minima ($g_{ij}(r) < 1$) reflect lower than average density of j , arising from exclusion, physical ordering etc.

$g_{ij}(r) = 0$ until some minimum r , reflecting physically excluded volume around atoms

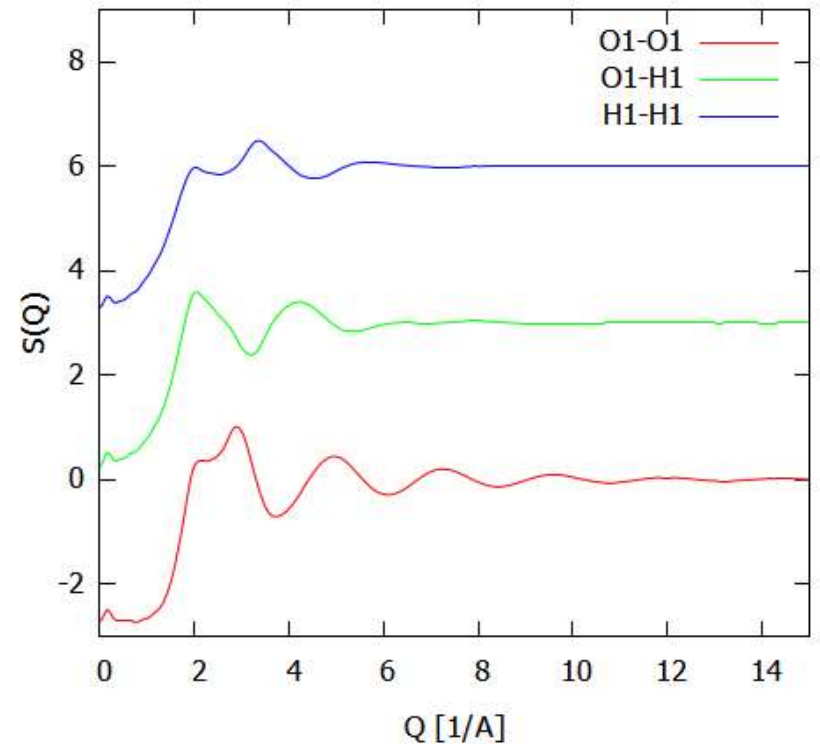
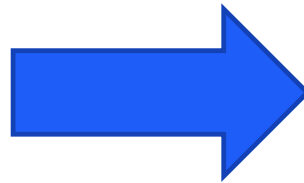


$g_{ij}(r) = 1$ corresponds to ‘expected density’ of j – in other words, nothing special

$g(r)$ to $S(Q)$ for water....



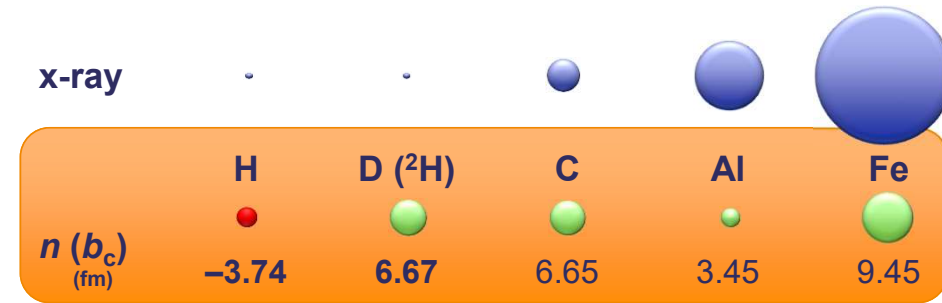
FT



$$S_{ij}(Q) = \rho \int_0^{\infty} 4\pi r^2 g_{ij}(r) \frac{\sin Qr}{Qr} dr$$

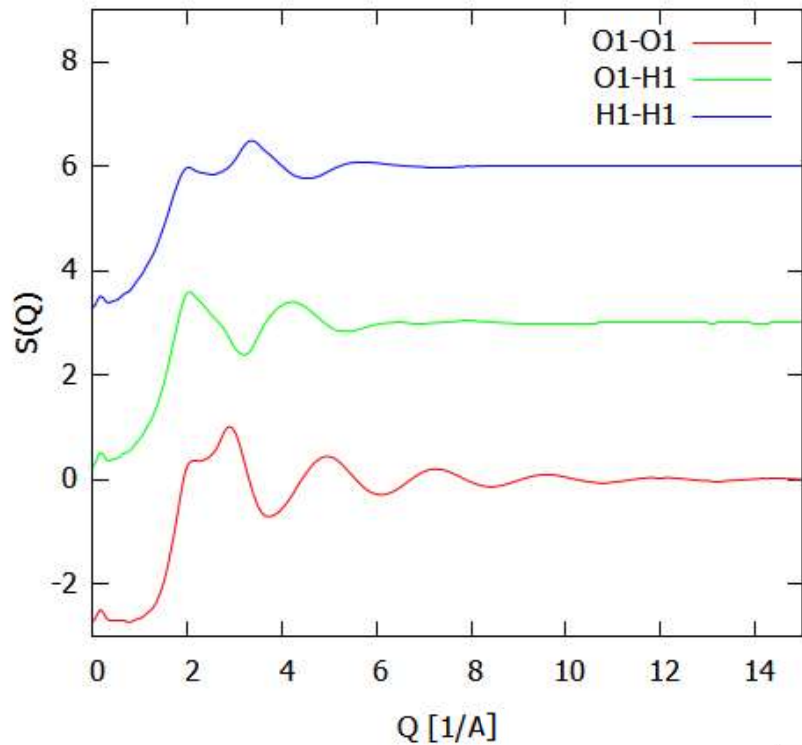
Isotopic Substitution

$$F(Q) = \sum_{i,j} (2 - \delta_{ij}) c_i c_j \underbrace{b_i b_j}_{\text{Scattering Lengths}} S_{ij}(Q)$$

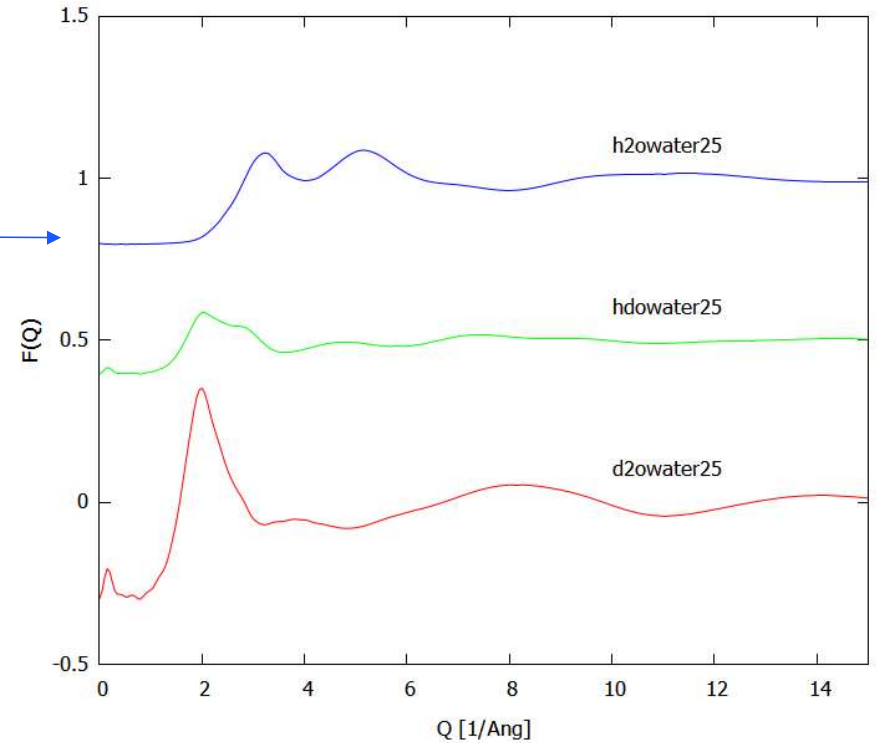


- Change weightings of partials in the measured $F(Q)$ by changing isotope
 - e.g. ^2H for H, ^{15}N for N, ^6Li for Li
- Key assumption: **structure is independent of isotopes used**
- Perform multiple measurements on the same system, with different isotopic substitutions

S(Q) to F(Q) for water....

