International Spring School Statistical Thermodynamics, Santiago de Chile
Tuesday, November 28, 2017
Lecture 17

On comparing simulated with experimental data

Prof. Dr. Wilfred F. van Gunsteren

ETH Zürich, Switzerland

On comparing molecular modelling results with experimental data

A. The experimental problem

1. Experimental data Q^{exp} are averages
2. Insufficient number of experimental data
3. Insufficient accuracy of experimental data
4. Experimental data Q^{exp} may be inconsistent

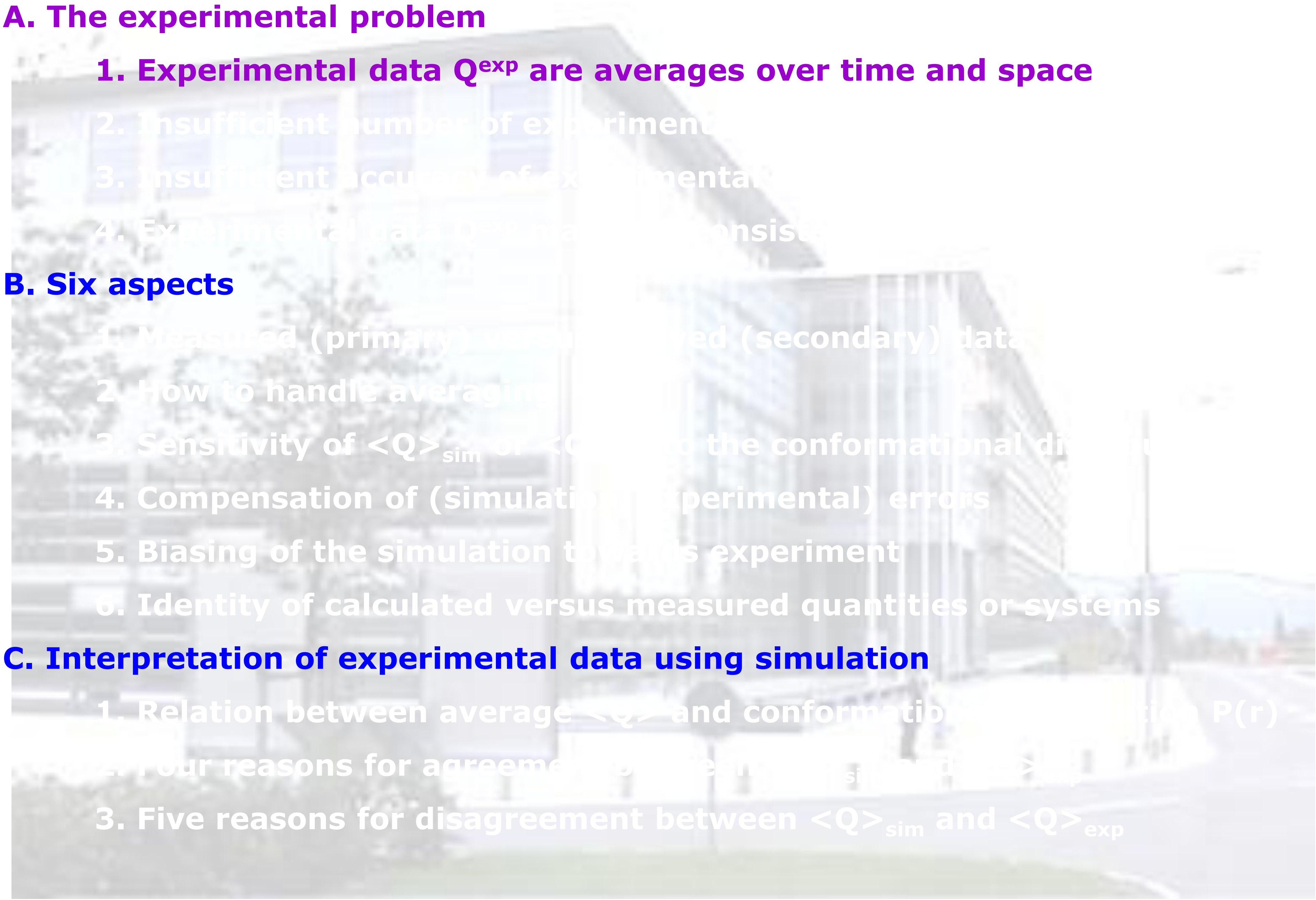
B. Six aspects

1. Measured (primary) versus derived (secondary) data
2. How to handle averaging
3. Sensitivity of $<Q>_{sim}$ or $<Q>_{exp}$ to the conformational distribution
4. Compensation of (simulation/experimental) errors
5. Biasing of the simulation towards experiment
6. Identity of calculated versus measured quantities or systems

C. Interpretation of experimental data using simulation

1. Relation between average $<Q>$ and conformational distribution P(r)
2. Four reasons for agreement between $<Q>_{sim}$ and $<Q>_{exp}$
3. Five reasons for disagreement between $<Q>_{sim}$ and $<Q>_{exp}$

On comparing molecular modelling results with experimental data



A. The experimental problem

1. Experimental data Q^{exp} are averages over time and space

2. Insufficient number of experimen[illegible]

3. Insufficient accuracy of experimental [illegible]

4. Experimental data Q^{exp} may be inconsistent

B. Six aspects

1. Measured (primary) versus derived (secondary) data

2. How to handle averaging

3. Sensitivity of $\langle Q\rangle_{sim}$ or $\langle Q\rangle_{exp}$ to the conformational distribution

4. Compensation of (simulation, experimental) errors

5. Biasing of the simulation towards experiment

6. Identity of calculated versus measured quantities or systems

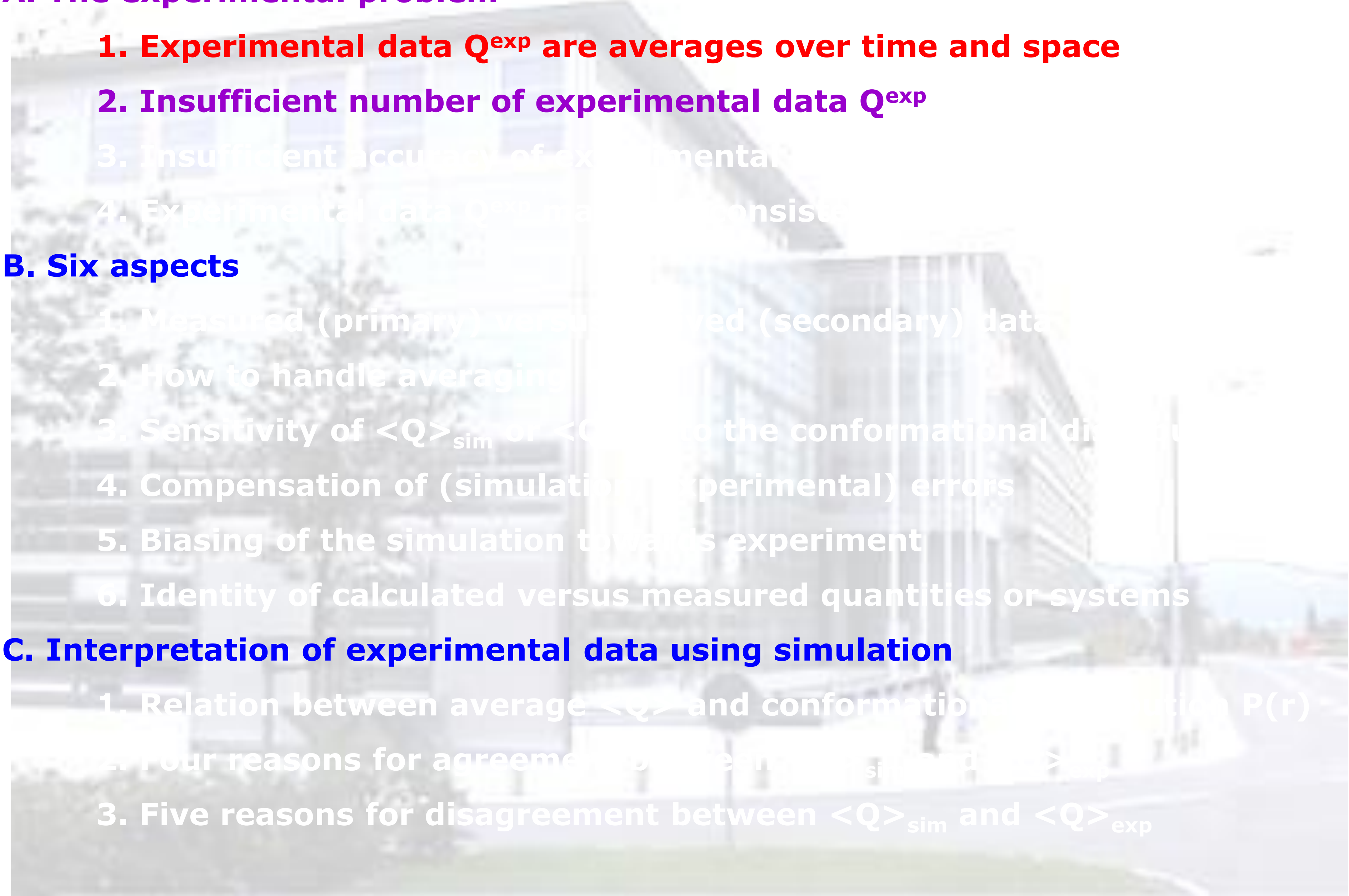
C. Interpretation of experimental data using simulation

1. Relation between average $\langle Q\rangle$ and conformational distribution P(r)

2. Four reasons for agreement between $\langle Q\rangle_{sim}$ and $\langle Q\rangle_{exp}$

3. Five reasons for disagreement between $\langle Q\rangle_{sim}$ and $\langle Q\rangle_{exp}$

On comparing molecular modelling results with experimental data



A. The experimental problem

1. Experimental data Q^{exp} are averages over time and space
2. Insufficient number of experimental data Q^{exp}
3. Insufficient accuracy of experimental data Q^{exp}
4. Experimental data Q^{exp} may be inconsistent

