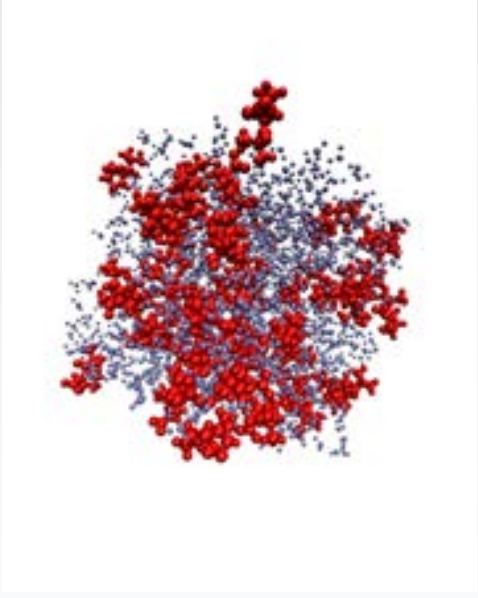
Modeling biomolecules in solution

pitfalls and challenges

Sylvia McLain

25th April 2013
Accuracy in Powder Diffraction IV



DEPARTMENT OF BIOCHEMISTRY
UNIVERSITY OF OXFORD

What is biology?

Biology
The study of living organisms, divided into many specialized fields that cover their morphology, physiology, anatomy, behavior, origin, and distribution
the plants and animals of a particular area as in "the biology of the Chesapeake Bay"
the physiology, behavior, and other qualities of a particular organism or class of organisms: "human biology"

Molecular biology
The branch of biology *that deals with the structure and function of the macromolecules (e.g. proteins and nucleic acids) essential to life.*

Structural biology
branch of molecular biology *which is concerned with macromolecular structure and how this effects function*

Challenge 1

Challenge 1: Convincing biologists understanding structures really is biology

Max Perutz

John Kendrew

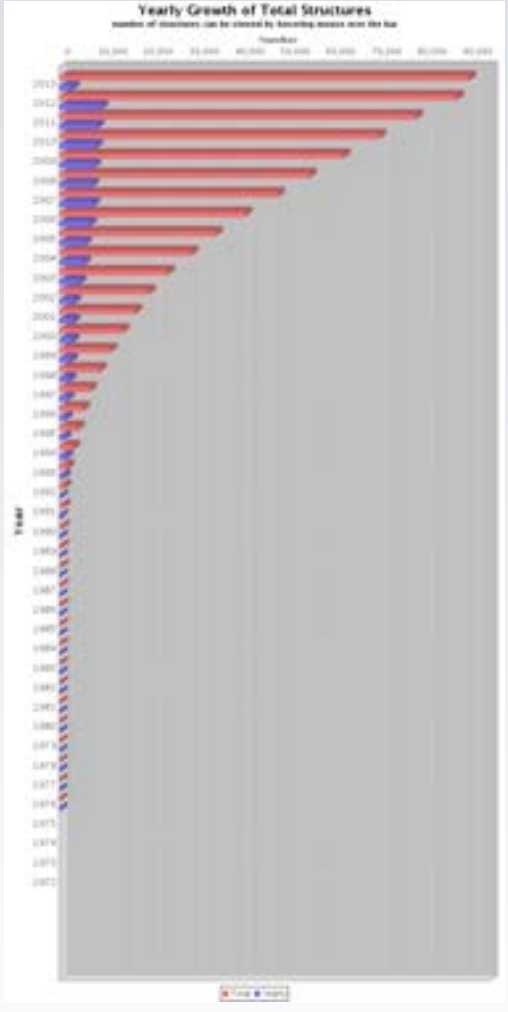
Dorothy Hodgkin + Linus Pauling

First published structure of globular myoglobin 1958
- Kendrew

First structure of myoglobin at 5.5 Å resolution
- Perutz

Insulin 1968 at 2.8 Å resolution
- Hodgkin

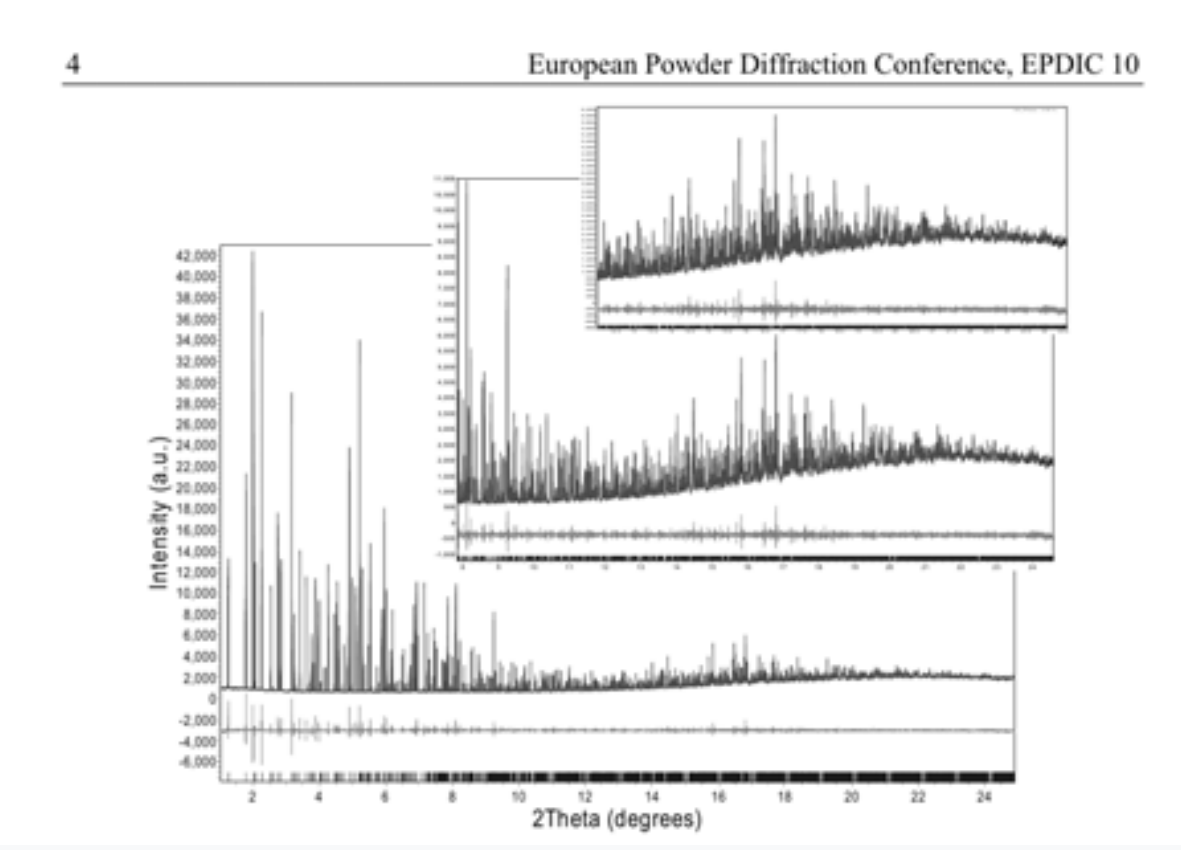
Protein crystallography



- Protein crystallography data base (PDB)
- In 1972 there were 2 structures in the PDB
- 89,740 structures in the PDB (many are repeats)
- 18 are from Powder diffraction (X-ray)
- number of known *human* proteins estimated at 50,000
- most biological molecules do not crystallize