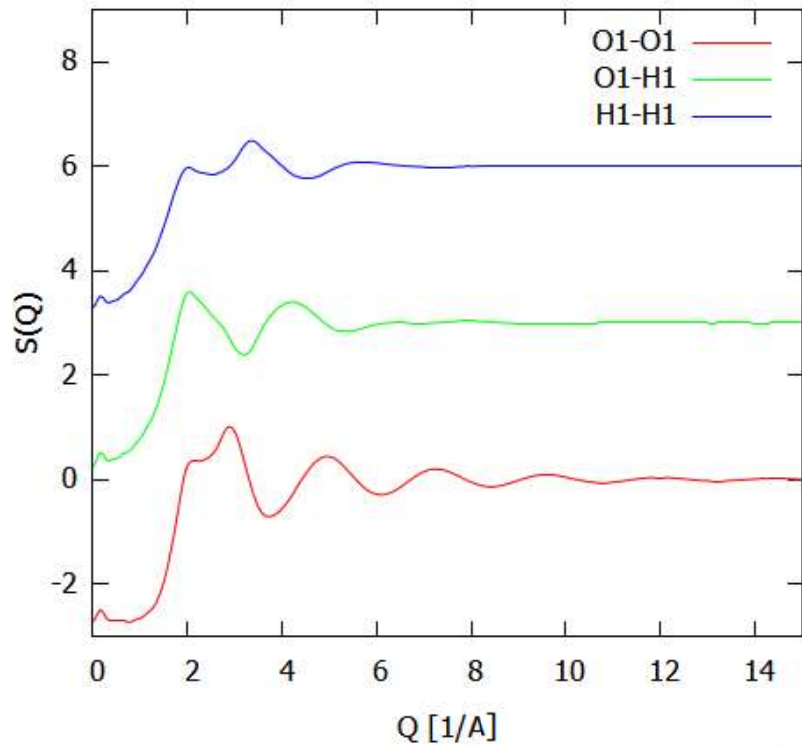


Weighted
sum

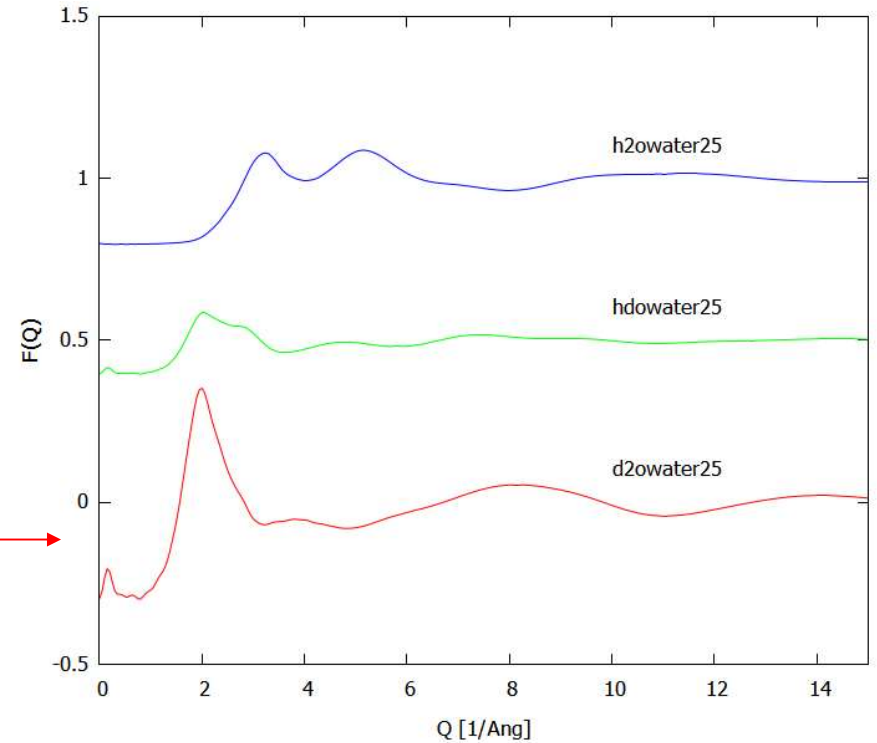


$$F(Q) = \sum_{i,j} (2 - \delta_{ij}) c_i c_j \underbrace{b_i b_j}_{b_H} S_{ij}(Q)$$

S(Q) to F(Q) for water....



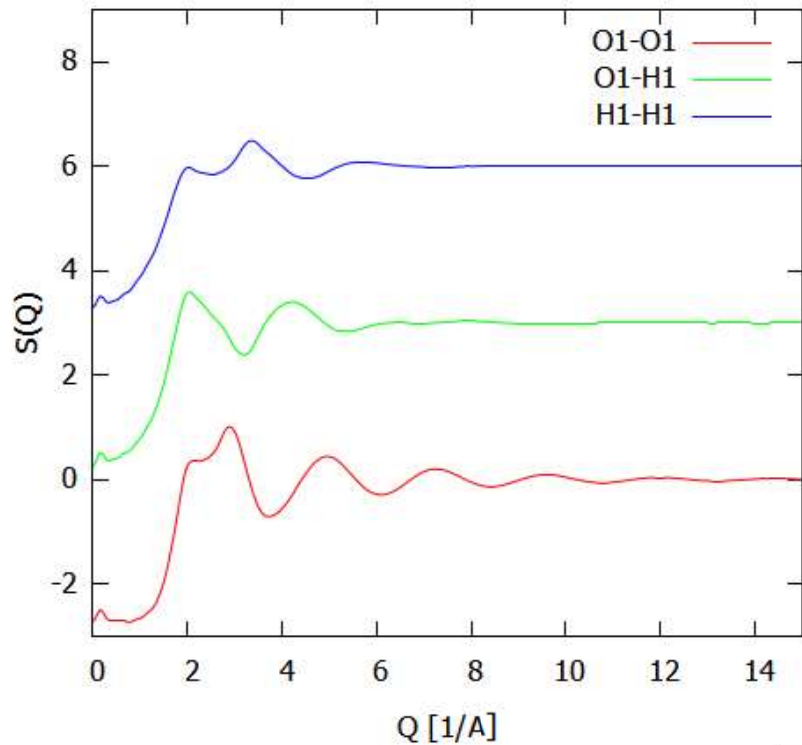
Weighted
sum



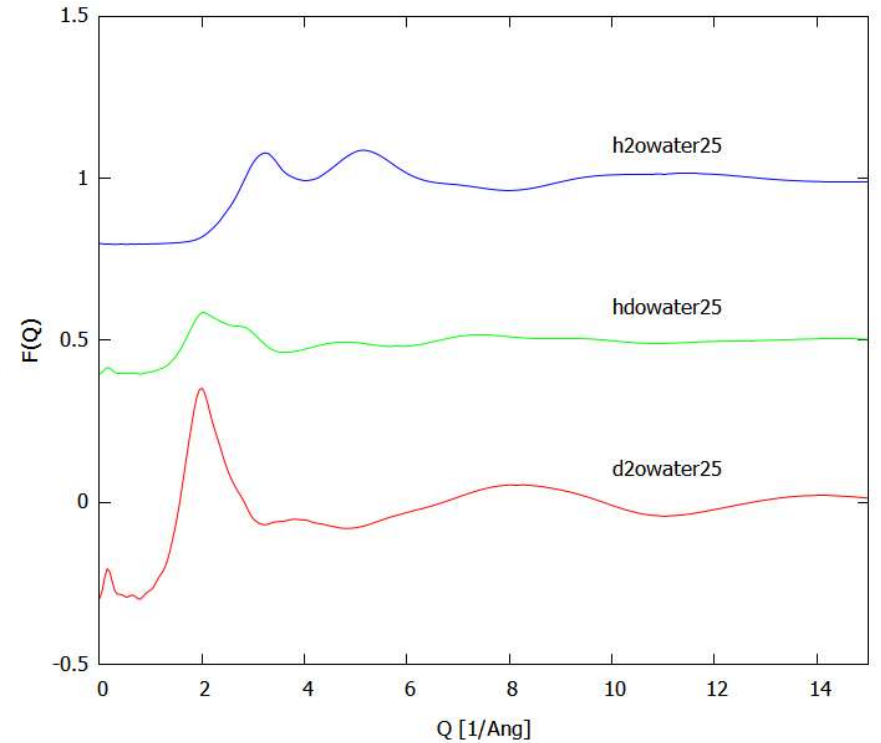
$$F(Q) = \sum_{i,j} (2 - \delta_{ij}) c_i c_j \boxed{b_i b_j} S_{ij}(Q)$$

b_D

S(Q) to F(Q) for water....



Weighted
sum



$$F(Q) = \sum_{i,j} (2 - \delta_{ij}) c_i c_j \boxed{b_i b_j} S_{ij}(Q)$$

b_H, b_D

Aside: Exchangeable hydrogens

O-H and N-H hydrogens can exchange in solution and therefore isotopic substitutions will be mixed up.

1:1 H₂O: D₂O

Mix of H₂O, D₂O and HDO

1:1 H₂O: d₄-methanol

**Mix of H₂O, D₂O, HDO, CD₃-OH
and CD₃-OD**

EPSR and Dissolve deal with this for you, but you must tell the code which hydrogens are exchangeable and which aren't

Inversion of the Scattering Matrix

$$\begin{bmatrix} F_{H_2O}(Q) \\ F_{D_2O}(Q) \\ F_{HDO}(Q) \end{bmatrix} = \begin{bmatrix} 0.0374 & -0.0964 & 0.0622 \\ 0.0374 & 0.1722 & 0.1980 \\ 0.0374 & 0.0378 & -0.0277 \end{bmatrix} \begin{bmatrix} S_{OO}(Q) \\ S_{OH}(Q) \\ S_{HH}(Q) \end{bmatrix} \quad w_{ij} = (2 - \delta_{ij})c_i c_j b_i b_j$$

Inversion of the Scattering Matrix

$$\begin{bmatrix} F_{H_2O}(Q) \\ F_{D_2O}(Q) \\ F_{HDO}(Q) \end{bmatrix} = \begin{bmatrix} 0.0374 & -0.0964 & 0.0622 \\ 0.0374 & 0.1722 & 0.1980 \\ 0.0374 & 0.0378 & -0.0277 \end{bmatrix} \begin{bmatrix} S_{OO}(Q) \\ S_{OH}(Q) \\ S_{HH}(Q) \end{bmatrix} \quad w_{ij} = (2 - \delta_{ij})c_i c_j b_i b_j$$