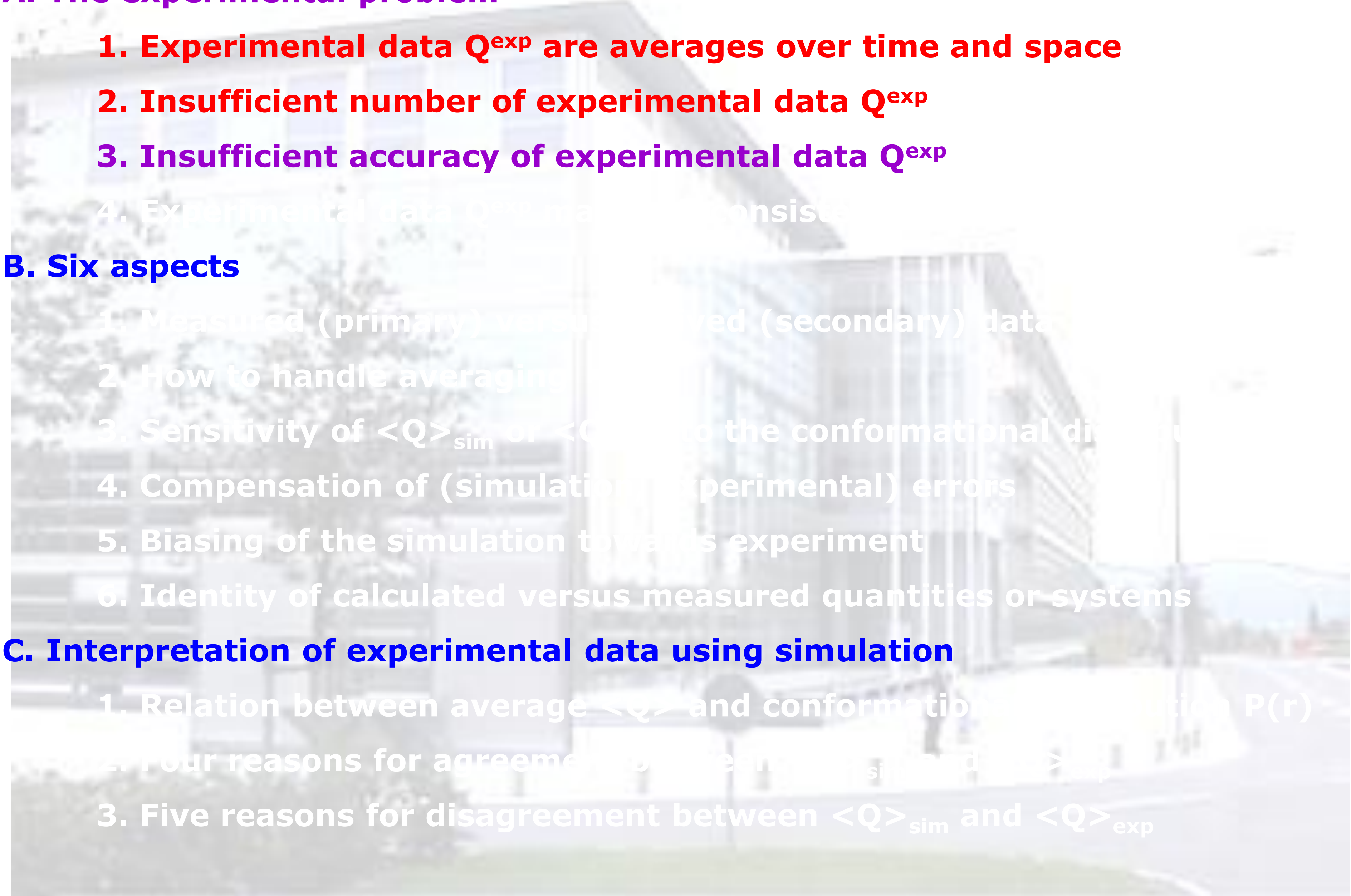
B. Six aspects

1. Measured (primary) versus derived (secondary) data
2. How to handle averaging
3. Sensitivity of $<Q>_{sim}$ or $<Q>_{exp}$ to the conformational distribution
4. Compensation of (simulation or experimental) errors
5. Biasing of the simulation towards experiment
6. Identity of calculated versus measured quantities or systems

C. Interpretation of experimental data using simulation

1. Relation between average $<Q>$ and conformational distribution P(r)
2. Four reasons for agreement between $<Q>_{sim}$ and $<Q>_{exp}$
3. Five reasons for disagreement between $<Q>_{sim}$ and $<Q>_{exp}$

On comparing molecular modelling results with experimental data



A. The experimental problem

1. Experimental data Q^{exp} are averages over time and space
2. Insufficient number of experimental data Q^{exp}
3. Insufficient accuracy of experimental data Q^{exp}
4. Experimental data Q^{exp} may be inconsistent

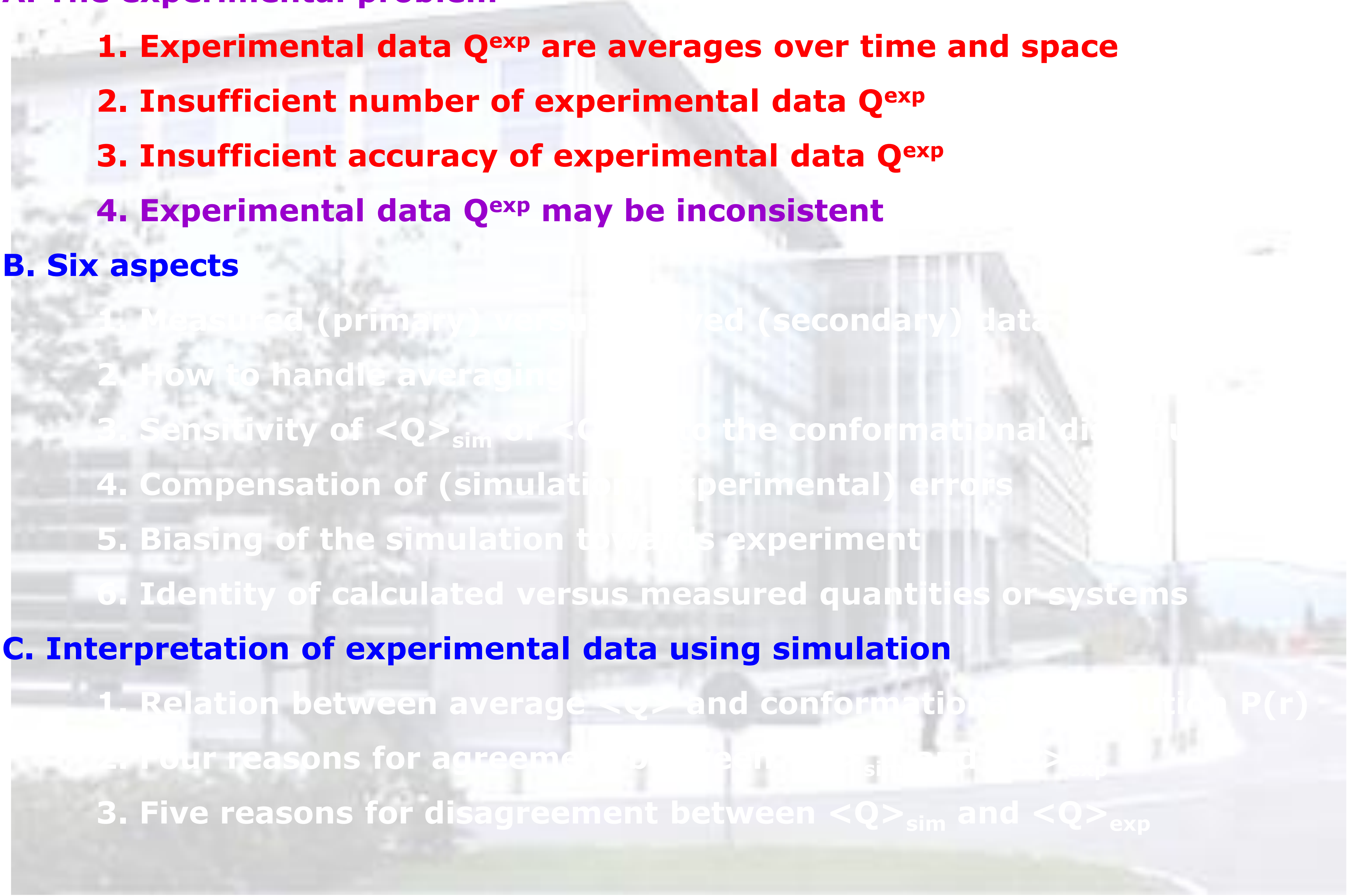
B. Six aspects

1. Measured (primary) versus derived (secondary) data
2. How to handle averaging
3. Sensitivity of $<Q>_{sim}$ or $<Q>_{exp}$ to the conformational distribution
4. Compensation of (simulation, experimental) errors
5. Biasing of the simulation towards experiment
6. Identity of calculated versus measured quantities or systems

C. Interpretation of experimental data using simulation

1. Relation between average $<Q>$ and conformational distribution P(r)
2. Four reasons for agreement between $<Q>_{sim}$ and $<Q>_{exp}$
3. Five reasons for disagreement between $<Q>_{sim}$ and $<Q>_{exp}$

On comparing molecular modelling results with experimental data



A. The experimental problem

1. Experimental data Q^{exp} are averages over time and space
2. Insufficient number of experimental data Q^{exp}
3. Insufficient accuracy of experimental data Q^{exp}
4. Experimental data Q^{exp} may be inconsistent

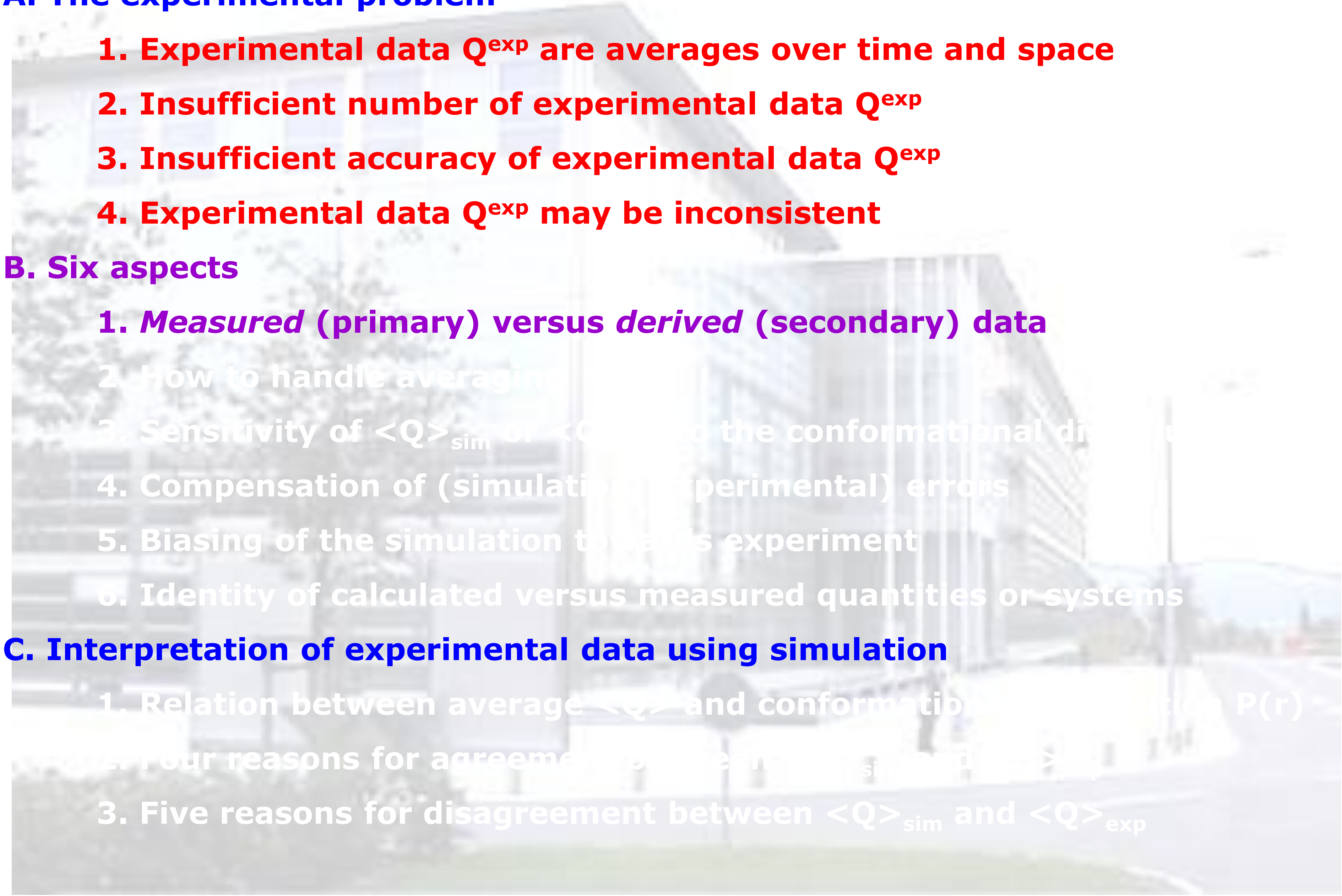
B. Six aspects

1. Measured (primary) versus derived (secondary) data
2. How to handle averaging
3. Sensitivity of $<Q>_{sim}$ or $<Q>_{exp}$ to the conformational distribution
4. Compensation of (simulation, experimental) errors
5. Biasing of the simulation towards experiment
6. Identity of calculated versus measured quantities or systems