Challenge 2

Challenge 2: Making protein powder diffraction a viable tool for structural biologists



I. Margiolaki *et al. Z. Krist.* Suppl. 26 (2007) 1-13

Challenge 3

Challenge 3: Crystallography (no matter how accurate) doesn't always work

Most biological molecules don't crystallize

Many biological molecules don't crystallize as single molecules - complexes

Many biological molecules crystallize with a high level of disorder (disordered loops)

Most biological molecules are disordered *in vivo*

Even when biological molecules crystallize, they may not be the same in solution

Measuring biological molecules in solution

Role of water in biology not well understood

- role in ligand binding?
- role in association (membranes, protein folding, amyloid fibers)
- hydrophobic/hydrophilic forces
- 'oil and water don't mix' but water crosses membranes!

Water in Biological Systems

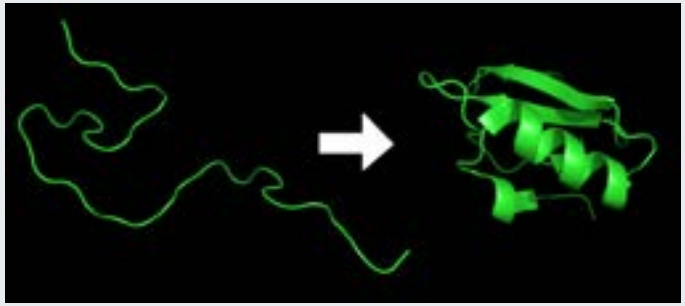
M. J. TAIT & F. FRANKS

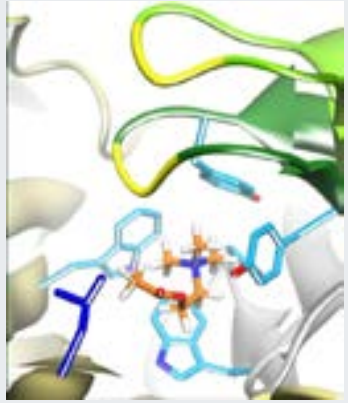
The object of this article is to draw attention to some of the major areas of investigation of the state of water in biochemical and biological systems. Deductions about the role of water in such systems are somewhat speculative, but the situation might be improved.

Measuring biological molecules in solution

In real life water is always around

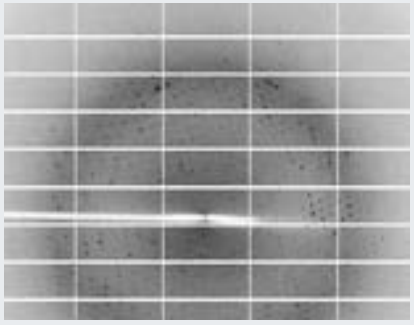
- Protein folding
- Protein/peptide association
- membrane formation
- DNA transcription
- receptor ligand binding interactions



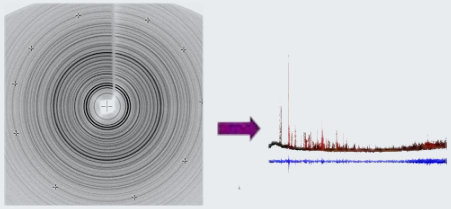


Challenge 4

Challenge 4: no Bragg scattering, disordered systems

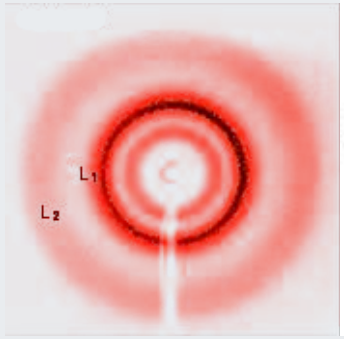


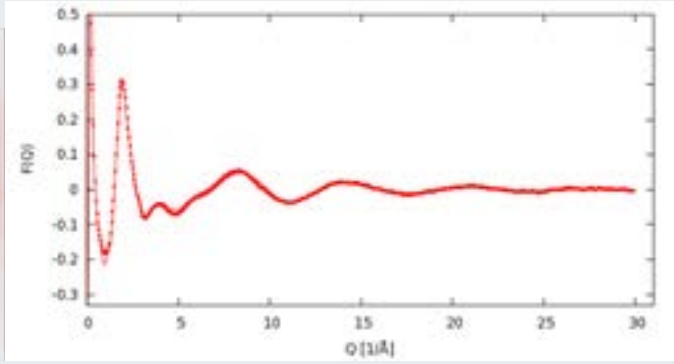
protein crystallography



protein powder crystallography

Liquid diffraction





DEPARTMENT OF BIOCHEMISTRY
UNIVERSITY OF OXFORD

Neutron Diffraction and Empirical Potential Structural Refinement (EPSR)

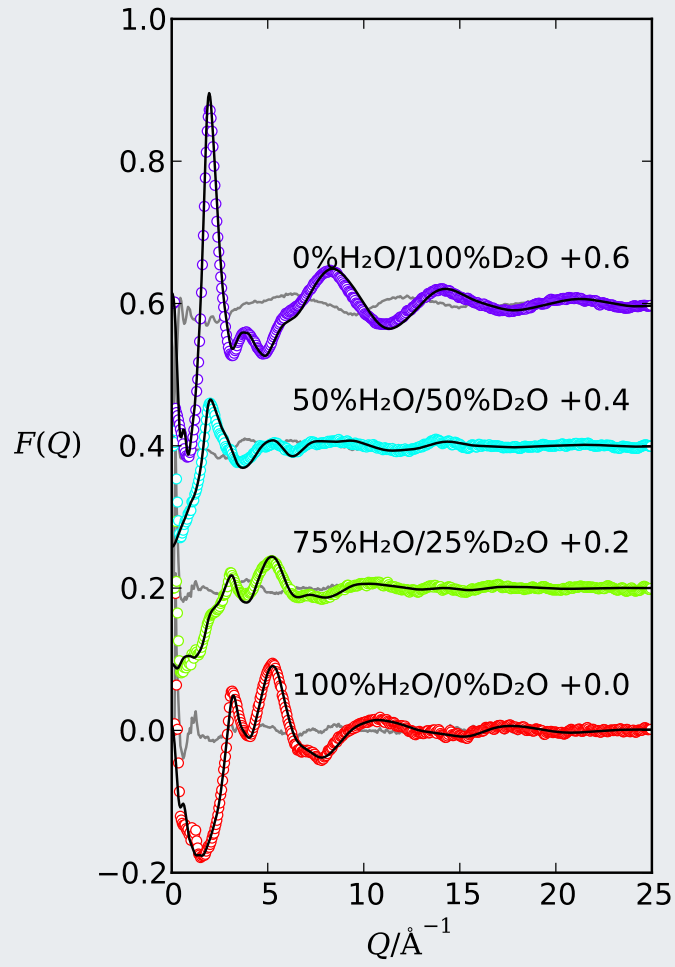
Neutron diffraction with isotopic substitution