Inversion of the Scattering Matrix

$$\begin{bmatrix} F_{H_2O}(Q) \\ F_{D_2O}(Q) \\ F_{HDO}(Q) \end{bmatrix} = \begin{bmatrix} 0.0374 & -0.0964 & 0.0622 \\ 0.0374 & 0.1722 & 0.1980 \\ 0.0374 & 0.0378 & -0.0277 \end{bmatrix} \begin{bmatrix} S_{OO}(Q) \\ S_{OH}(Q) \\ S_{HH}(Q) \end{bmatrix} \quad w_{ij} = (2 - \delta_{ij})c_i c_j b_i b_j$$

Inversion of the Scattering Matrix

$$\begin{bmatrix} F_{H_2O}(Q) \\ F_{D_2O}(Q) \\ F_{HDO}(Q) \end{bmatrix} = \begin{bmatrix} 0.0374 & -0.0964 & 0.0622 \\ 0.0374 & 0.1722 & 0.1980 \\ 0.0374 & 0.0378 & -0.0277 \end{bmatrix} \begin{bmatrix} S_{OO}(Q) \\ S_{OH}(Q) \\ S_{HH}(Q) \end{bmatrix} \quad w_{ij} = (2 - \delta_{ij})c_i c_j b_i b_j$$

Inversion of the Scattering Matrix

$$\begin{bmatrix} F_{H_2O}(Q) \\ F_{D_2O}(Q) \\ F_{HDO}(Q) \end{bmatrix} = \begin{bmatrix} 0.0374 & -0.0964 & 0.0622 \\ 0.0374 & 0.1722 & 0.1980 \\ 0.0374 & 0.0378 & -0.0277 \end{bmatrix} \begin{bmatrix} S_{OO}(Q) \\ S_{OH}(Q) \\ S_{HH}(Q) \end{bmatrix} \quad w_{ij} = (2 - \delta_{ij})c_i c_j b_i b_j$$



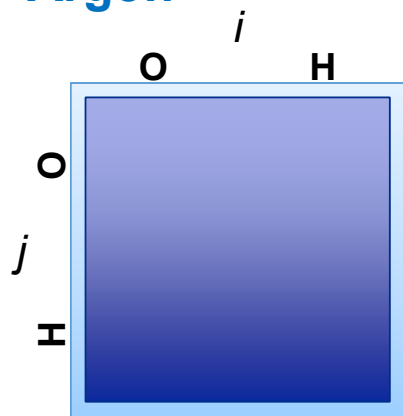
$$\begin{bmatrix} S_{OO}(Q) \\ S_{OH}(Q) \\ S_{HH}(Q) \end{bmatrix} = \begin{bmatrix} 4.8238 & -2.7037 & 24.6061 \\ -5.8227 & 1.6265 & 4.1962 \\ 4.1525 & 4.1525 & -8.3050 \end{bmatrix} \begin{bmatrix} F_{H_2O}(Q) \\ F_{D_2O}(Q) \\ F_{HDO}(Q) \end{bmatrix} \quad w_{ij}^{-1}$$

- Direct extraction of individual $S(Q)$ is possible for 'simple' systems

Partial Pair Correlations

- An “atom type” is a specific kind of atom in the system
 - Depends at least on the element. Can be split by chemical environment.
 - Does not depend on isotope.

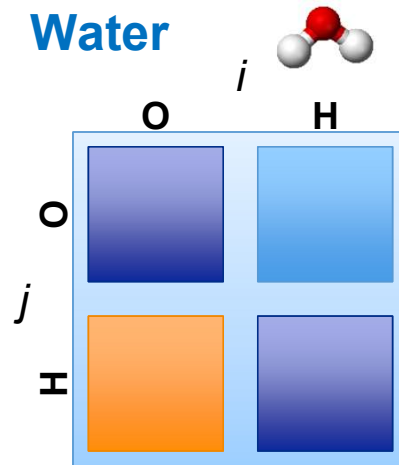
Argon



$N = 1$

Also N_2 , O_2 ...

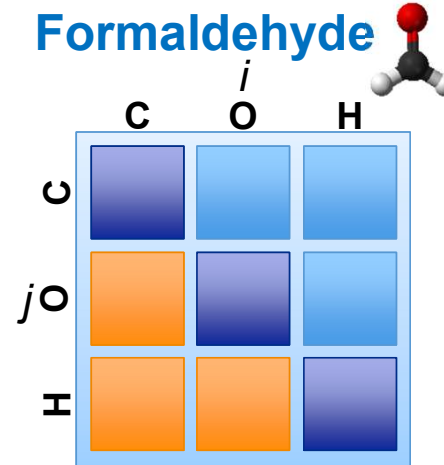
Water



$N = 3$

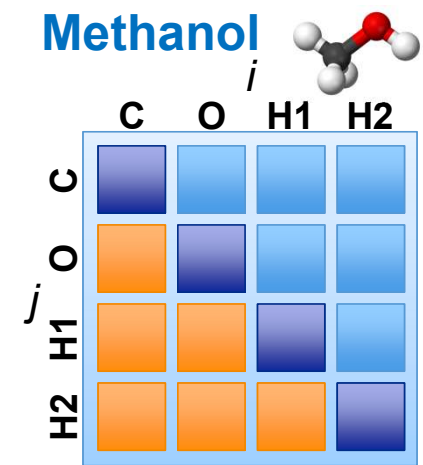
Also benzene, silica...

Formaldehyde



$N = 6$

Methanol



$N = 10$

G(r): FT of data F(Q)

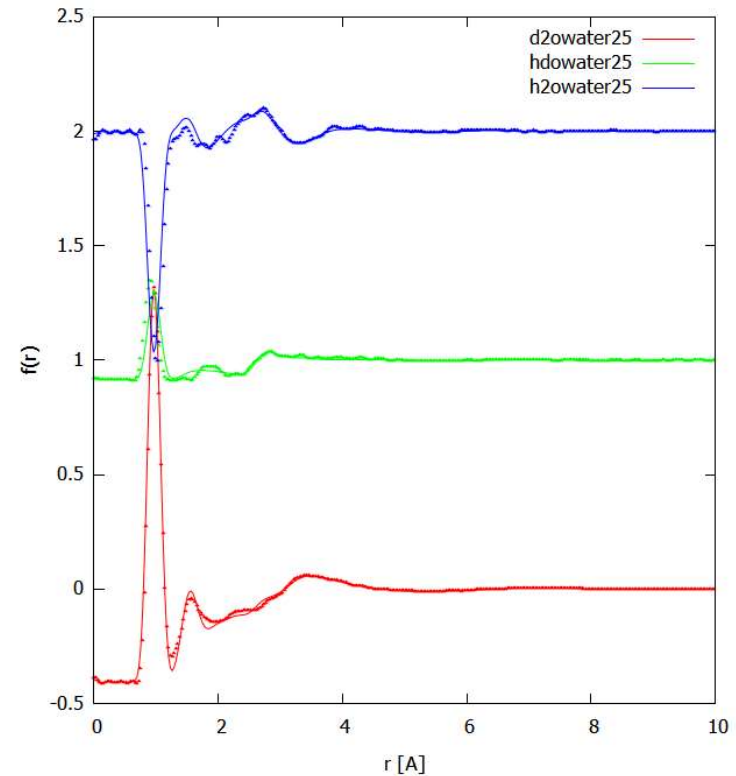
- The total pair distribution function, obtained as a FT of the data, can also be useful to compare to simulation – particularly for local intramolecular structure (bonds, angles etc) that are **not** refined to the data

From data

$$G(r) = \frac{1}{(2\pi)^3 \rho_0} \int_0^\infty 4\pi Q^2 F(Q) \frac{\sin Qr}{Qr} dQ.$$

From Simulation

$$G(r) = \sum_{i,j=1}^n c_i c_j \bar{b}_i \bar{b}_j [g_{ij}(r) - 1],$$



Bragg Scattering

- By usual method scattering calculation limited to **local structure only** (i.e. half the box length)
- Can be calculated by different method for low Q which then includes periodic structure based on boundary conditions
- Calculation is more computationally expensive at higher Q , so need to set sensible upper Q limit.
- Peak widths are another adjustable parameter (based on crystallinity of sample)



Science and
Technology
Facilities Council

ISIS Neutron and
Muon Source

Molecular Simulation: Key concepts



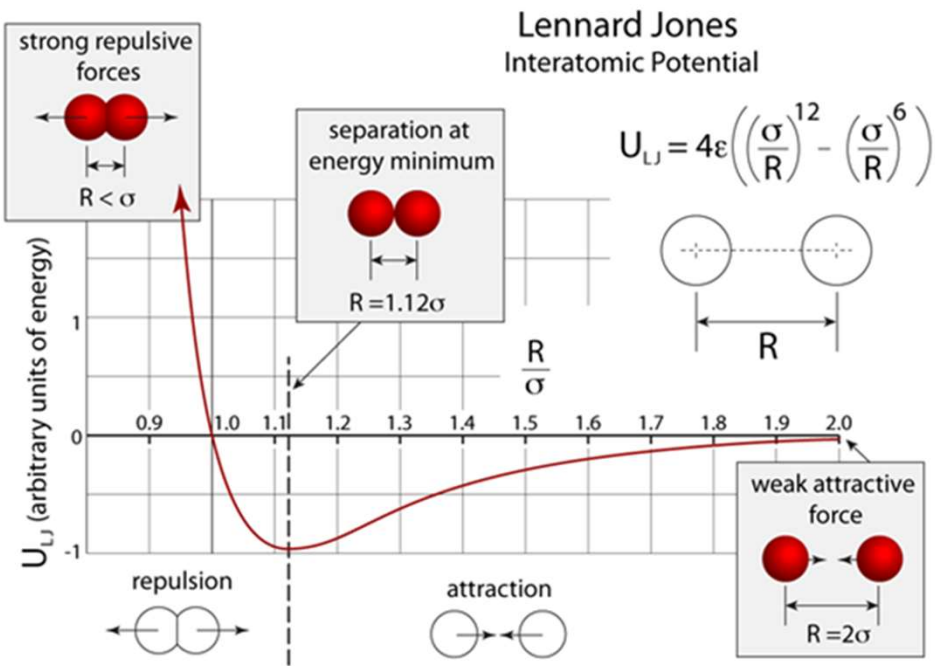
VERY quick background on molecular simulation