C. Interpretation of experimental data using simulation

1. Relation between average $<Q>$ and conformational distribution P(r)
2. Four reasons for agreement between $<Q>_{sim}$ and $<Q>_{exp}$
3. Five reasons for disagreement between $<Q>_{sim}$ and $<Q>_{exp}$

On comparing molecular modelling results with experimental data



A. The experimental problem

1. Experimental data Q^{exp} are averages over time and space
2. Insufficient number of experimental data Q^{exp}
3. Insufficient accuracy of experimental data Q^{exp}
4. Experimental data Q^{exp} may be inconsistent

B. Six aspects

1. *Measured* (primary) versus *derived* (secondary) data
2. How to handle averaging
3. Sensitivity of $<Q>_{sim}$ or $<Q>_{exp}$ to the conformational distribution
4. Compensation of (simulation, experimental) errors
5. Biasing of the simulation towards experiment
6. Identity of calculated versus measured quantities or systems

C. Interpretation of experimental data using simulation

1. Relation between average $<Q>$ and conformational distribution P(r)
2. Four reasons for agreement between $<Q>_{sim}$ and $<Q>_{exp}$
3. Five reasons for disagreement between $<Q>_{sim}$ and $<Q>_{exp}$

Comparison with Experimental Data for a Quantity Q

$\langle Q \rangle_{sim} \quad \leftrightarrow \quad \langle Q \rangle_{exp}$

W.F. van Gunsteren, J. Dolenc, &
A.E. Mark, Curr. Opin. Struct. Biol.,
*18 (**2008**) 149-153*

Distinguish between:

1. primary experimental data $\mathbf{Q^{measured}}$: *observable* quantities Q that are ***directly measured***

   Examples: peak location and intensity from X-ray diffraction or NMR spectroscopic measurements (a.o. ^{3}J–values)

2. secondary (***derived using a model***) "experimental" data $\mathbf{Q^{derived}}$: quantities Q for which (*non-observed*) values are ***derived from*** (observed) values of primary experimental data $\mathbf{Q^{measured}}$ by applying a particular **procedure f**: $\mathbf{Q^{derived} = f\,(Q^{measured})}$ which involves assumptions and approximations

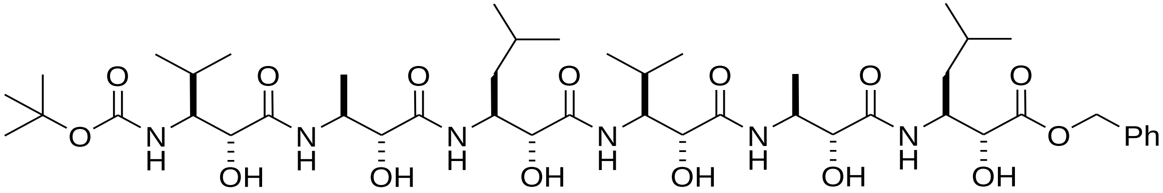
   Examples: *molecular structures* (a.o. torsional angle values) NMR order parameters

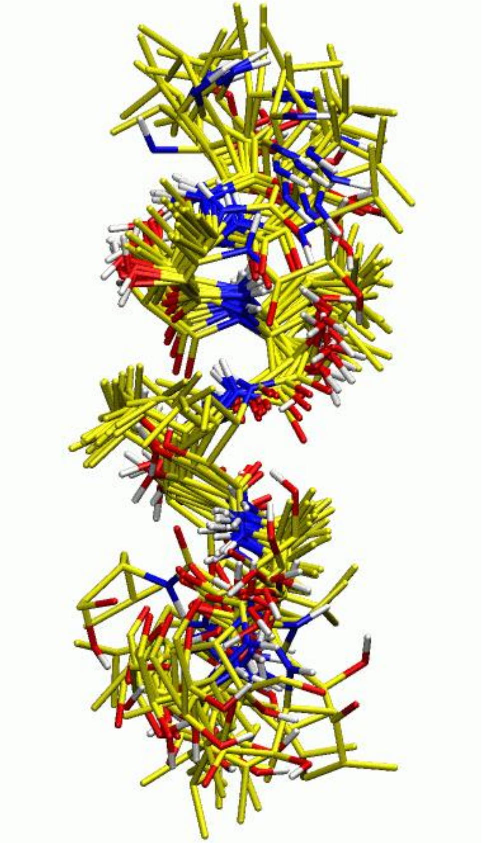
Comparison of		may reflect the quality of:
a. $\langle Q^{measured} \rangle_{sim}$	with $\langle Q^{measured} \rangle_{exp}$	the simulation
b. $\langle Q^{derived} \rangle_{sim}$	with $\langle Q^{derived} \rangle_{exp} = f\,(\langle Q^{measured} \rangle_{exp})$	the procedure f

In reality $\langle Q^{derived} \rangle_{exp}$ may carry little experimental information

A β-hexapeptide

Two non-overlapping conformational ensembles reproduce the experimental data: Which one is realistic?





- β-hexapeptide with hydroxyl groups attached to the α-carbons
- NMR single-structure refinement *based on NOE and* ^{3}*J-coupling data* suggests the formation of a 2_8-***P***-helix
- MD simulation from totally extended conformation at two different temperatures (298 K & 340K) using the GROMOS 45A3 force field *without any NOE-distance or* ^{3}*J-value restraining* suggests the formation of a 2.5_{12}-***P***-helix

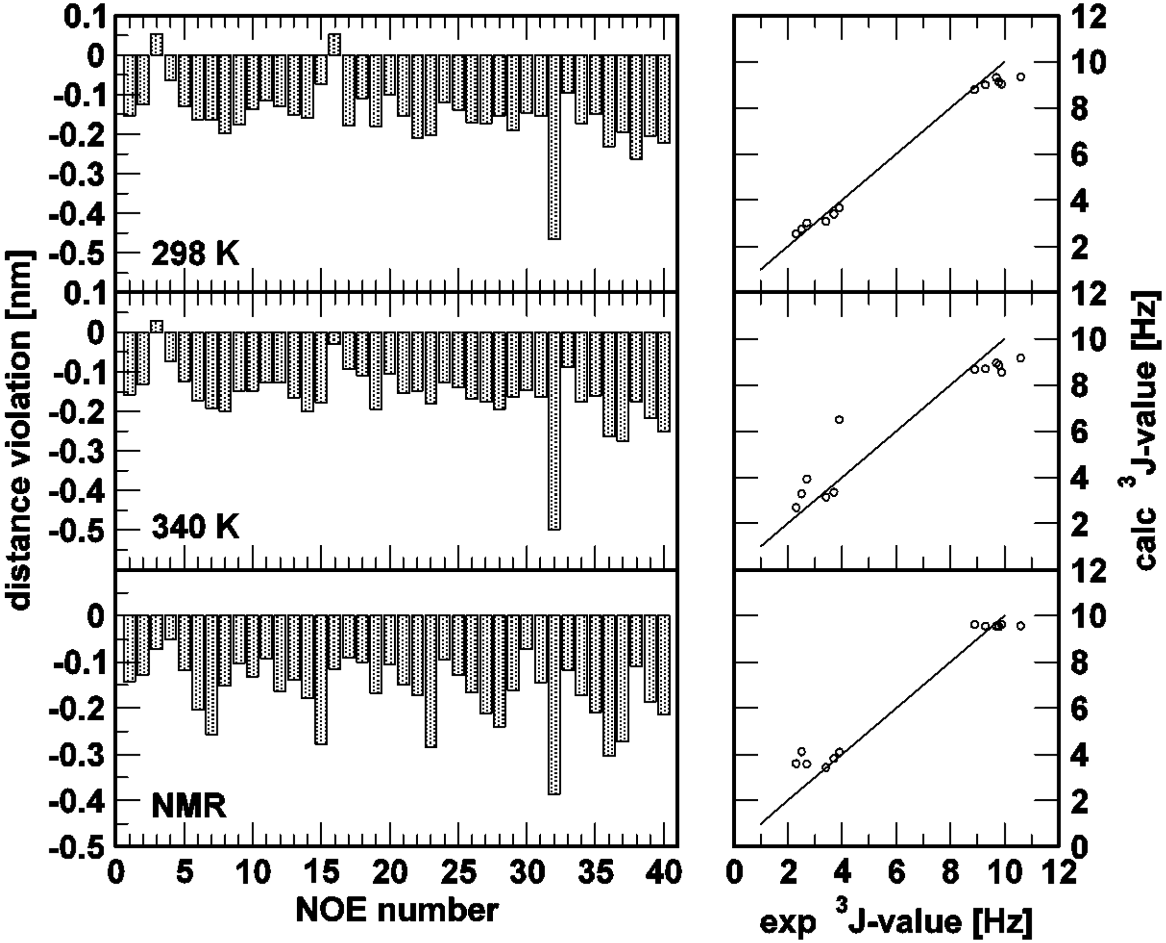
Bundle of 20 NMR model structures

(protection groups not shown)

*Glaettli & van Gunsteren, Angew. Chem. Int. Ed. Engl. 43 (**2004**) 6312*

*Gademann et al., Angew. Chem. Int. Ed. Engl. 42 (**2003**) 1534*

NOE Distance Violations & Backbone ^{3}J-values



- MD at 298 K
  2 violations (~0.05 nm)

  average deviation from exp. J-values: 0.44 Hz

- MD at 340 K
  1 violation (~ 0.03 nm)