- $F(Q)$ - structure factor
- $F(Q) = \sum c_\alpha c_\beta b_\alpha b_\beta S_{\alpha\beta}(Q)$

b - neutron scattering length

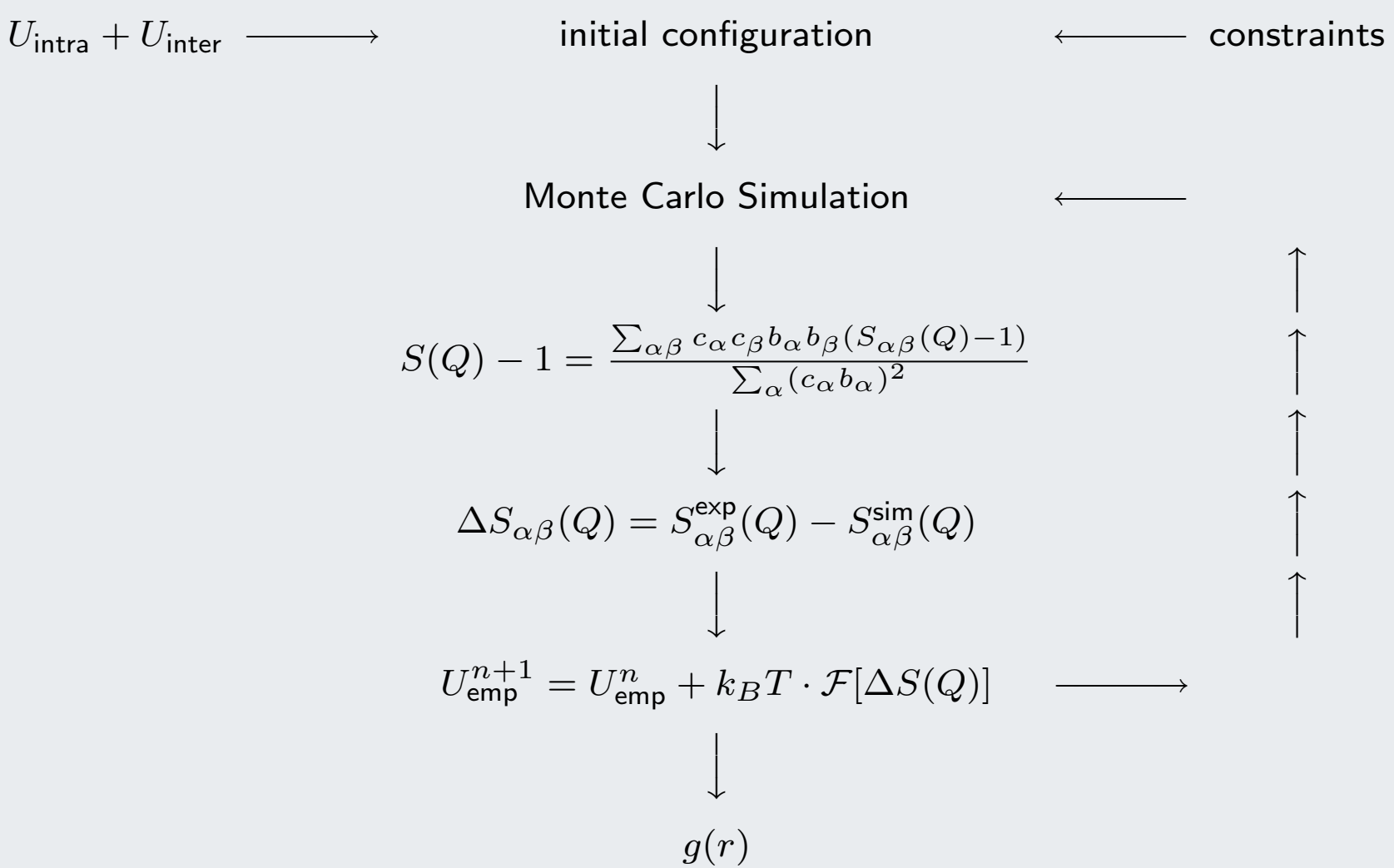
- Different isotopes scatter with different intensity
- Measurement of chemically equivalent isotopically unique samples
- model with EPSR

gpg-peptide in water



S Busch,* *et al.*, *manuscript submitted* (2013)

Empirical Potential Structural Refinement - computer modeling

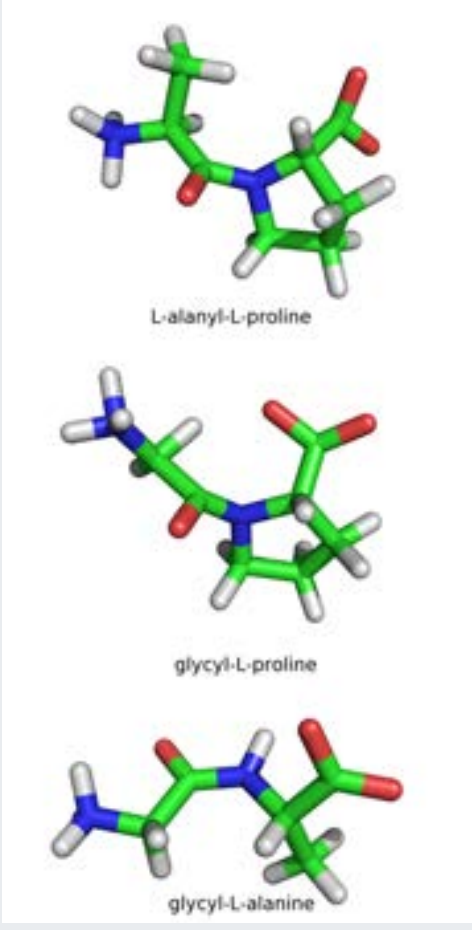


- Model specifically designed for amorphous systems
- Fits a set of neutron data
- Structural model only!

DEPARTMENT OF BIOCHEMISTRY
UNIVERSITY OF OXFORD

Structure of peptides in solution (a.k.a: how we usually tell the story)