We need:

1. A force-field that describes the energy/force on an atom as function of position w.r.t. other atoms
2. A way evolving the system that correctly represents thermodynamic state point
 - a) Molecular dynamics: essentially iteratively evolving Newton's equations of Motion
 - b) **Monte Carlo simulation:**
 1. Make a "move"
 2. Calculate new energy, U , based on FF
 3. If $U < 0$ accept move.
 4. If $U > 0$ accept move with probability

$$\beta = \exp\left(\frac{-\Delta U}{k_B T}\right)$$

Forcefields: Intermolecular* (through space)



Coulomb:

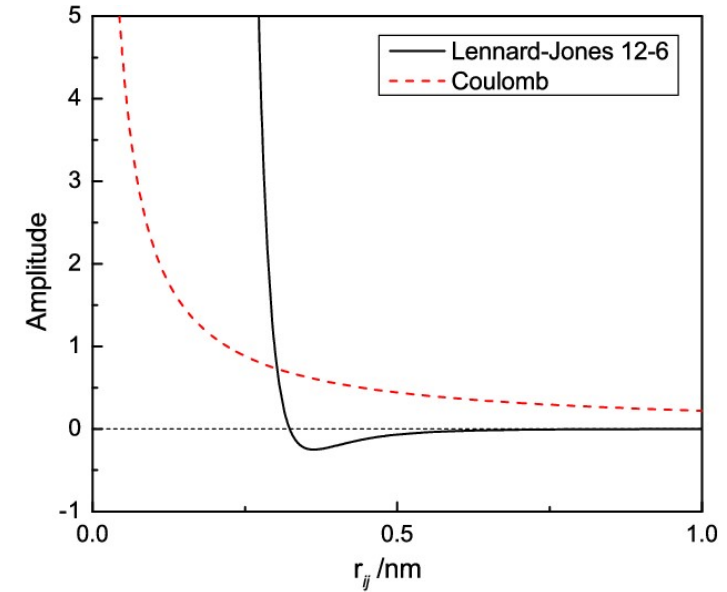
$$U_{Coul} = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}^2}$$

Three terms (per atom type):

ϵ – LJ energy term

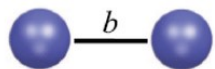
σ – LJ distance term

q – (partial) charge



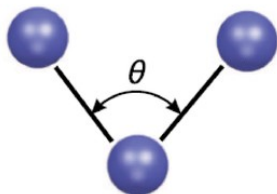
Forcefields: Intramolecular

$$V_{bond} = k_b(r - r_0)^2$$

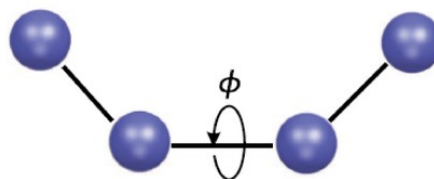


Bond stretching

$$V_{angle} = k_\theta(\theta - \theta_0)^2$$

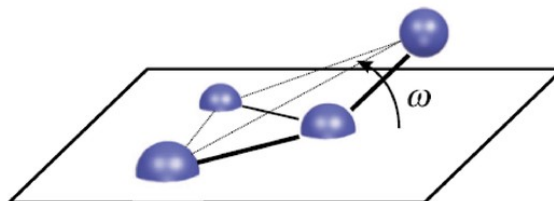


Angle bending



Proper dihedral torsion

$$V_{dih} = k_\phi(1 + \cos(n\phi + 1))$$



Improper dihedral torsion

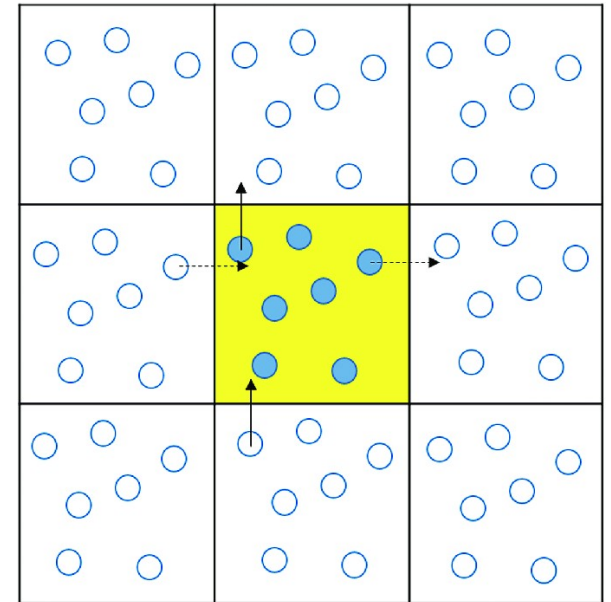
$$V_{imp} = k_{imp}(\omega - \omega_0)^2$$

NOTE: EPSR and *Dissolve* treat intermolecular potentials in different ways

EPSR defines the strength of the bond/angle/dihedral with an effective potential for **all atoms** weighted by the reduced mass of the pair – This is “floppier” than standard bonded potentials to account for zero point energy – meaning neutron “sees” a more disordered atom locations.

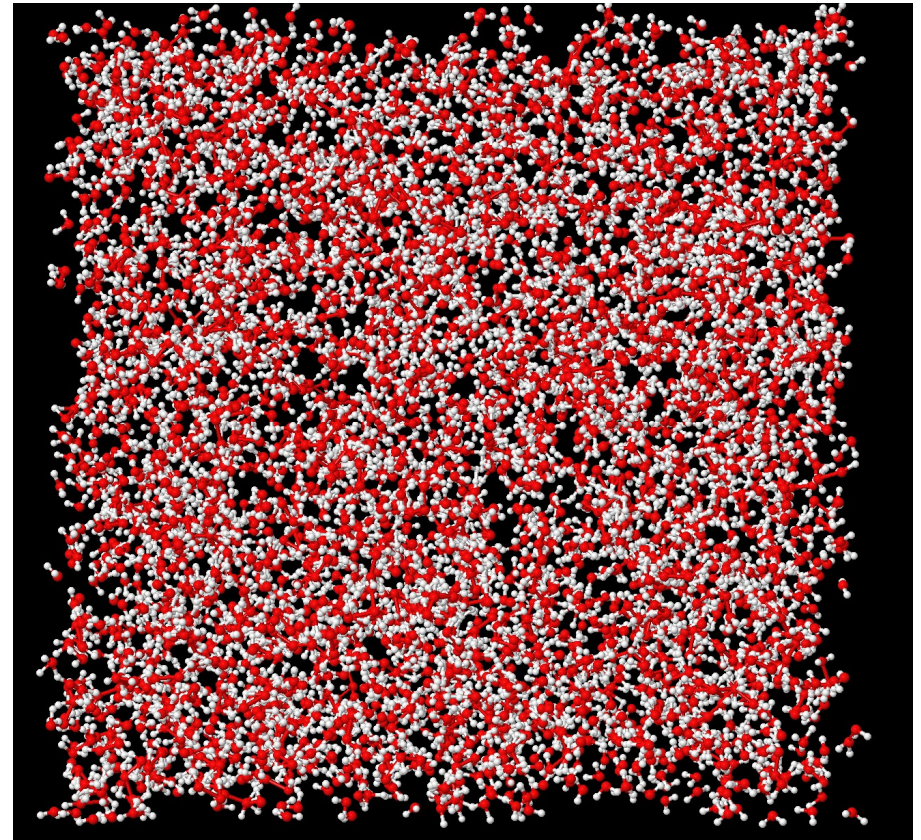
Periodic Boundary Conditions

- We need to represent the bulk with a small box of atoms (3-10nm)
- Achieve this with periodic boundary conditions
- Potential is calculated between atom and the **nearest** version of another atom within these conditions.
- Potentials must therefore be limited “**cut-off**” at a distance shorter than half the box length