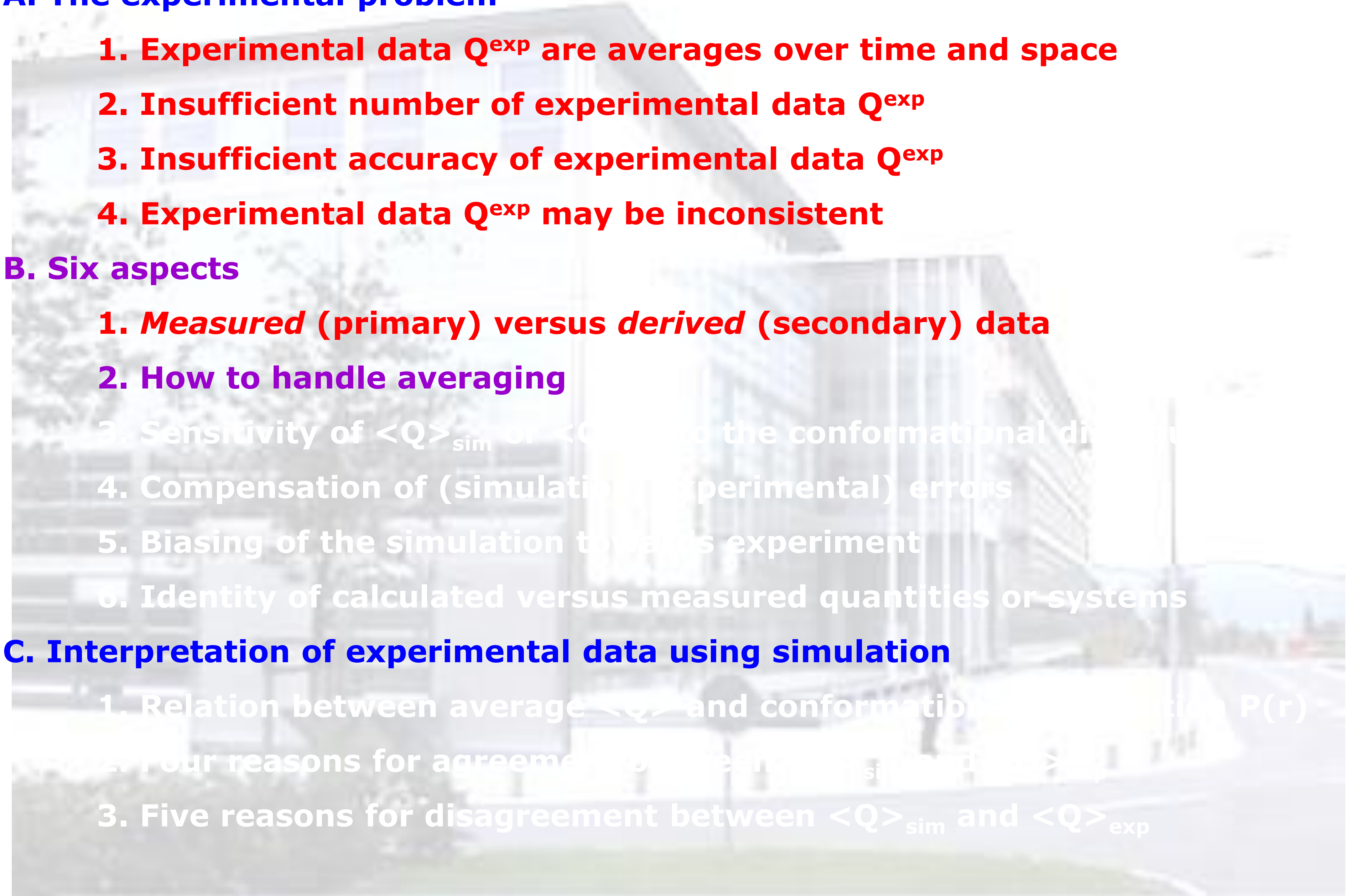
  average deviation from exp. J-values: 0.91 Hz

- NMR set of structures
  no violation (0.0 nm)

  average deviation from exp. J-values: 0.57 Hz

**Two different methods to derive a set of peptide structures produce non-overlapping ensembles that each reproduce the *measured* data.
However MD simulation (*ensemble*) predicts a well known 2.5_{12}-helix, whereas the NMR *single-structure* refinement predicts an unknown 2_8-helix**

On comparing molecular modelling results with experimental data



A. The experimental problem

1. Experimental data Q^{exp} are averages over time and space
2. Insufficient number of experimental data Q^{exp}
3. Insufficient accuracy of experimental data Q^{exp}
4. Experimental data Q^{exp} may be inconsistent

B. Six aspects

1. *Measured* (primary) versus *derived* (secondary) data
2. How to handle averaging
3. Sensitivity of $<Q>_{sim}$ or $<Q>_{exp}$ to the conformational distribution
4. Compensation of (simulation, experimental) errors
5. Biasing of the simulation towards experiment
6. Identity of calculated versus measured quantities or systems

C. Interpretation of experimental data using simulation

1. Relation between average $<Q>$ and conformational distribution P(r)
2. Four reasons for agreement between $<Q>_{sim}$ and $<Q>_{exp}$
3. Five reasons for disagreement between $<Q>_{sim}$ and $<Q>_{exp}$

The Molecular Modelling Approach

How to calculate a quantity or observable $Q(\vec{r})$?

Choose:

1. (essential) degrees of freedom $\vec{r}$

   ↳ for $Q(\vec{r})$ electronic
   atomic
   solvent

2. interaction function $V^{phys}(\vec{r})$

   between degrees of freedom (force field, e.g. GROMOS)

3. equations of motion or sampling method

   to generate a Boltzmann-weighted ensemble of conformers:

   probability $P(\vec{r}) = \exp(-V^{phys}(\vec{r})/k_B T) / \int \exp(-V^{phys}(\vec{r})/k_B T)\ d\vec{r}$

4. function $Q(\vec{r})$ (contains approximations and assumptions)

Ensemble averages

$$\langle Q \rangle_{\vec{r}} \equiv \int Q(\vec{r}) P(\vec{r}) d\vec{r}$$

are to be compared:

$$\langle Q \rangle_{sim}$$

is to be compared to

$$\langle Q \rangle_{exp} \equiv Q^{exp}$$

If

1. $V^{phys}(\vec{r})$ and $Q(\vec{r})$ ***are correct***
2. ***infinite*** sampling

problem solved

Otherwise

make other choices and repeat

Effects of Ensemble (Motional) Averaging

$$\left\langle Q(\vec{r})\right\rangle_{\vec{r}} \equiv \int Q(\vec{r})\, P(\vec{r}) d\vec{r} \equiv \frac{\int Q(r) e^{-V^{phys}(\vec{r})/k_B T}\, d\vec{r}}{\int e^{-V^{phys}(\vec{r})/k_B T}\, d\vec{r}}$$

$$\left\langle Q(\vec{r})\right\rangle_{\vec{r}} \neq Q(\vec{r})$$

averaged Q — single structure Q

$$\neq Q\left(\left\langle \vec{r}\right\rangle_{\vec{r}}\right)$$

mean structure Q

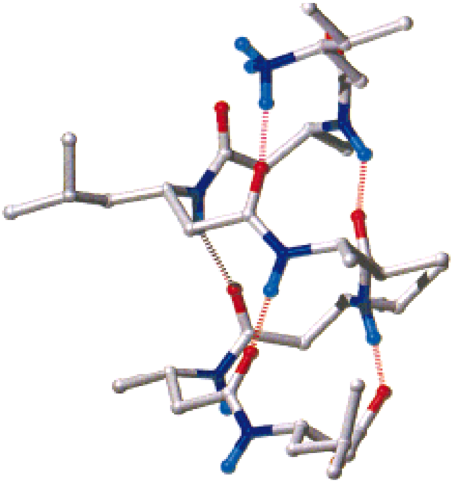
Examples of (*observable*) quantities *Q(r)*:

- NOE intensities (NMR)
- ^{3}J-coupling constants (NMR)
- Residual dipolar couplings (NMR)
- Chemical shifts (NMR)
- Structure factors (amplitudes) (X-ray)
- CD spectra (CD)
- ……..

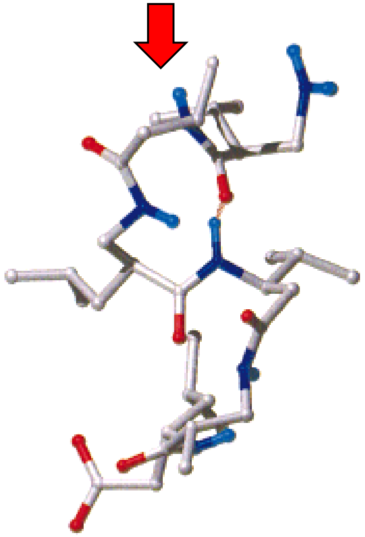
Boltzmann weighting is non-linear
Function Q(r) may be non-linear

Effect of Ensemble (Motional) Averaging

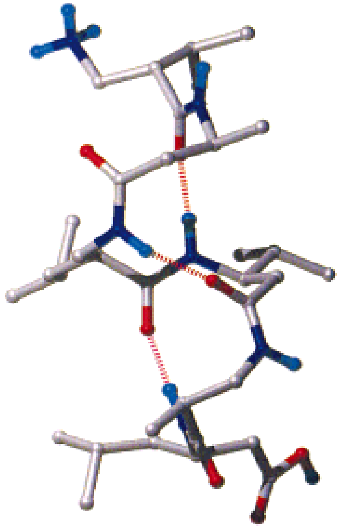
The average structure $<\mathbf{r}>_r$ is highly strained
for a 6-ß-peptide in methanol: 34 NOE's



left-handed **3_{14}-helix**
of a similar peptide in MeOH
H-bonds: NH(i) – O(i+2)



average structure:
distorted right-handed helix in MeOH only one
H-bond: NH(4) – O(1)



right-handed **helix in pyridine**
H-bonds: NH(i) – O(i+1, i-3)