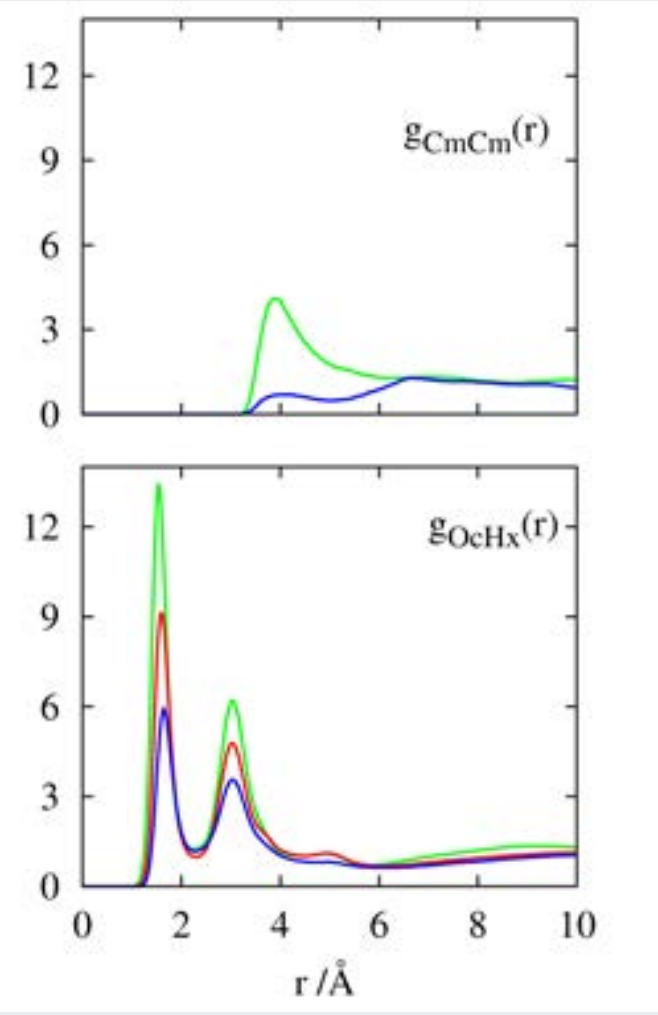
Dipeptides as a model system



- Series of soluble peptides with increasing hydrophobicity
- Very soluble in pure water
- Can use H/D substitution on H atoms
- Measured at high concentrations (2.5 M)

Association between peptides in solution

Peptide-peptide correlations

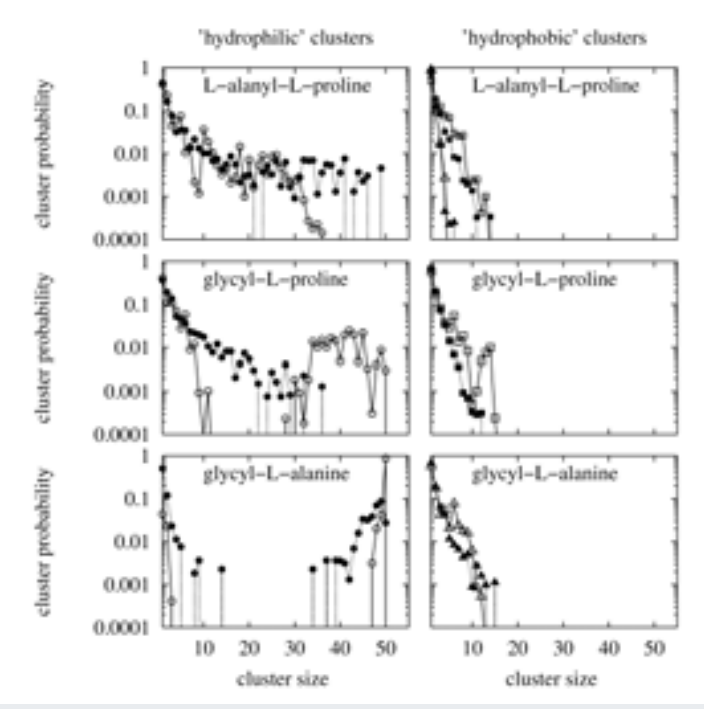


SE McLain,* *et al.*, *Agnew. Chem., Int. Ed.*, **47**, 9059 (2008)

- $g_{OcHx(r)}$ coordination highest in gly-ala (0.74)
- hydrophobic coordination highest in gly-ala (least hydrophobic peptide)

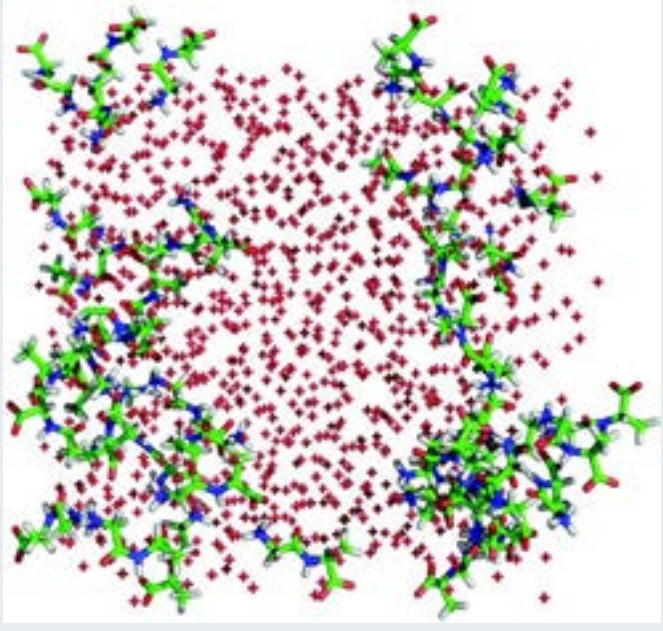
Association of peptides in solution

Clustering analysis of peptides in solution



Open symbols MD, closed symbols EPSR

Association of glycyl-L-alanine in solution from MD simulations



charge-charge interactions drive association in solution

Challenge 4

Challenge 4: Most functional biological molecules are larger than dipeptides

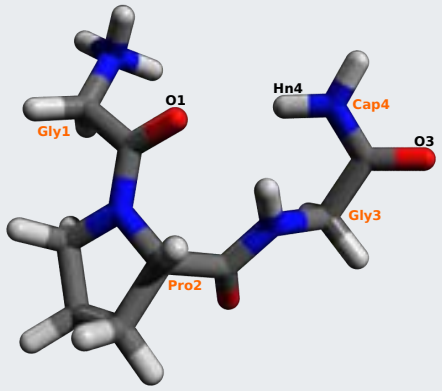
higher level of complexity

lower level of solubility (less material in solution)

disorder, disorder, disorder

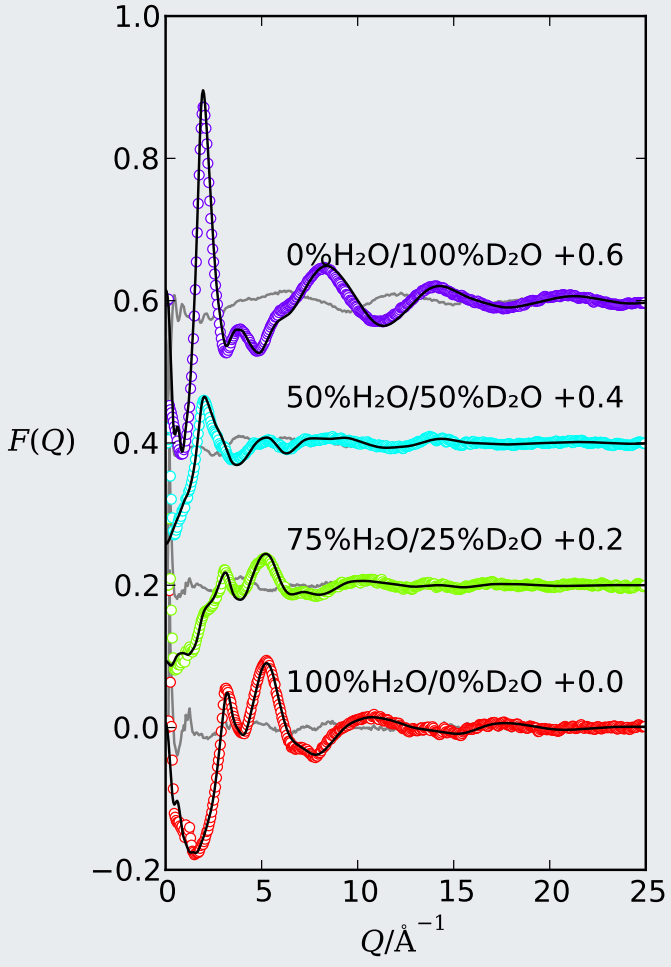
Even slightly larger molecules start to cause problems

glycyl-L-prolyl-glycinamide in water



- Protein folding model
- β-hairpin turn motif
- role of water in folding initiation?