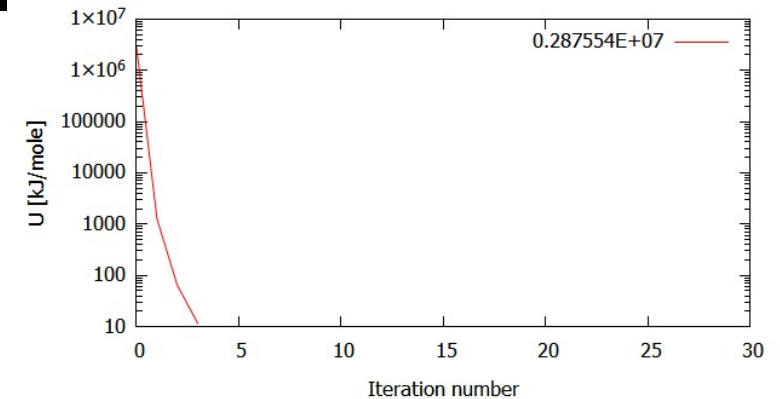
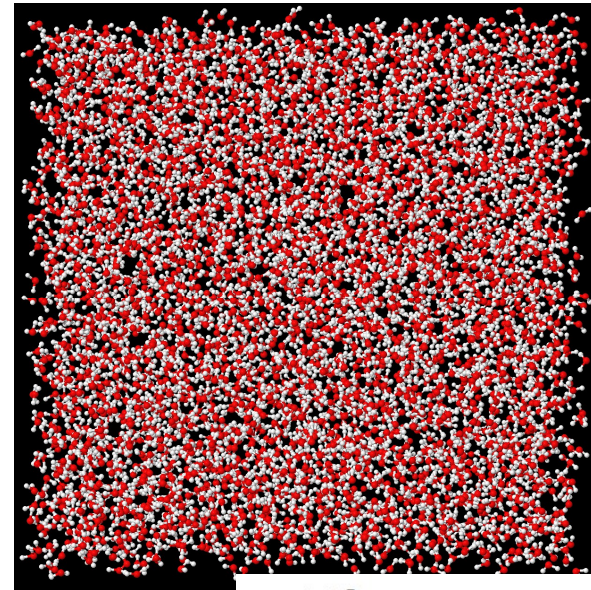
Starting and running a simulation

1. Define your atoms/molecules:
Bonding and FF
2. Fill a box of an appropriate **size, density and composition** - this can be random placement



Starting and running a simulation

1. Define your atoms/molecules:
Bonding and FF
2. Fill a box of an appropriate size, density and composition - this can be random placement
3. **Evolve the system to reach equilibrium**

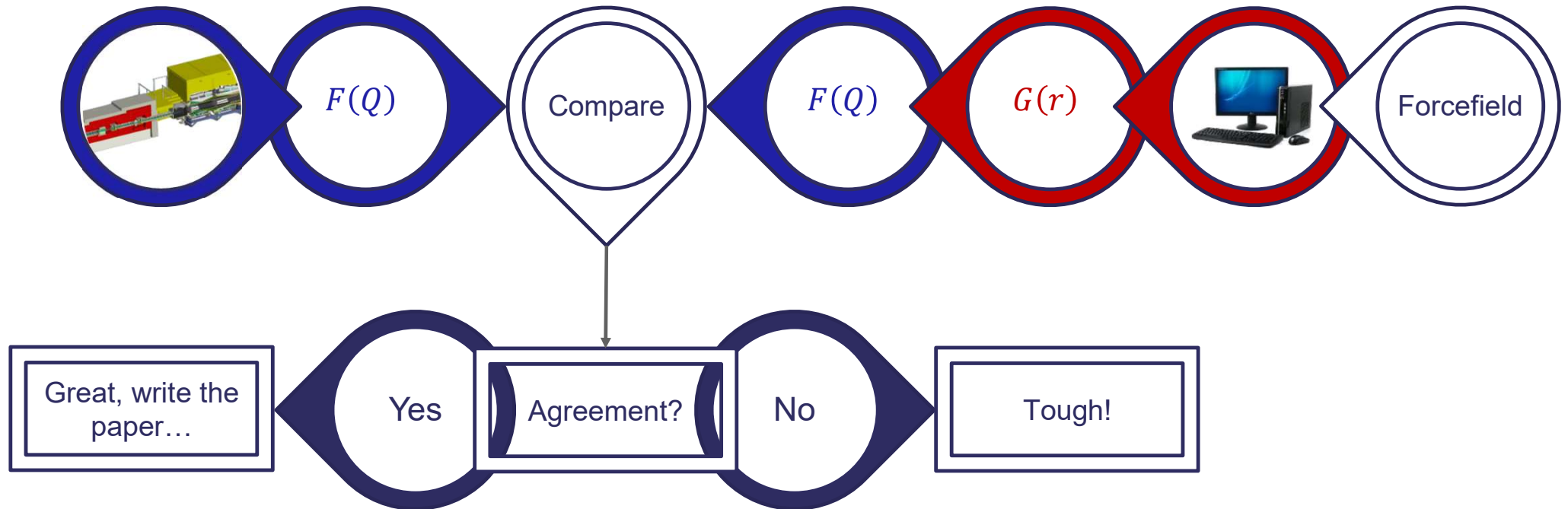


Starting and running a simulation

1. Define your atoms/molecules: Bonding and FF
2. Fill a box of an appropriate size, density and composition - this can be random placement
3. Evolve the system to reach equilibrium
4. **Accumulate and calculate**

Basic Simulation

- Published forcefield parameters describing many systems, so...



Improving on a Basic Simulation

- Need to have a forcefield-based simulation to interpret data in vast majority of cases – literature parameters exist for a variety of systems
- How are these forcefield parameters obtained?
 - Reproduction of thermodynamic properties
 - Extrapolated / inferred from other systems
 - Chosen to reproduce experimental structural data? NO!
- Change these somehow to improve the agreement between simulation experiment?
 - Tweaking by hand almost impossible
 - Fitting algorithm could work, but parameter space is large



Science and
Technology
Facilities Council

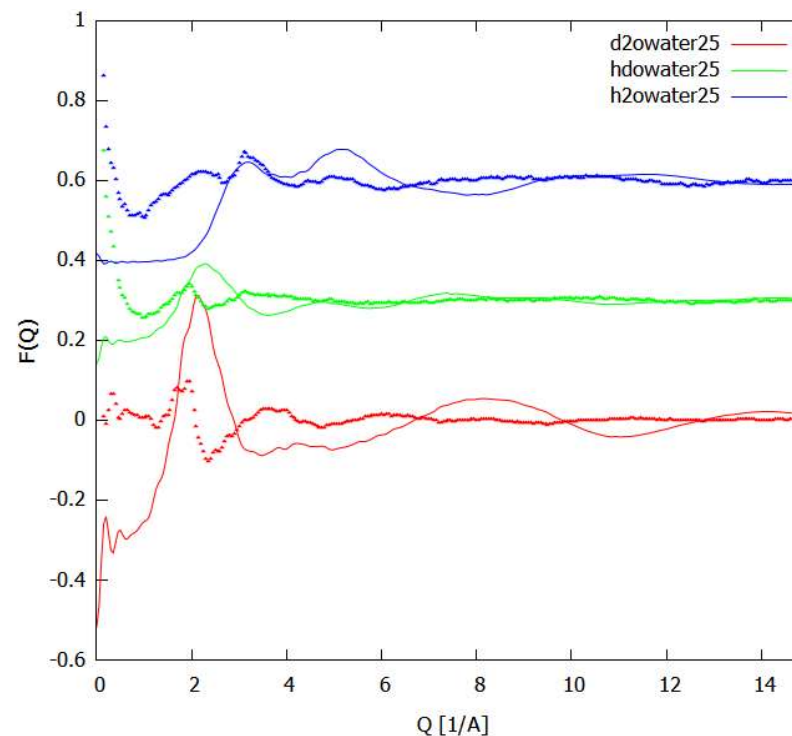
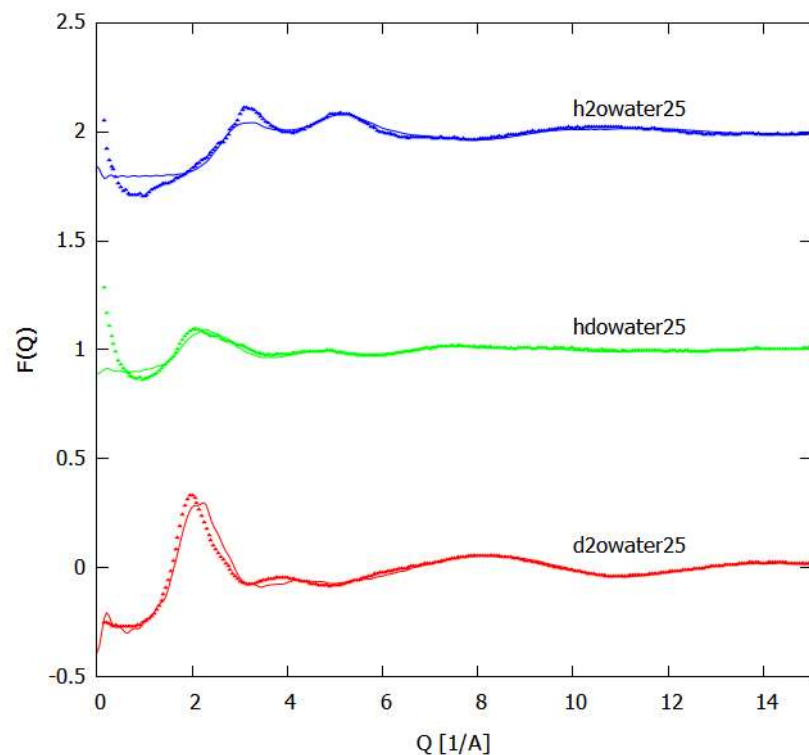
ISIS Neutron and
Muon Source

Potential Refinement



Potential Refinement: EPSR method

1) Take differences between experimental and simulated $F(Q)$ to get $\Delta F(Q)$