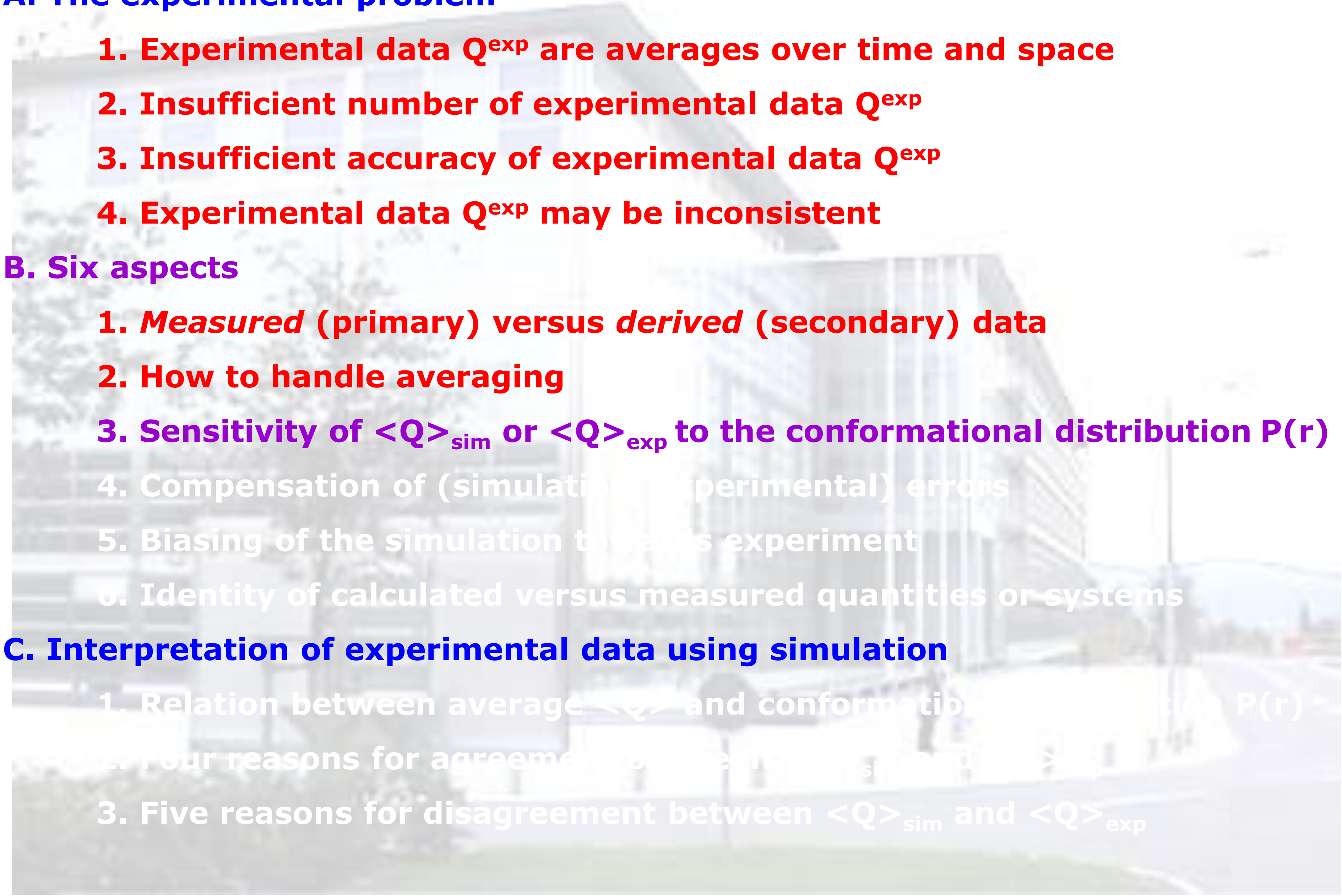
due to 3 NOE's characteristic → not observed in pyridine
for a left–handed 3_{14}-helix

MD simulation: - satisfies all NOE bounds
- shows 35% *right*-handed helix
1.3% *left*-handed helix

Conclusion: - average structure may be meaningless
- use ***primary*** (***observed***) exp. data (NOE's),
not secondary (***derived***) exp. data (structures) ***to compare with***

*X. Daura et al., Angew. Chem. Int. Ed. 38 (**1999**) 236-240*

On comparing molecular modelling results with experimental data



A. The experimental problem

1. Experimental data Q^{exp} are averages over time and space
2. Insufficient number of experimental data Q^{exp}
3. Insufficient accuracy of experimental data Q^{exp}
4. Experimental data Q^{exp} may be inconsistent

B. Six aspects

1. *Measured* (primary) versus *derived* (secondary) data
2. How to handle averaging
3. Sensitivity of $<Q>_{sim}$ or $<Q>_{exp}$ to the conformational distribution P(r)
4. Compensation of (simulation, experimental) errors
5. Biasing of the simulation towards experiment
6. Identity of calculated versus measured quantities or systems

C. Interpretation of experimental data using simulation

1. Relation between average <Q> and conformational distribution P(r)
2. Four reasons for agreement between $<Q>_{sim}$ and $<Q>_{exp}$
3. Five reasons for disagreement between $<Q>_{sim}$ and $<Q>_{exp}$

Interpretation of Experimental Data using Simulation

Relation between average $\langle Q \rangle$ and conformational distribution $P(\vec{r})$

When relating the *average* of a property over a given conformational distribution P(r), whether from a simulation (<Q>$_{sim}$) or measured experimentally (<Q>$_{exp}$), to the conformational distribution itself, ***three general cases*** can be distinguished:

Q1 <Q> does not reflect the shape of P(**r**) as <Q> is ***insensitive*** to conformation

Q2 <Q> does not reflect the shape of P(**r**) as <Q> is determined by ***rarely sampled*** conformations with small (*irrelevant*) Boltzmann weights

Q3 <Q> ***does reflect*** the dominant conformations of P(**r**)

Only in case **Q3** can <Q>$_{sim}$ carry information relevant to the interpretation of <Q>$_{exp}$ at a molecular level

*W.F. van Gunsteren, J. Dolenc, & A.E. Mark, Curr. Opin. Struct. Biol., 18 (**2008**) 149-153*

Interpretation of Experimental Data using Simulation

Relation between average $\langle Q \rangle$ and conformational distribution $P(\vec{r})$

When relating the *average* of a property over a given conformational distribution P(r), whether from a simulation (<Q>$_{sim}$) or measured experimentally (<Q>$_{exp}$), to the conformational distribution itself, ***three general cases*** can be distinguished:

Q1 <Q> does not reflect the shape of P(**r**) as <Q> is ***insensitive*** to conformation

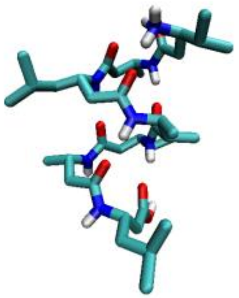
Q2 <Q> does not reflect the shape of P(**r**) as <Q> is determined by ***rarely sampled*** conformations with small (*irrelevant*) Boltzmann weights

Q3 <Q> ***does reflect*** the dominant conformations of P(**r**)

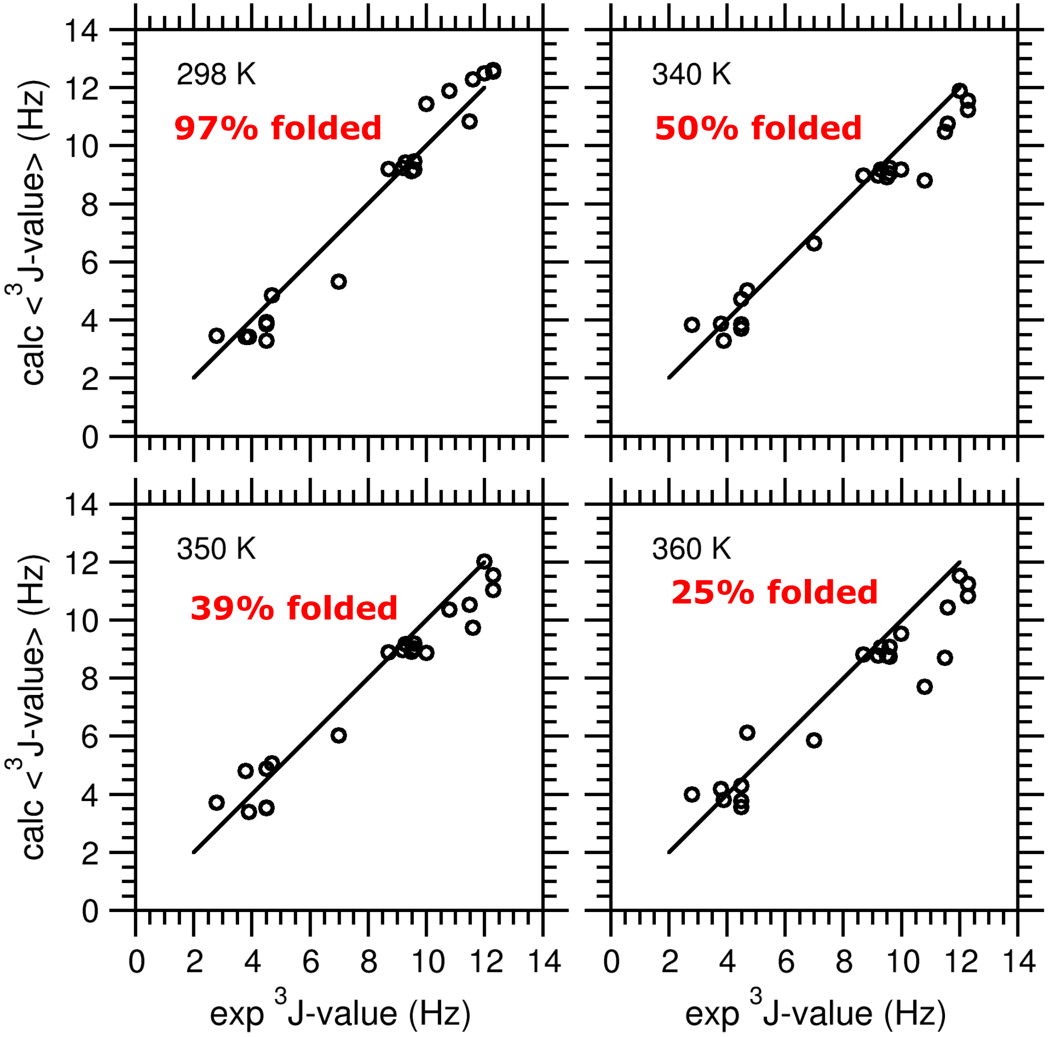
Only in case **Q3** can <Q>$_{sim}$ carry information relevant to the interpretation of <Q>$_{exp}$ at a molecular level

*W.F. van Gunsteren, J. Dolenc, & A.E. Mark, Curr. Opin. Struct. Biol., 18 (**2008**) 149-153*

Different Ensembles of a 7-β-peptide in solution



3_{14}-L-helix



$^3J(H_N\text{-}H_{\alpha\text{ or }\beta})$-couplings are *insensitive* to the conformational distribution

*X. Daura et al., Proteins 36 (**1999**) 542-555*

Interpretation of Experimental Data using Simulation

Relation between average $\langle Q \rangle$ and conformational distribution $P(\vec{r})$

When relating the *average* of a property over a given conformational distribution P(r), whether from a simulation ($<Q>_{sim}$) or measured experimentally ($<Q>_{exp}$), to the conformational distribution itself, ***three general cases*** can be distinguished:

Q1 <Q> does not reflect the shape of P(**r**) as <Q> is ***insensitive*** to conformation

Q2 <Q> does not reflect the shape of P(**r**) as <Q> is determined by ***rarely sampled*** conformations with small (*irrelevant*) Boltzmann weights