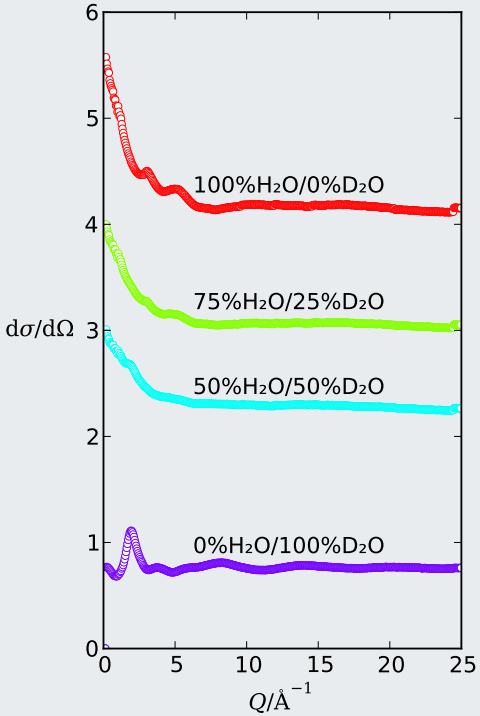
gpg-peptide in water



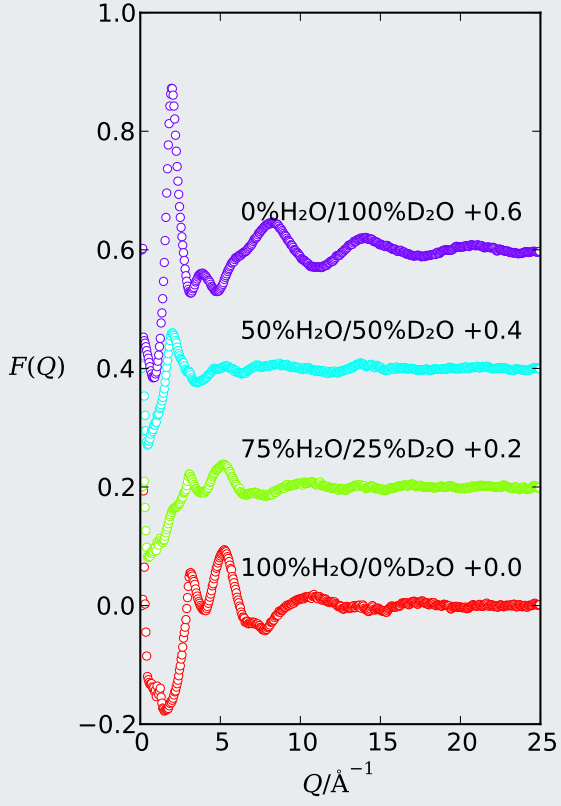
S Busch,* *et al.*, *manuscript submitted* (2013)

Liquid diffraction - what it really looks like

What you really measure



$$\frac{d\sigma}{d\Omega} = \frac{d\sigma}{d\Omega_s} + \frac{d\sigma}{d\Omega_d} = \sum_a c_\alpha b_\alpha^2 + P(Q,\theta) + F(Q)$$



DEPARTMENT OF BIOCHEMISTRY
UNIVERSITY OF OXFORD

gpg - EPSR fit to the data

- box of molecules at ρ, T, P of measurement
- reasonable R factor (Rf)

$$Rf = \frac{1}{M} \sum_i \left\{ \frac{\sum_Q (D_i(Q) - fit_i(Q))^2}{N_Q(i)} \right\}^*$$