Potential Refinement: EPSR method

- 1) Take differences between experimental and simulated $F(Q)$ to get $\Delta F(Q)$
- 2) Enter these $\Delta F(Q)$ into the inverse scattering matrix to generate $\Delta S(Q)$

$$\begin{bmatrix} \Delta S_{OO}(Q) \\ \Delta S_{OH}(Q) \\ \Delta S_{HH}(Q) \end{bmatrix} = \begin{bmatrix} 4.8238 & -2.7037 & 24.6061 \\ -5.8227 & 1.6265 & 4.1962 \\ 4.1525 & 4.1525 & -8.3050 \end{bmatrix} \begin{bmatrix} \Delta F_{H_2O}(Q) \\ \Delta F_{D_2O}(Q) \\ \Delta F_{HDO}(Q) \end{bmatrix}$$

$$\begin{bmatrix} S_{OO}(Q) \\ S_{OH}(Q) \\ S_{HH}(Q) \end{bmatrix} = \begin{bmatrix} 4.8238 & -2.7037 & 24.6061 \\ -5.8227 & 1.6265 & 4.1962 \\ 4.1525 & 4.1525 & -8.3050 \end{bmatrix} \begin{bmatrix} F_{H_2O}(Q) \\ F_{D_2O}(Q) \\ F_{HDO}(Q) \end{bmatrix}$$

Potential Refinement: EPSR method

- 1) Take differences between experimental and simulated $F(Q)$ to get $\Delta F(Q)$
- 2) Enter these $\Delta F(Q)$ into the inverse scattering matrix to generate $\Delta S(Q)$

$$\begin{bmatrix} \Delta S_{OO}(Q) \\ \Delta S_{OH}(Q) \\ \Delta S_{HH}(Q) \end{bmatrix} = \begin{bmatrix} 4.8238 & -2.7037 & 24.6061 \\ -5.8227 & 1.6265 & 4.1962 \\ 4.1525 & 4.1525 & -8.3050 \end{bmatrix} \begin{bmatrix} \Delta F_{H_2O}(Q) \\ \Delta F_{D_2O}(Q) \\ \Delta F_{HDO}(Q) \end{bmatrix}$$

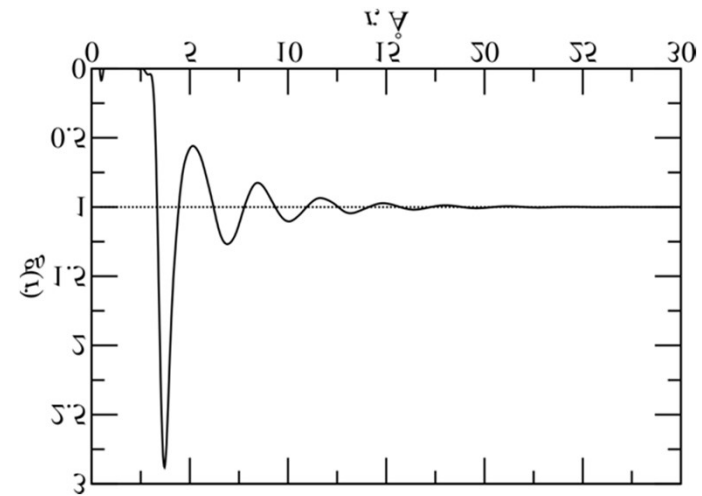
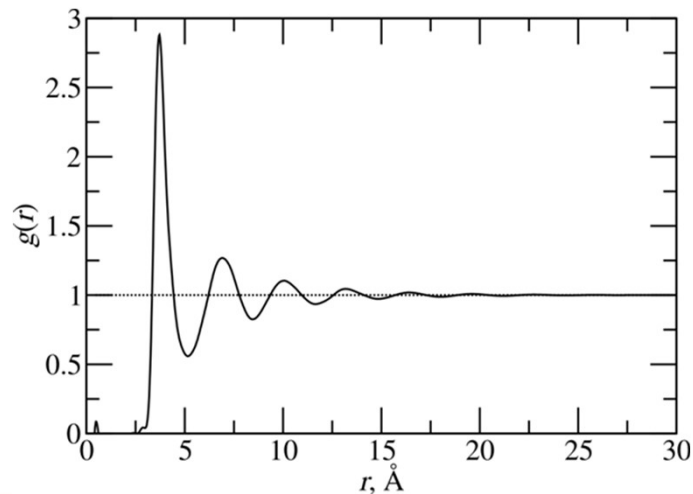
- 3) Fourier transform the $\Delta S(Q)$ into $\Delta g(r)$ and use these to form an additional, empirical potential for each atom type pair

$$\Delta S_{ij}(Q) = \rho \int_0^{\infty} 4\pi r^2 \Delta g_{ij}(r) \frac{\sin Qr}{Qr} dr$$

Aside: Linking $g(r)$ and potential $U(r)$

Potential of mean force:

$$U(r) = -kT \ln g(r)$$



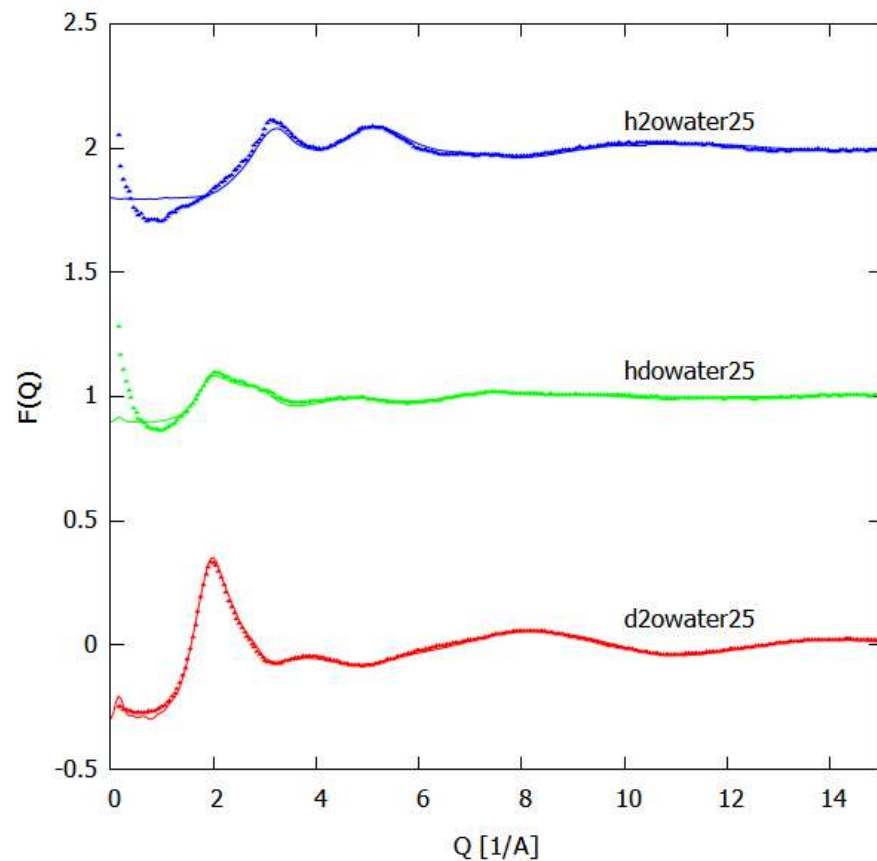
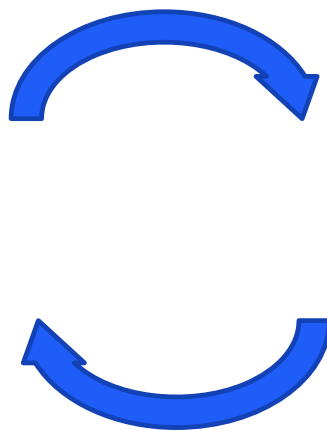
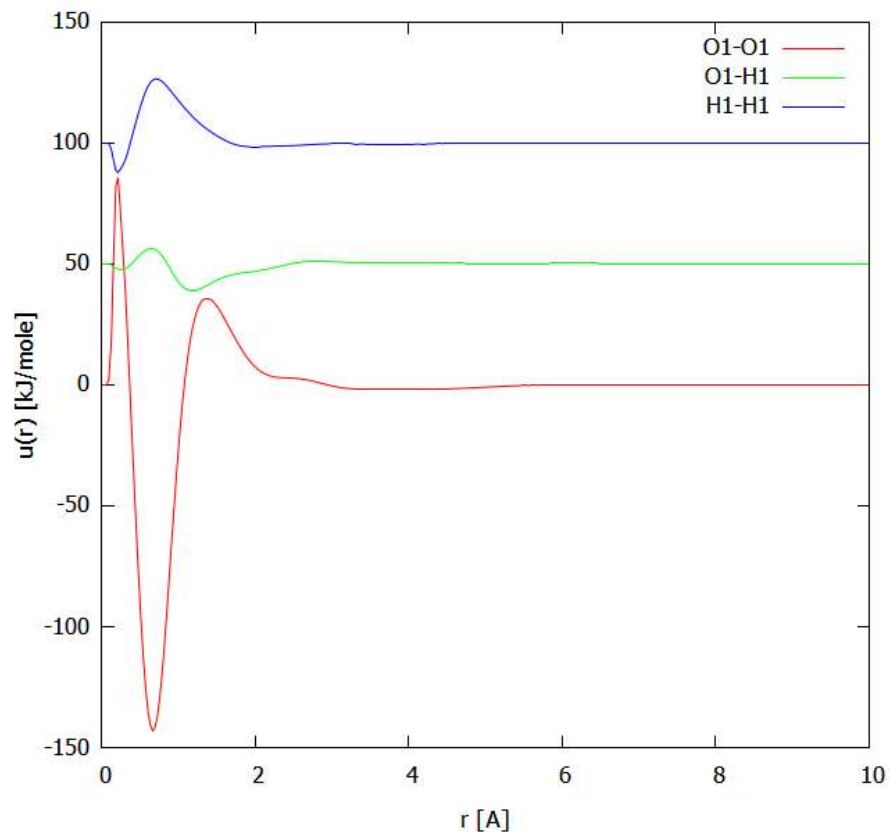
Potential Refinement: EPSR method