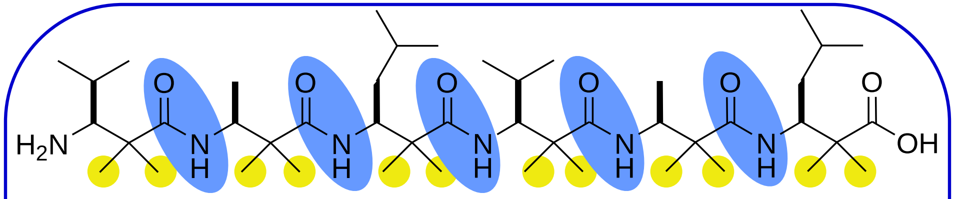
Q3 <Q> ***does reflect*** the dominant conformations of P(**r**)

Only in case **Q3** can $<Q>_{sim}$ carry information relevant to the interpretation of $<Q>_{exp}$ at a molecular level

*W.F. van Gunsteren, J. Dolenc, & A.E. Mark, Curr. Opin. Struct. Biol., 18 (**2008**) 149-153*

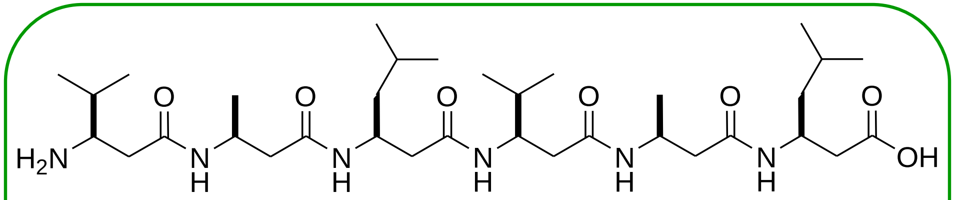
Calculation of Circular Dichroism (CD) Spectra

Two molecules with *similar* CD spectra, but cannot have a *similar* dominant stucture



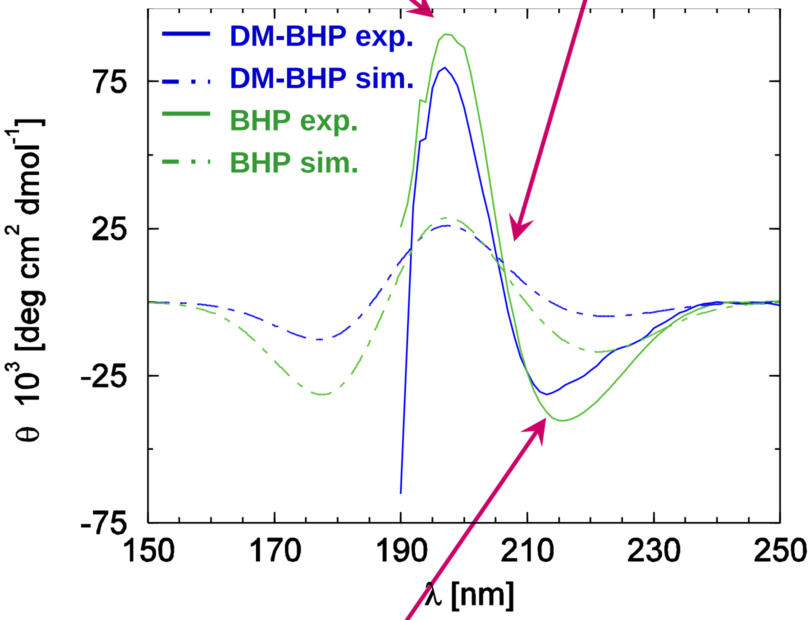
Peptide A: DM-BHP *(methyls (yellow))*

- geminal dimethylation inhibits the formation of a 3_{14} helix
- no NMR data available
- CD spectrum shows a pattern, which is “typical” for a 3_{14} helix



Peptide B: BHP *(no methyls)*

- can adopt a 3_{14} helix, confirmed by NMR experiments, CD spectrum similar



positive Cotton effect at ~200 nm

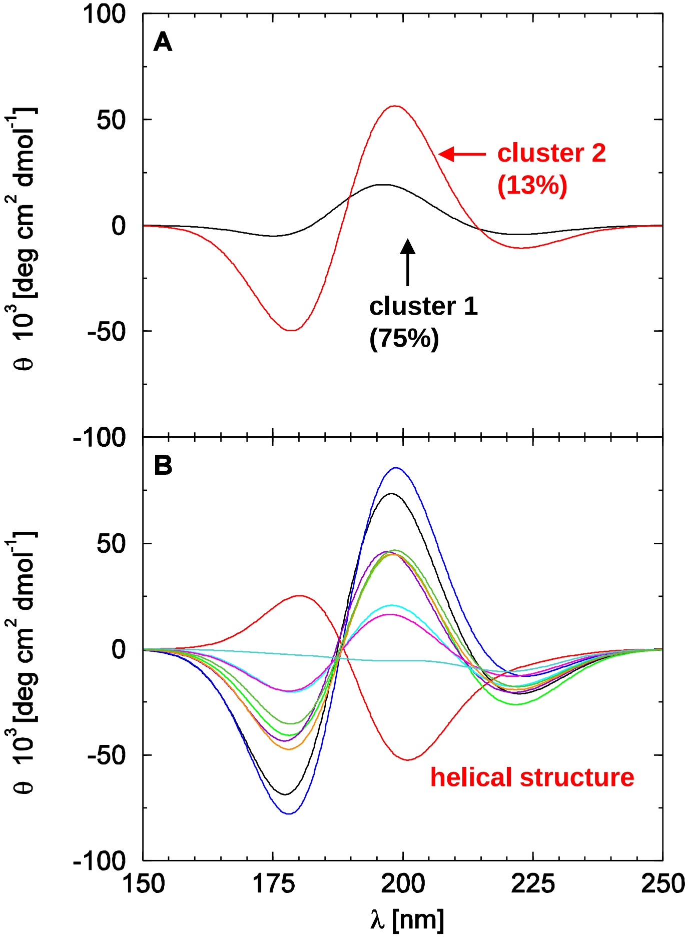
zero crossing between 205 and 210 nm

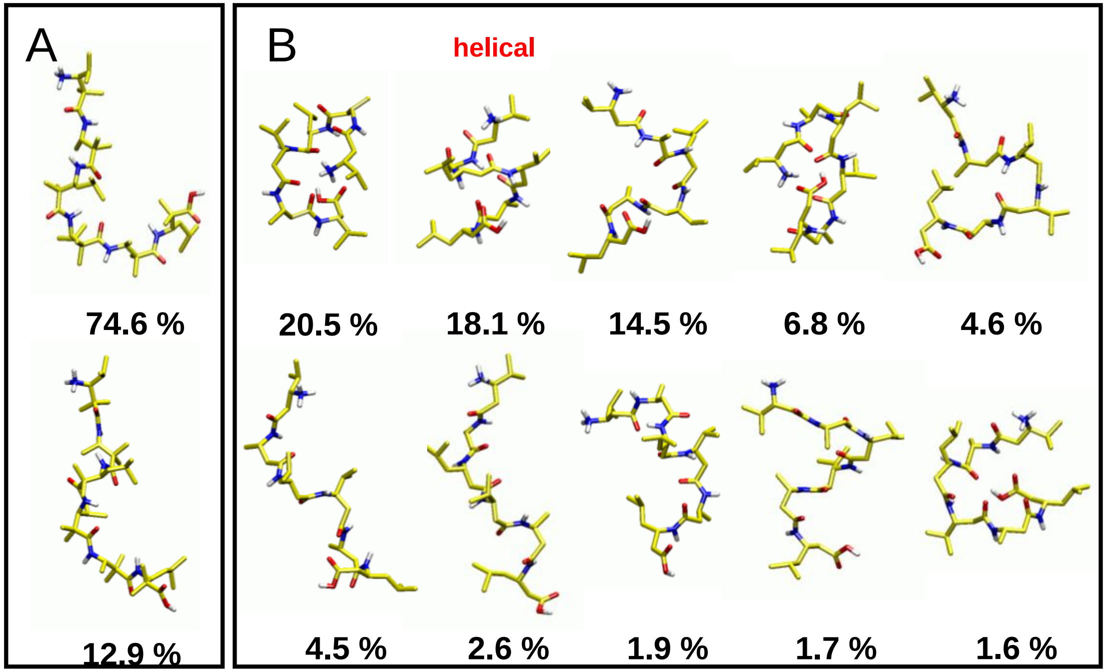
negative Cotton effect between 215 and 220 nm

*A. Glaettli et al., JACS 124 (**2002**) 12972-12978*

CD Spectra per Conformational Cluster

Similarity criterion: backbone RMSD ≤ 0.09nm
10000 structures, 10 psec apart





Non-helical conformers exhibit the CD pattern assigned to the 3_{14} helix, the “helical” conformer doesn’t.

→ virtually NO OVERLAP between the conformational ensembles of both molecules, which have similar CD spectra !

→ spectrum *not representative* for the dominant conformation !

*A. Glaettli et al., JACS 124 (**2002**) 12972-12978*

Interpretation of Experimental Data using Simulation

Relation between average $\langle Q \rangle$ and conformational distribution $P(\vec{r})$

When relating the *average* of a property over a given conformational distribution P(r), whether from a simulation ($<Q>_{sim}$) or measured experimentally ($<Q>_{exp}$), to the conformational distribution itself, ***three general cases*** can be distinguished:

Q1 <Q> does not reflect the shape of P(**r**) as <Q> is ***insensitive*** to conformation