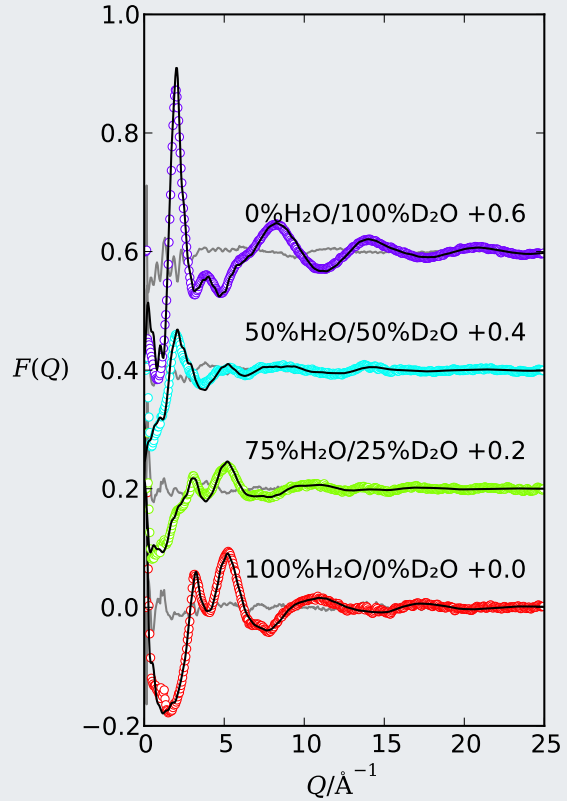
M - number of data sets

$D_i(Q)$ - data at point Q,

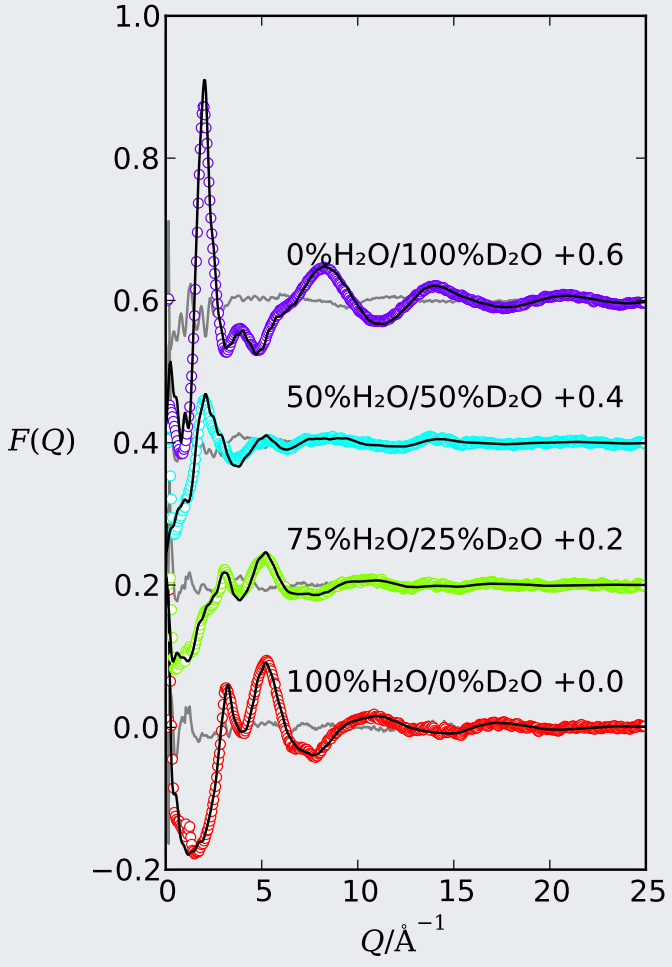
ith data set

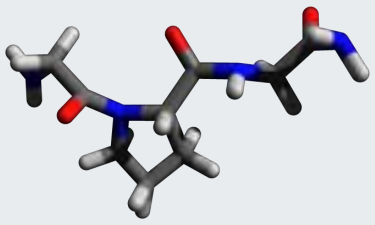
$N_Q(i)$ - number of points in Q

*deliberatly ignores statistical errors
systematic effects unknown



gpg - configuration



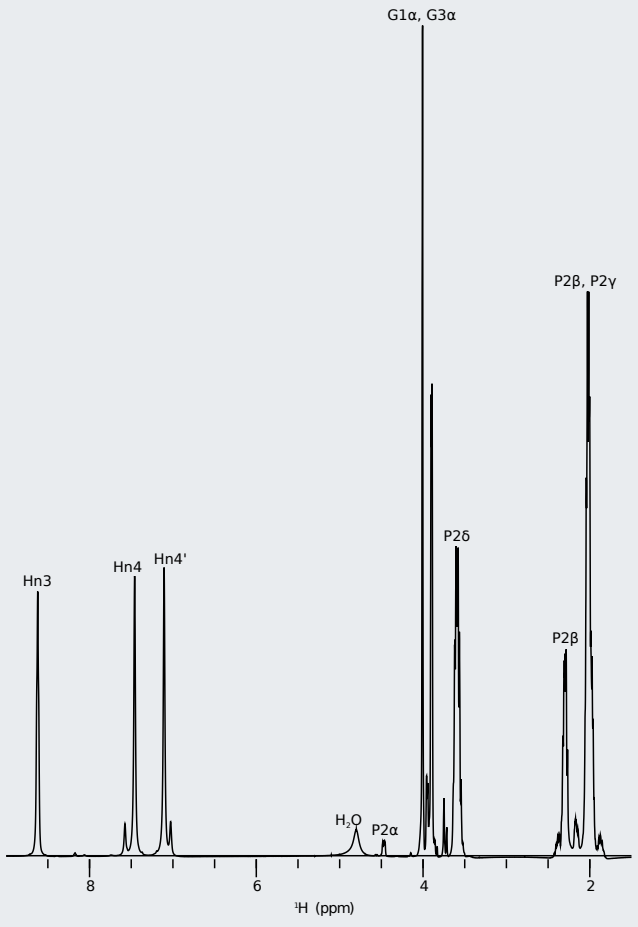


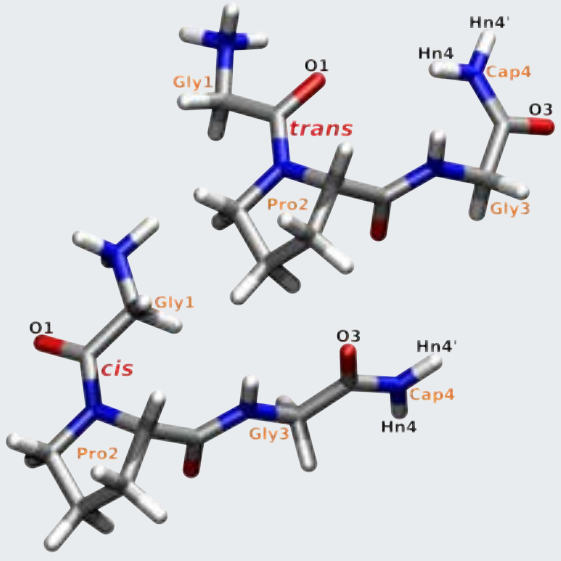
Peptide bonds not flat!

Need additional constraints

Need additional measurements

gpg - NMR





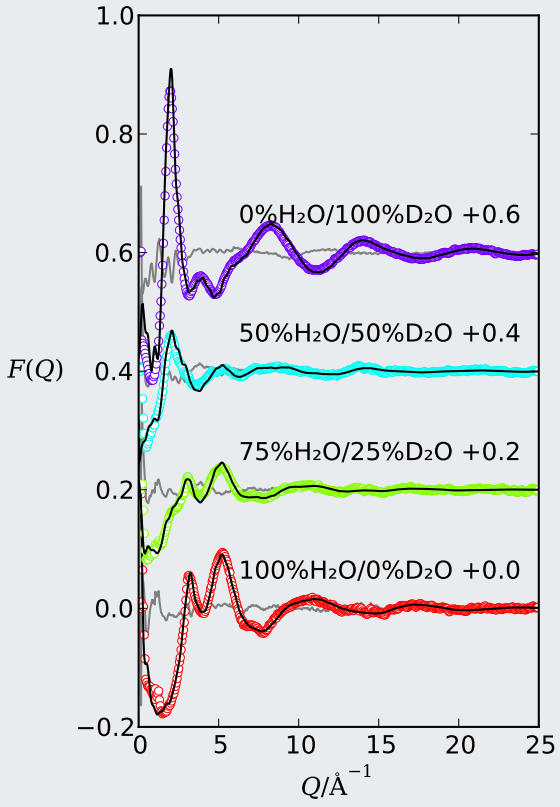
two average conformations in solution

cis vs trans with respect to the glycyl-prolyl bond

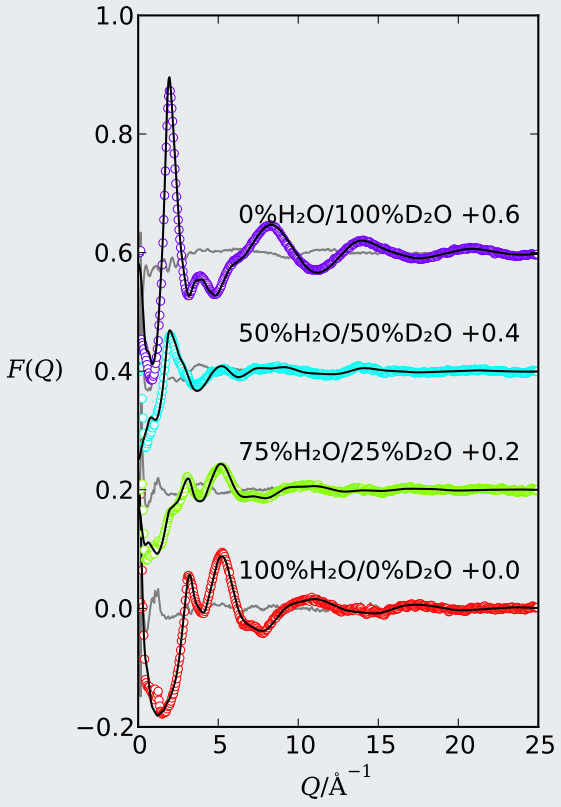
DEPARTMENT OF BIOCHEMISTRY
UNIVERSITY OF OXFORD

gpg - which fits are better?