- 1) Take differences between experimental and simulated $F(Q)$ to get $\Delta F(Q)$
- 2) Enter these $\Delta F(Q)$ into the inverse scattering matrix to generate $\Delta S(Q)$

$$\begin{bmatrix} \Delta S_{OO}(Q) \\ \Delta S_{OH}(Q) \\ \Delta S_{HH}(Q) \end{bmatrix} = \begin{bmatrix} 4.8238 & -2.7037 & 24.6061 \\ -5.8227 & 1.6265 & 4.1962 \\ 4.1525 & 4.1525 & -8.3050 \end{bmatrix} \begin{bmatrix} \Delta F_{H_2O}(Q) \\ \Delta F_{D_2O}(Q) \\ \Delta F_{HDO}(Q) \end{bmatrix}$$

- 3) **Fourier transform the $\Delta S(Q)$ into $\Delta g(r)$ and use these to form an additional, empirical potential for each atom type pair**
- 4) Repeatedly run the simulation and generate additional potentials until the experimental and simulated $F(Q)$ 'match'

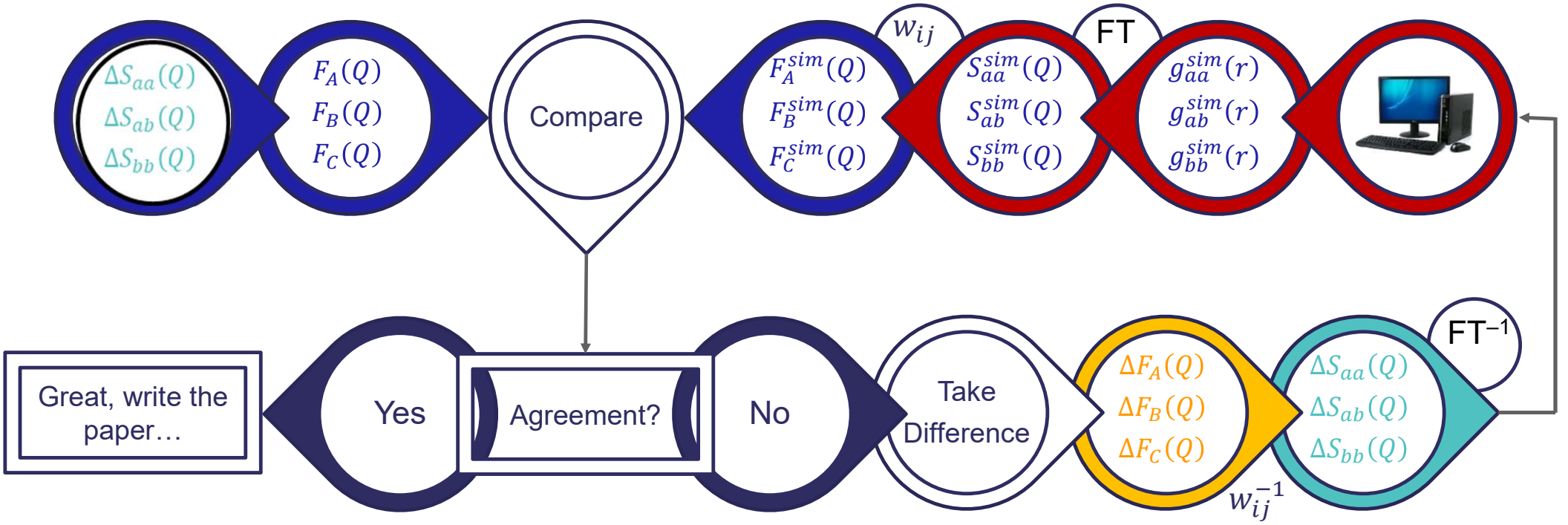
Potential Refinement: EPSR results



ereq: Maximum energy of the empirical potential

- This parameter is set in EPSR or *Dissolve* software to provide a maximum limit for the empirical potential energy
- You may need to try a few values to get the best fit to the data
- Too low: empirical potential energy not sufficient to reach a good level of fit
- Too high: “weird stuff can happen”. The empirical potential is meant to be a perturbation on the seed potential. For example it can make the system too cohesive – leading to voids.

“Enhanced” Simulation



- Experimental datasets actively used to derive new pair potentials

Isotope Availability & Access

- For many systems, there are too few isotopic sample possibilities available
 - Elements for which there are no suitable isotope pairs
 - Isotopic exchange prohibited by difficulty in synthesising target compounds
 - System complexity means there may be too many measurements to make in the timescale of a reasonable experiment

- For instance, silica:

$$\left[F_{SiO_2}(Q) \right] = \begin{bmatrix} 0.0191 & 0.1070 & 0.1497 \end{bmatrix} \begin{bmatrix} S_{SiSi}(Q) \\ S_{SiO}(Q) \\ S_{OO}(Q) \end{bmatrix}$$

- No inversion possible, so no access to partial $S(Q)$

Generalisation

- What is a general method to allow inversion of the underdetermined case?

$$\left[F_{Si_2}(Q) \right] = \begin{bmatrix} 0.0191 & 0.1070 & 0.1497 \end{bmatrix} \begin{bmatrix} S_{SiSi}(Q) \\ S_{SiO}(Q) \\ S_{OO}(Q) \end{bmatrix}$$

- Use the simulation? It has all the partials we need, which are hopefully a good “guess” for the experimental ones
- Define a feedback factor, $0 < f < 1$ (typically 0.7-0.9), and write new weights factors...

$$w_{ij} = f(2 - \delta_{ij})c_i c_j b_i b_j$$

$$x_{ij} = (1 - f)$$

The Augmented Scattering Matrix

- For each partial in the simulation, add a row to the scattering matrix weighted by x_{ij}
- Typically, $f \sim 0.9$

$$\begin{bmatrix} F_{SiO_2}(Q) \\ S_{SiSi}^{sim}(Q) \\ S_{SiO}^{sim}(Q) \\ S_{OO}^{sim}(Q) \end{bmatrix} = \begin{bmatrix} 0.0172 & 0.0963 & 0.1347 \\ 0.1 & 0.0 & 0.0 \\ 0.0 & 0.1 & 0.0 \\ 0.0 & 0.0 & 0.1 \end{bmatrix} \begin{bmatrix} S_{SiSi}(Q) \\ S_{SiO}(Q) \\ S_{OO}(Q) \end{bmatrix}$$

- Now have an overdetermined matrix for which we can find a (pseudo) inverse



Science and
Technology
Facilities Council

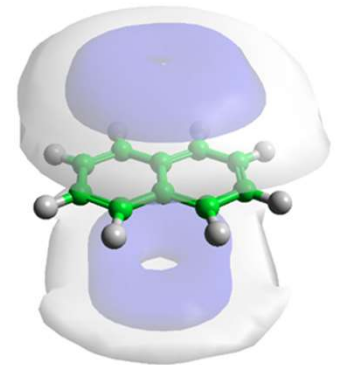
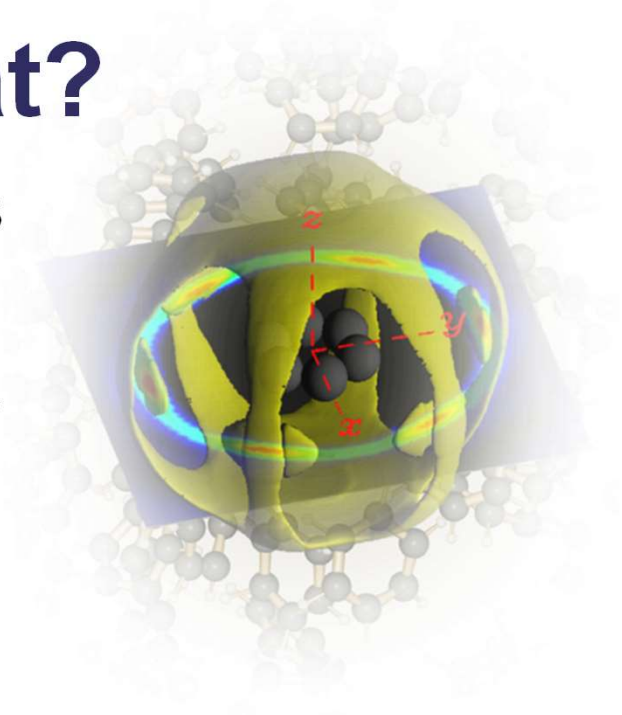
ISIS Neutron and
Muon Source