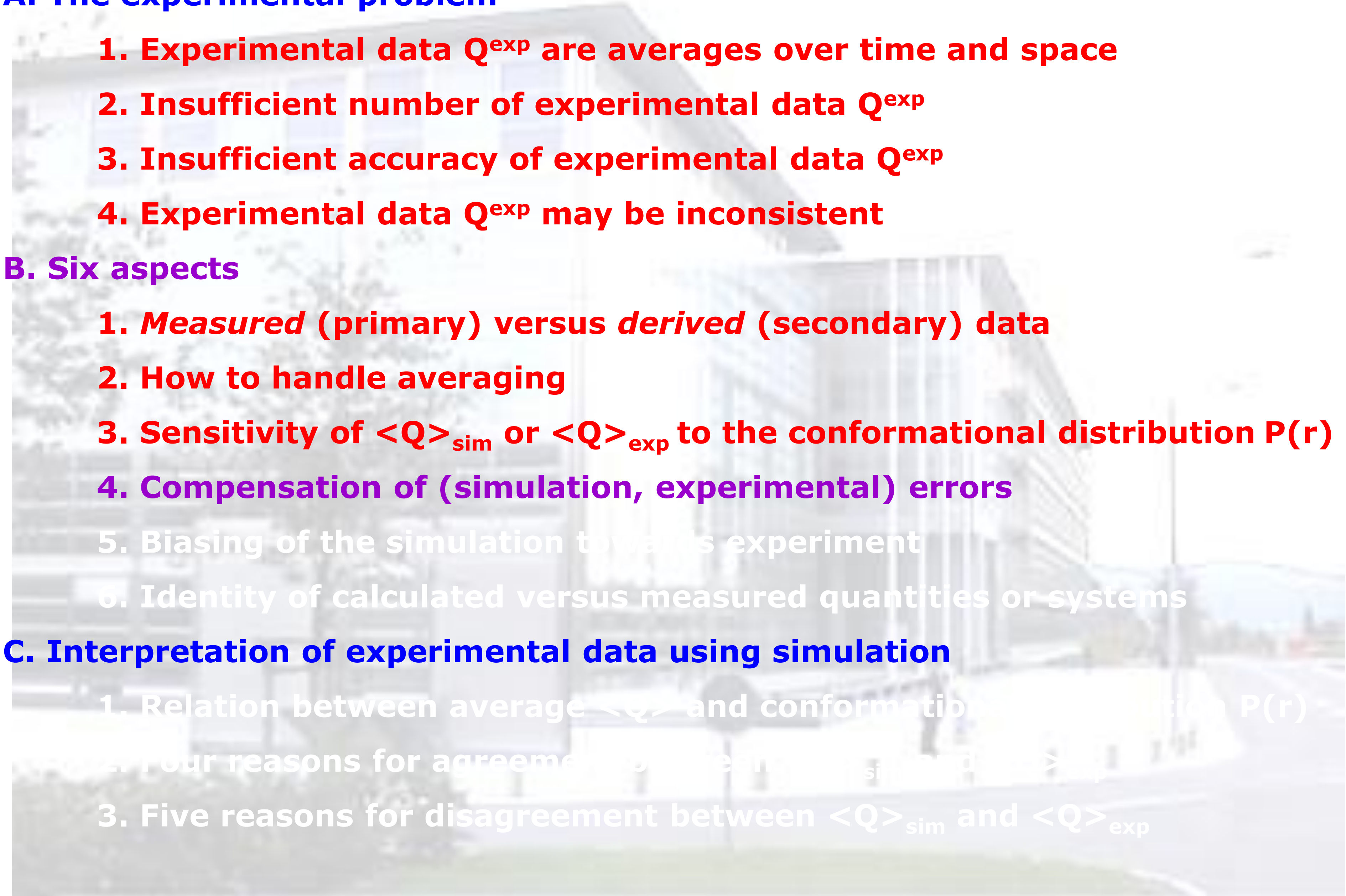
Q2 <Q> does not reflect the shape of P(**r**) as <Q> is determined by ***rarely sampled*** conformations with small (*irrelevant*) Boltzmann weights

Q3 <Q> ***does reflect*** the dominant conformations of P(**r**)

Only in case **Q3** can $<Q>_{sim}$ carry information relevant to the interpretation of $<Q>_{exp}$ at a molecular level

*W.F. van Gunsteren, J. Dolenc, & A.E. Mark, Curr. Opin. Struct. Biol., 18 (**2008**) 149-153*

On comparing molecular modelling results with experimental data



A. The experimental problem

1. Experimental data Q^{exp} are averages over time and space
2. Insufficient number of experimental data Q^{exp}
3. Insufficient accuracy of experimental data Q^{exp}
4. Experimental data Q^{exp} may be inconsistent

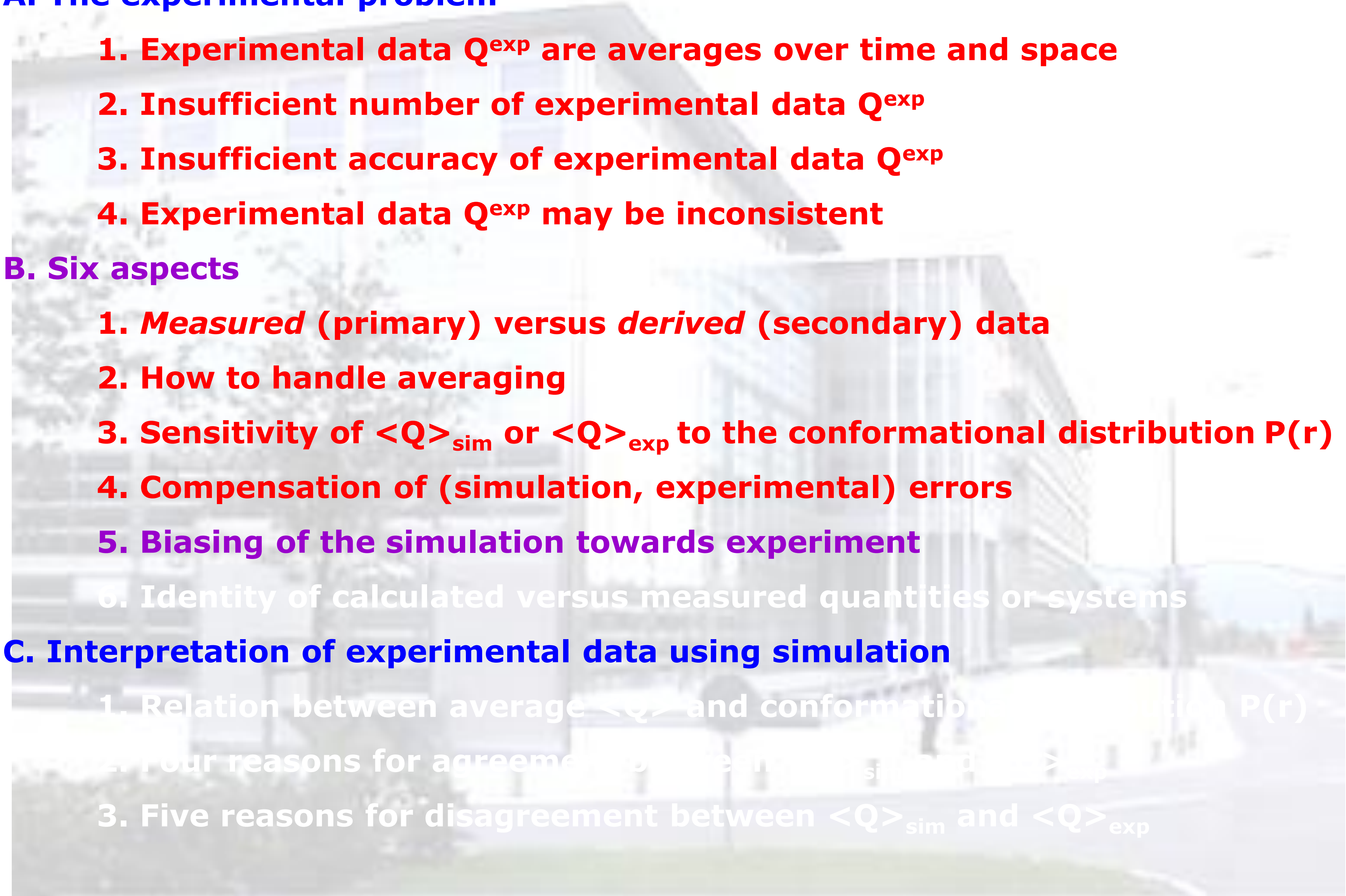
B. Six aspects

1. *Measured* (primary) versus *derived* (secondary) data
2. How to handle averaging
3. Sensitivity of $<Q>_{sim}$ or $<Q>_{exp}$ to the conformational distribution P(r)
4. Compensation of (simulation, experimental) errors
5. Biasing of the simulation towards experiment
6. Identity of calculated versus measured quantities or systems

C. Interpretation of experimental data using simulation

1. Relation between average $<Q>$ and conformational distribution P(r)
2. Four reasons for agreement between $<Q>_{sim}$ and $<Q>_{exp}$
3. Five reasons for disagreement between $<Q>_{sim}$ and $<Q>_{exp}$

On comparing molecular modelling results with experimental data



A. The experimental problem

1. Experimental data Q^{exp} are averages over time and space
2. Insufficient number of experimental data Q^{exp}
3. Insufficient accuracy of experimental data Q^{exp}
4. Experimental data Q^{exp} may be inconsistent

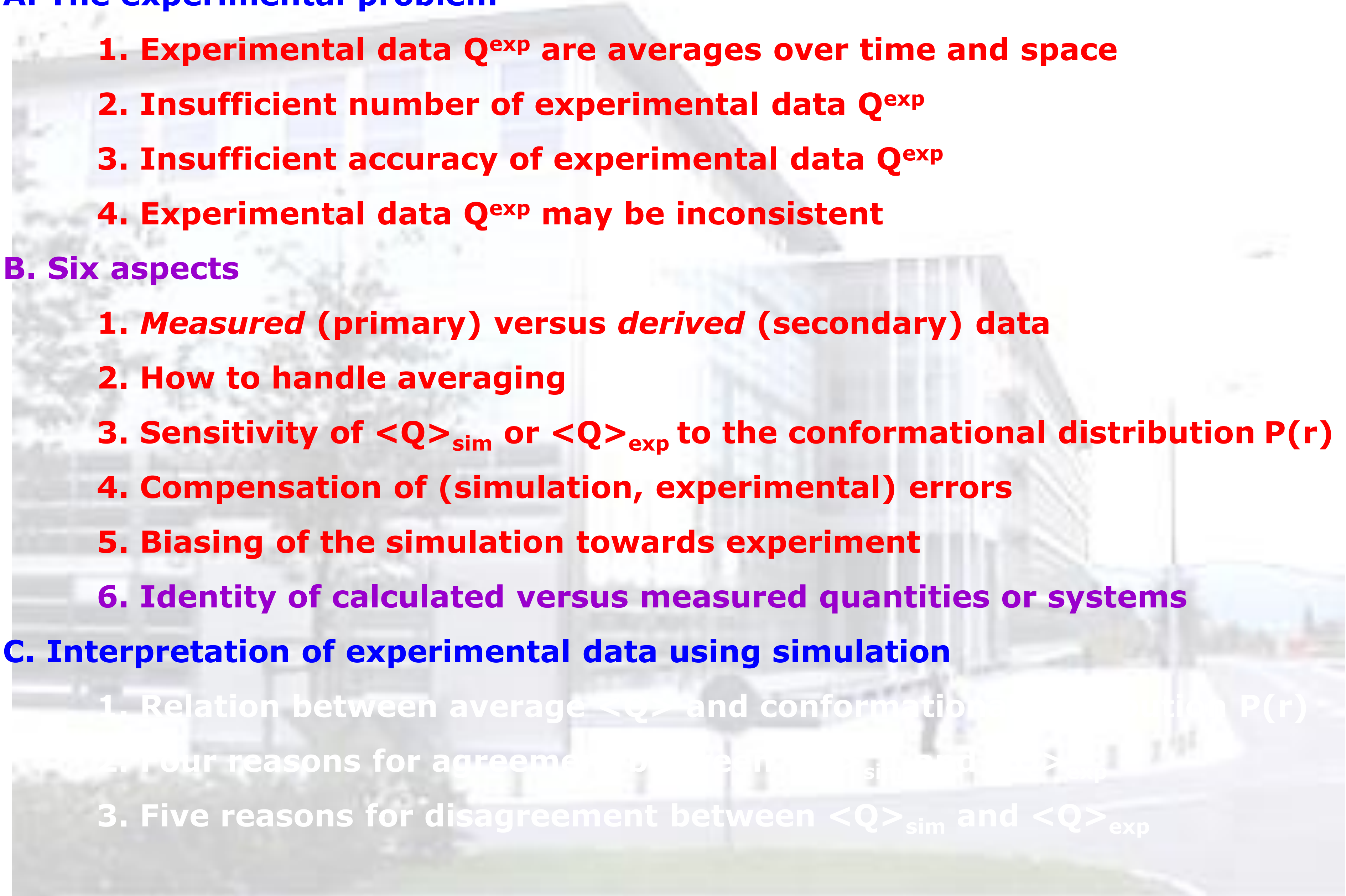
B. Six aspects

1. *Measured* (primary) versus *derived* (secondary) data
2. How to handle averaging
3. Sensitivity of $<Q>_{sim}$ or $<Q>_{exp}$ to the conformational distribution P(r)
4. Compensation of (simulation, experimental) errors
5. Biasing of the simulation towards experiment
6. Identity of calculated versus measured quantities or systems