before molecular constraint

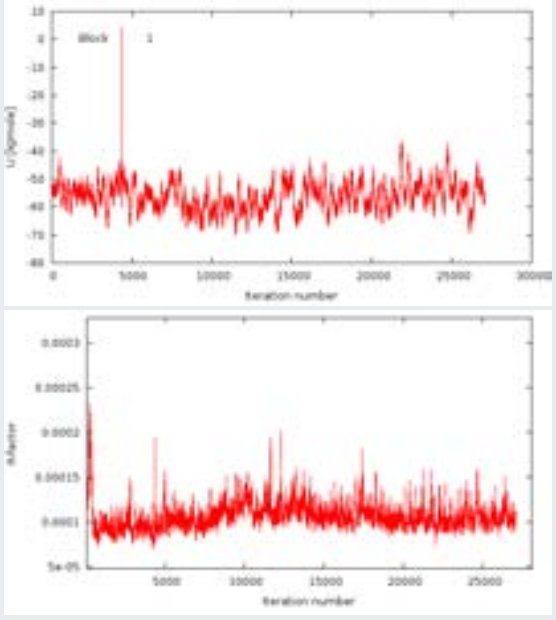


after molecular constraint

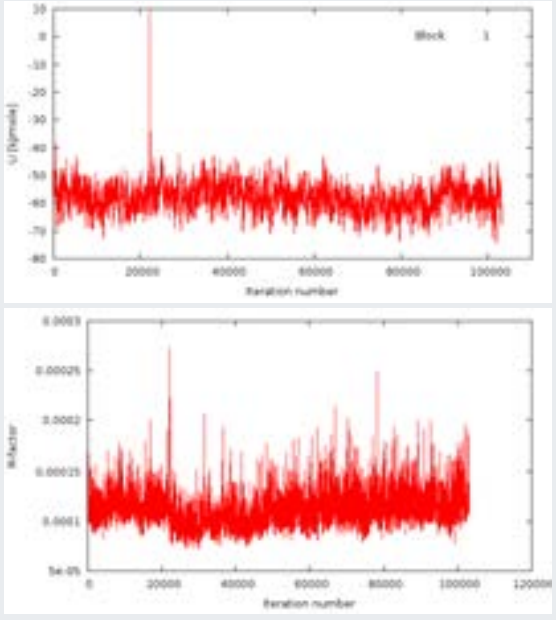


gpg - which fits are better?