What next: Simulation analysis, Common pitfalls, Additional features



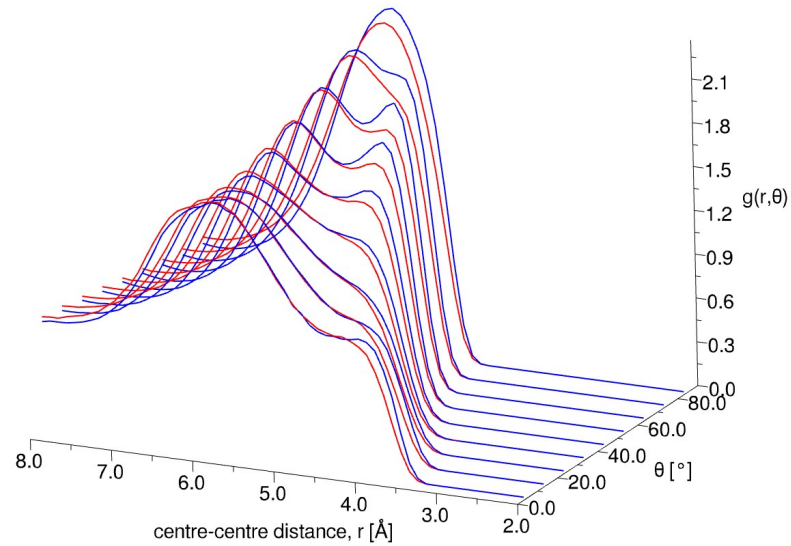
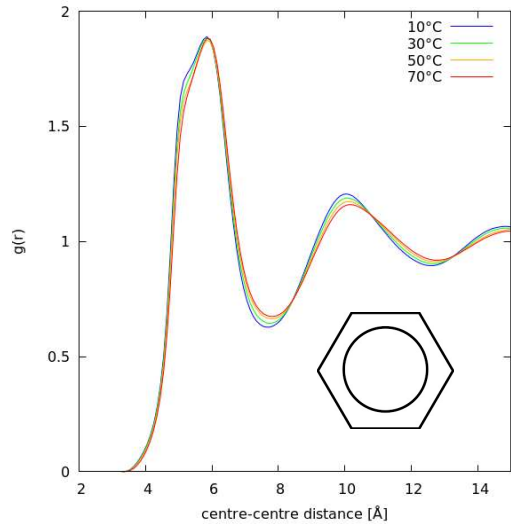
Got a good fit? Now what?

- Run simulation for a large number of iterations (or time steps) for good statistics
- You can analyse the save ensemble/trajectory for any structural information you want
- Tools are available in EPSR and *Dissolve*
- Or the atomistic coordinates can be saved and analysed using external software



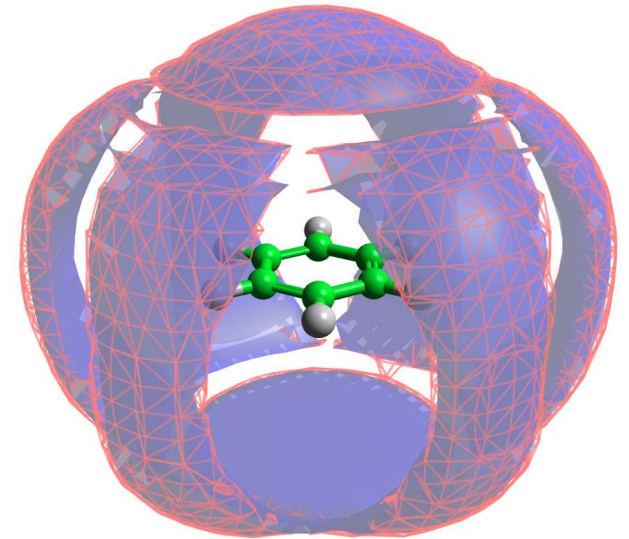
Options for analysis:

1D $g(r)$'s – not necessarily from atom sites



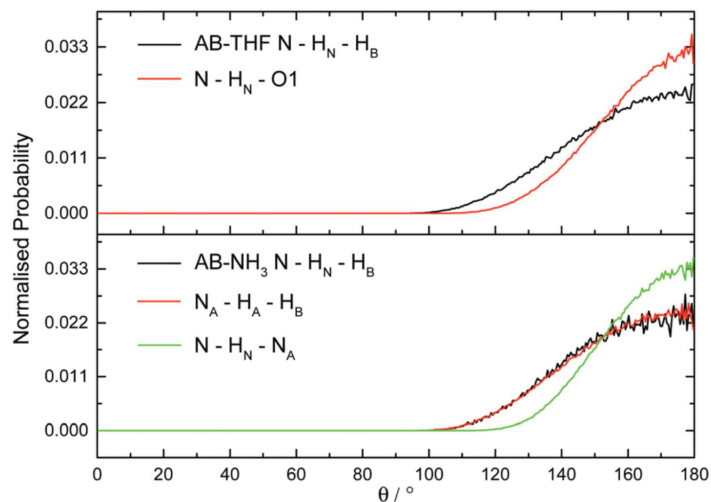
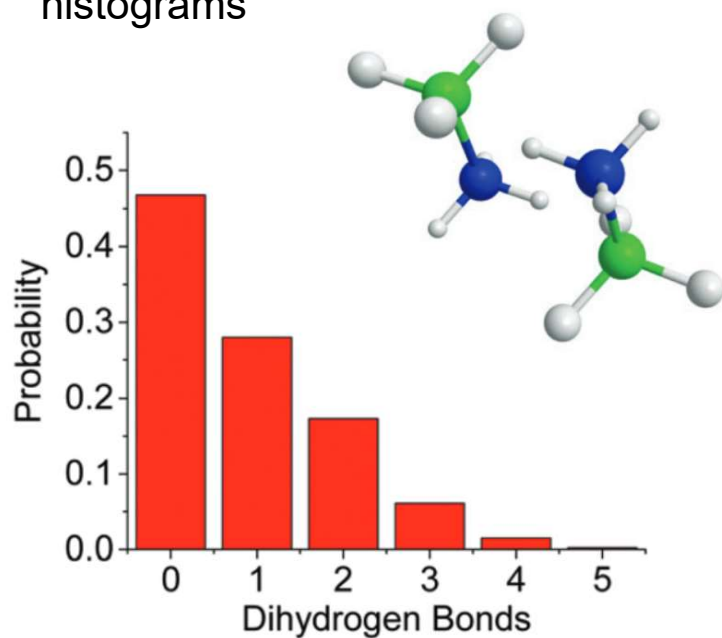
2D angular radial distribution functions

3D spatial density functions



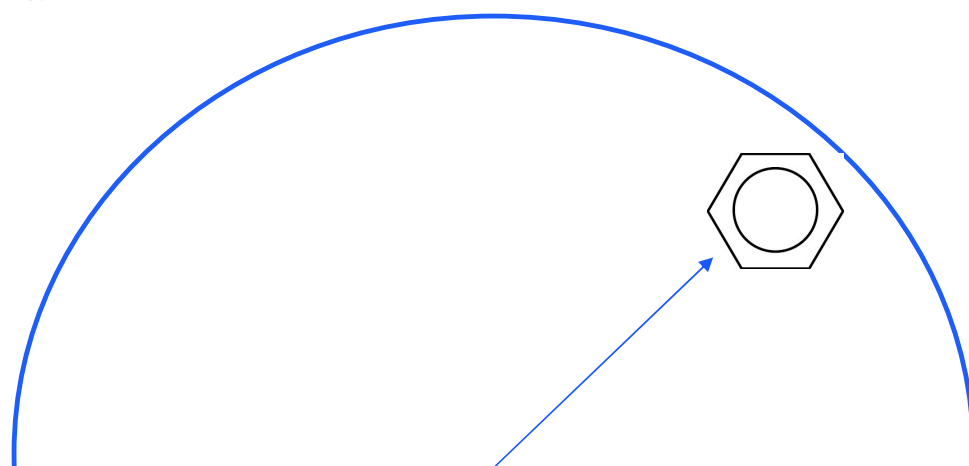
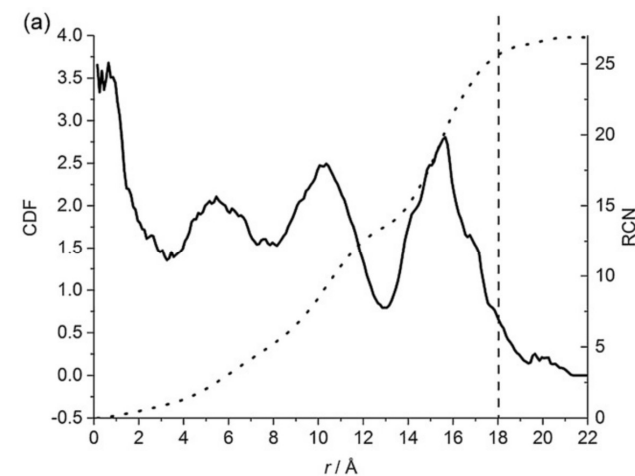
Options for analysis:

Coordination number histograms



Bond angle histograms

Cylindrical distribution functions



Garbage in, garbage out

- Double check input parameters for forcefields – automation of this process can help (as long as that process works!)
- Check all output functions are behaving as expected e.g. $g(r)$'s – have in mind what the expected behaviour is

Finite size effects

- Is the size of the simulation box large enough to accurately represent the system of interest?
- Check for finite size effects: Increase box size, any change in structure?
- Is the cut-off long enough to represent structure in the system?
- If the box is too large, this can make the simulation too slow.

Reaching equilibrium

- Simulation can get stuck in a local (high energy) minimum
- Simulations of mesoscale systems (e.g. surfactants in solution) may take a long time to reach equilibrium - may need build the simulation box close to the “solution” rather than randomly
- Recommend checking the output coordinate files to check for appropriate sampling



Additional features

- Bragg scattering calculation: For crystalline and mixed phase systems (e.g. confined fluids in porous media) – *Dissolve* and EPSR
- X-ray scattering calculation: Allows use of HEXRD data as additional contrast - *Dissolve* and EPSR

Acknowledgements

- Tristan Youngs
- Alan Soper
- Daniel Bowron and DM Group

- Chris Howard, Neal Skipper (UCL)