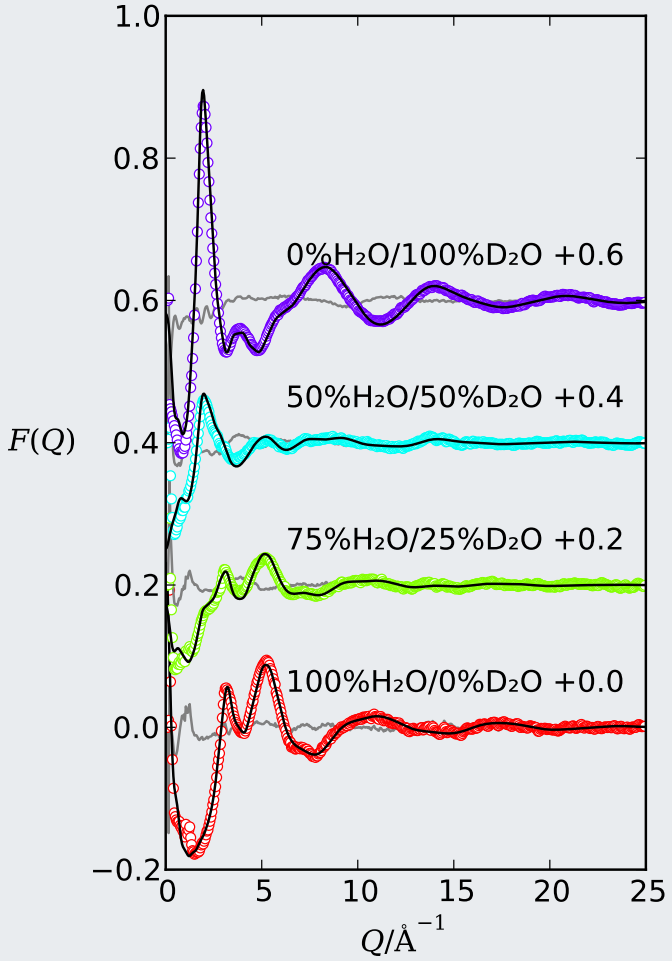
before molecular constraint

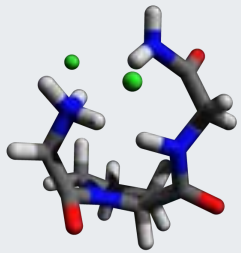


after molecular constraint



gpg - configuration

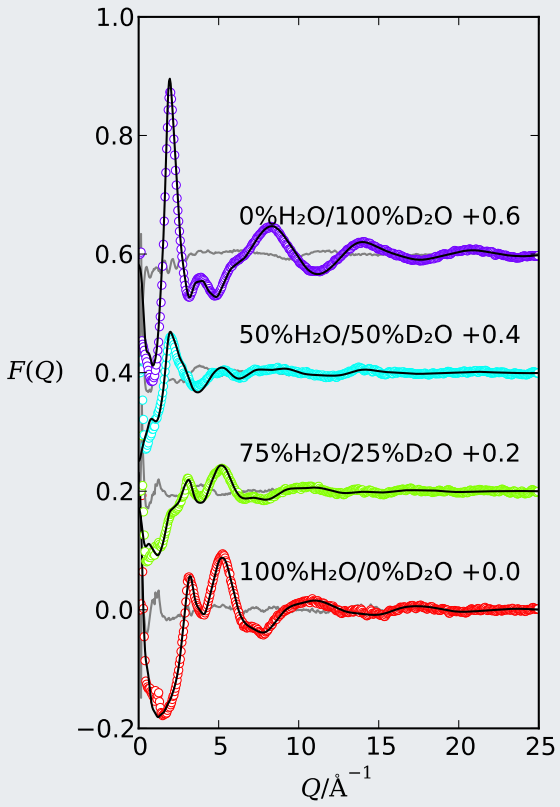




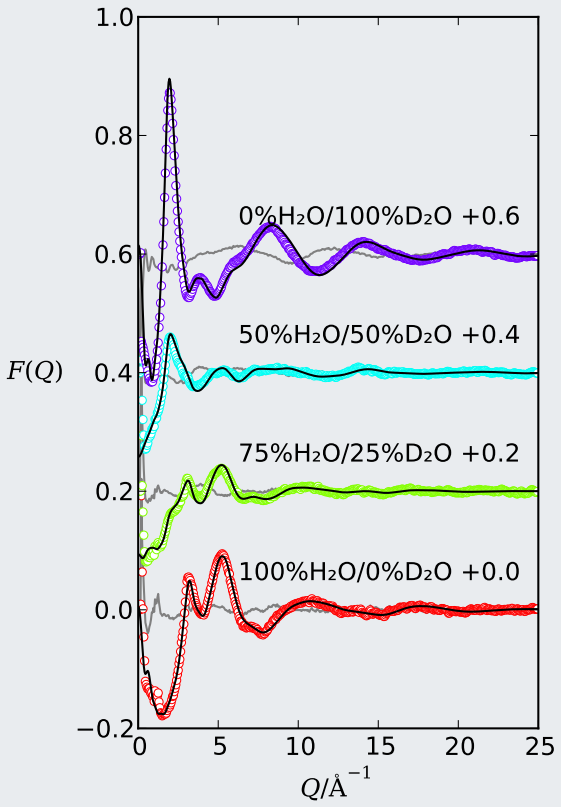
- unphysical distances
- Na^+ association ?
- hard to detect
- EPSR perfectly 'happy'

gpg - which fits are better?

before reduced charges



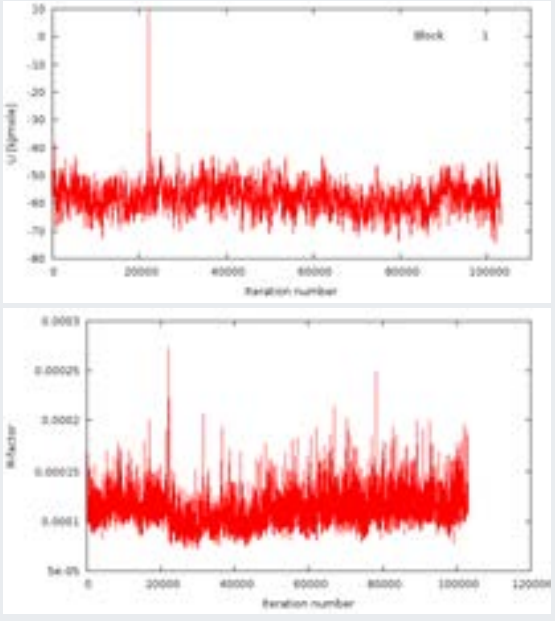
after reduced charges



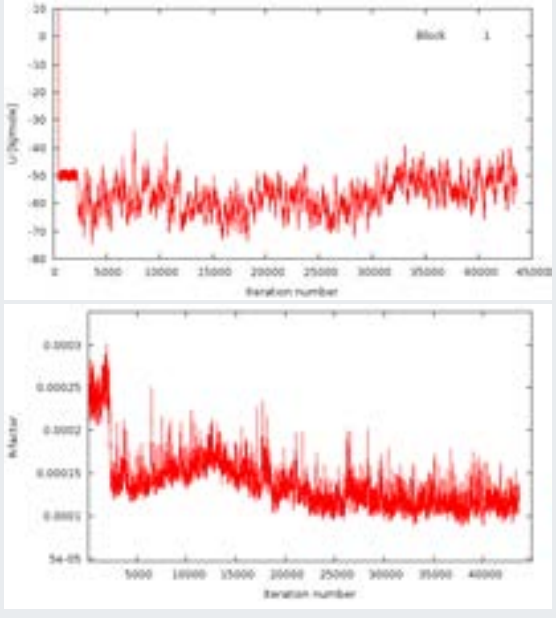
DEPARTMENT OF BIOCHEMISTRY
UNIVERSITY OF OXFORD

gpg - which fits are better?

before reduced charges



after reduced charges

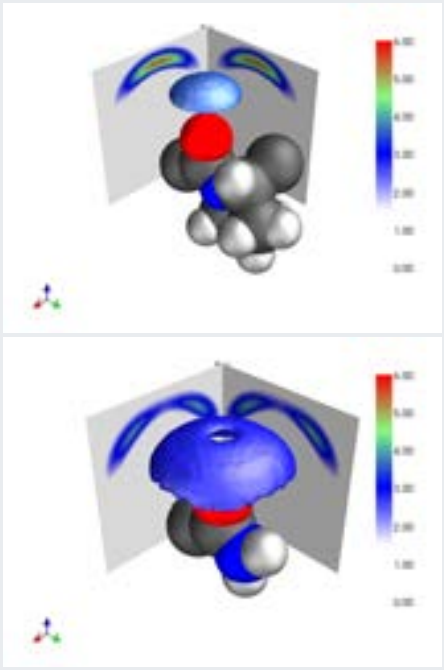


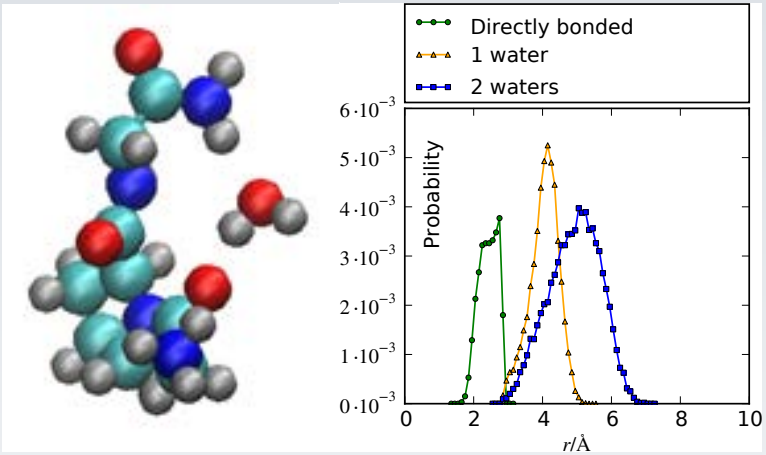
Challenge 5: Building consistent models

Liquid diffraction data on its own for complex systems not enough

- more neutron diffraction data sets (more isotopic substitutions)
- add X-ray diffraction
- NMR
- other simulation techniques - MD, DFT (other?)

water-mediated turning in gpg - the consistent result





consistent evidence

diffraction

NMR - chemical shifts

EPSR simulations

MD simulations

Challenge 6

Modeling constrained by more experimental data

- improving reverse Monte Carlo methods
- more data to be 'fit' such as NMR
- using potentials from EPSR to inform MD
- link structure with dynamics

Challenge 6b: Moving towards more complex systems

- boundary between polycrystalline systems and amorphous systems
- 'disordered refinement' - PDF for biomolecules
- Spanning from the Å scale to macromolecular scale
- New neutron instrumentation - NIMROD and SANS2d at ISIS (UK)

Acknowledgments

Oxford

Sebastian Busch
Andrew Johnston
Christina Redfeild

Alan Soper (ISIS)
Chris Lorenz (King's College London)

Funding

UK Engineering and Physical Sciences Research Council

Wellcome Trust

DEPARTMENT OF BIOCHEMISTRY
UNIVERSITY OF OXFORD