C. Interpretation of experimental data using simulation

1. Relation between average <Q> and conformational distribution P(r)
2. Four reasons for agreement between $<Q>_{sim}$ and $<Q>_{exp}$
3. Five reasons for disagreement between $<Q>_{sim}$ and $<Q>_{exp}$

On comparing molecular modelling results with experimental data



A. The experimental problem

1. Experimental data Q^{exp} are averages over time and space
2. Insufficient number of experimental data Q^{exp}
3. Insufficient accuracy of experimental data Q^{exp}
4. Experimental data Q^{exp} may be inconsistent

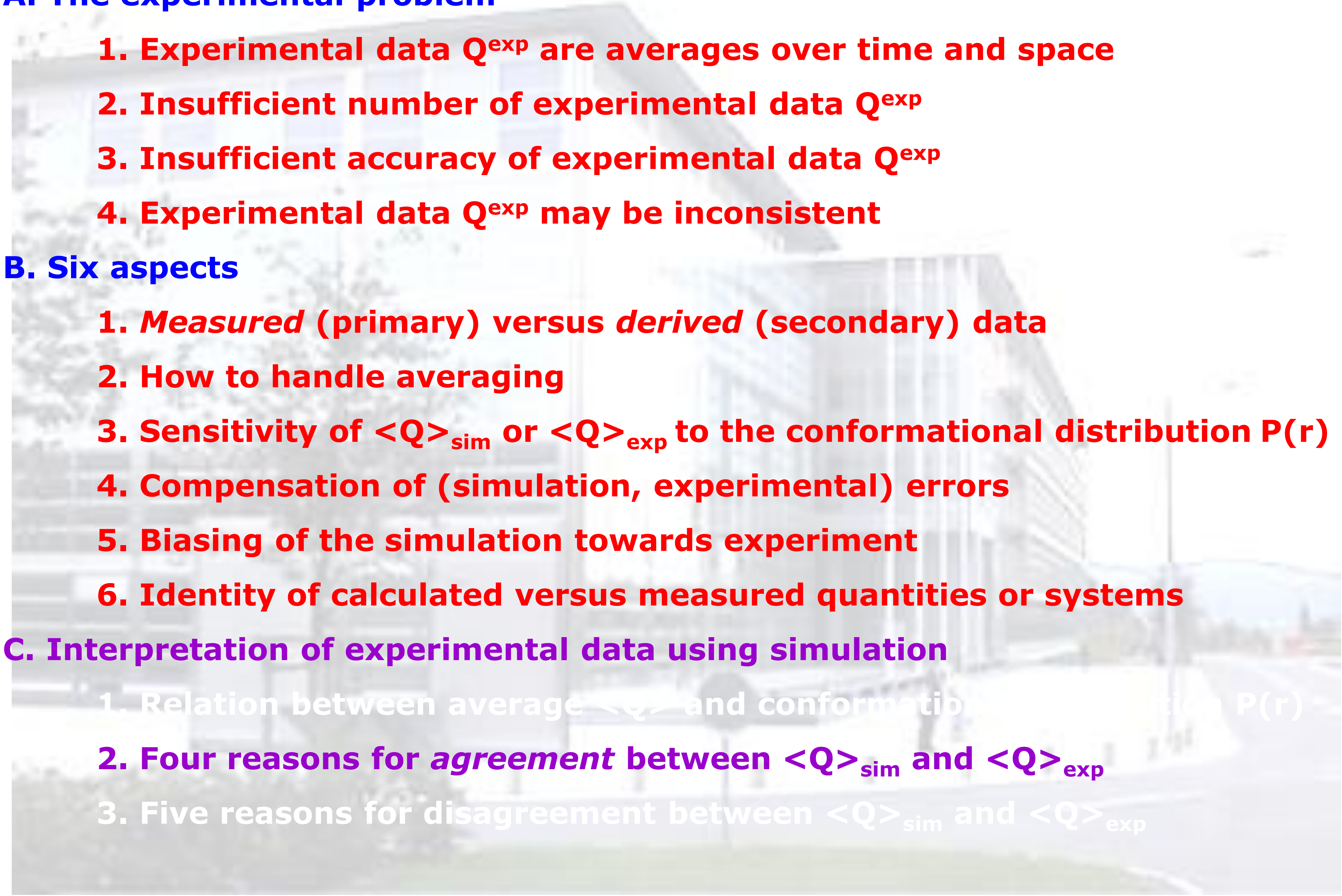
B. Six aspects

1. *Measured* (primary) versus *derived* (secondary) data
2. How to handle averaging
3. Sensitivity of $<Q>_{sim}$ or $<Q>_{exp}$ to the conformational distribution P(r)
4. Compensation of (simulation, experimental) errors
5. Biasing of the simulation towards experiment
6. Identity of calculated versus measured quantities or systems

C. Interpretation of experimental data using simulation

1. Relation between average <Q> and conformational distribution P(r)
2. Four reasons for agreement between $<Q>_{sim}$ and $<Q>_{exp}$
3. Five reasons for disagreement between $<Q>_{sim}$ and $<Q>_{exp}$

On comparing molecular modelling results with experimental data



A. The experimental problem

1. Experimental data Q^{exp} are averages over time and space
2. Insufficient number of experimental data Q^{exp}
3. Insufficient accuracy of experimental data Q^{exp}
4. Experimental data Q^{exp} may be inconsistent

B. Six aspects

1. *Measured* (primary) versus *derived* (secondary) data
2. How to handle averaging
3. Sensitivity of $<Q>_{sim}$ or $<Q>_{exp}$ to the conformational distribution P(r)
4. Compensation of (simulation, experimental) errors
5. Biasing of the simulation towards experiment
6. Identity of calculated versus measured quantities or systems

C. Interpretation of experimental data using simulation

1. Relation between average $<Q>$ and conformational distribution P(r)
2. Four reasons for *agreement* between $<Q>_{sim}$ and $<Q>_{exp}$
3. Five reasons for disagreement between $<Q>_{sim}$ and $<Q>_{exp}$

Interpretation of Experimental Data using Simulation

Reasons for ***agreement*** between $\langle Q \rangle_{sim}$ and $\langle Q \rangle_{exp}$

Agreement between $<Q>_{sim}$ and $<Q>_{exp}$ may be obtained if:

A1 <Q> is insensitive to P(**r**), i.e. $<Q>_{sim}$ matches $<Q>_{exp}$ irrespective of the conformational distribution P(**r**) simulated

A2 There are compensating errors in the simulation model, procedure or experimental set-up

A3 The experimental data of interest, $<Q>_{exp}$, has been used to bias the simulation

A4 $<Q(\mathbf{r})>_{sim}$ is sensitive to the distribution P(**r**)

Only in the case **A4** can the degree of agreement between $<Q>_{sim}$ and $<Q>_{exp}$ be used to validate the simulation and/or to interpret the experimental results

*W.F. van Gunsteren, J. Dolenc, & A.E. Mark, Curr. Opin. Struct. Biol., 18 (**2008**) 149-153*

Interpretation of Experimental Data using Simulation

Reasons for agreement between $\langle Q \rangle_{sim}$ and $\langle Q \rangle_{exp}$

Agreement between $<Q>_{sim}$ and $<Q>_{exp}$ may be obtained if:

A1 <Q> ***is insensitive*** to P(**r**), i.e. $<Q>_{sim}$ matches $<Q>_{exp}$ irrespective of the P(**r**) simulated

A2 There are compensating errors in the simulation model, procedure or experimental set-up

A3 The experimental data of interest, $<Q>_{exp}$, has been used to bias the simulation

A4 $<Q(\mathbf{r})>_{sim}$ is sensitive to the distribution P(**r**)

Only in the case **A4** can the degree of agreement between $<Q>_{sim}$ and $<Q>_{exp}$ be used to validate the simulation and/or to interpret the experimental results

*W.F. van Gunsteren, J. Dolenc, & A.E. Mark, Curr. Opin. Struct. Biol., 18 (**2008**) 149-153*

Interpretation of Experimental Data using Simulation

Reasons for agreement between $\langle Q \rangle_{sim}$ and $\langle Q \rangle_{exp}$

Agreement between $<Q>_{sim}$ and $<Q>_{exp}$ may be obtained if:

A1 <Q> is insensitive to P(**r**), i.e. $<Q>_{sim}$ matches $<Q>_{exp}$ irrespective of the P(**r**) simulated

A2 There are ***compensating errors*** in the simulation model, procedure or experimental set-up

A3 The experimental data of interest, $<Q>_{exp}$, has been used to bias the simulation

A4 $<Q(\mathbf{r})>_{sim}$ is sensitive to the distribution P(**r**)

Only in the case **A4** can the degree of agreement between $<Q>_{sim}$ and $<Q>_{exp}$ be used to validate the simulation and/or to interpret the experimental results

*W.F. van Gunsteren, J. Dolenc, & A.E. Mark, Curr. Opin. Struct. Biol., 18 (**2008**) 149-153*

Interpretation of Experimental Data using Simulation

Reasons for agreement between $\langle Q \rangle_{sim}$ and $\langle Q \rangle_{exp}$

Agreement between $<Q>_{sim}$ and $<Q>_{exp}$ may be obtained if:

A1 <Q> is insensitive to P(**r**), i.e. $<Q>_{sim}$ matches $<Q>_{exp}$ irrespective of the P(**r**) simulated

A2 There are compensating errors in the simulation model, procedure or experimental set-up

A3 The experimental data of interest, $<Q>_{exp}$, has been used to ***bias*** the simulation

A4 $<Q(\mathbf{r})>_{sim}$ is sensitive to the distribution P(**r**)

Only in the case **A4** can the degree of agreement between $<Q>_{sim}$ and $<Q>_{exp}$ be used to validate the simulation and/or to interpret the experimental results

*W.F. van Gunsteren, J. Dolenc, & A.E. Mark, Curr. Opin. Struct. Biol., 18 (**2008**) 149-153*

Interpretation of Experimental Data using Simulation

Reasons for agreement between $\langle Q \rangle_{sim}$ and $\langle Q \rangle_{exp}$

Agreement between $<Q>_{sim}$ and $<Q>_{exp}$ may be obtained if:

A1 <Q> is insensitive to P(**r**), i.e. $<Q>_{sim}$ matches $<Q>_{exp}$ irrespective of the P(**r**) simulated

A2 There are compensating errors in the simulation model, procedure or experimental set-up

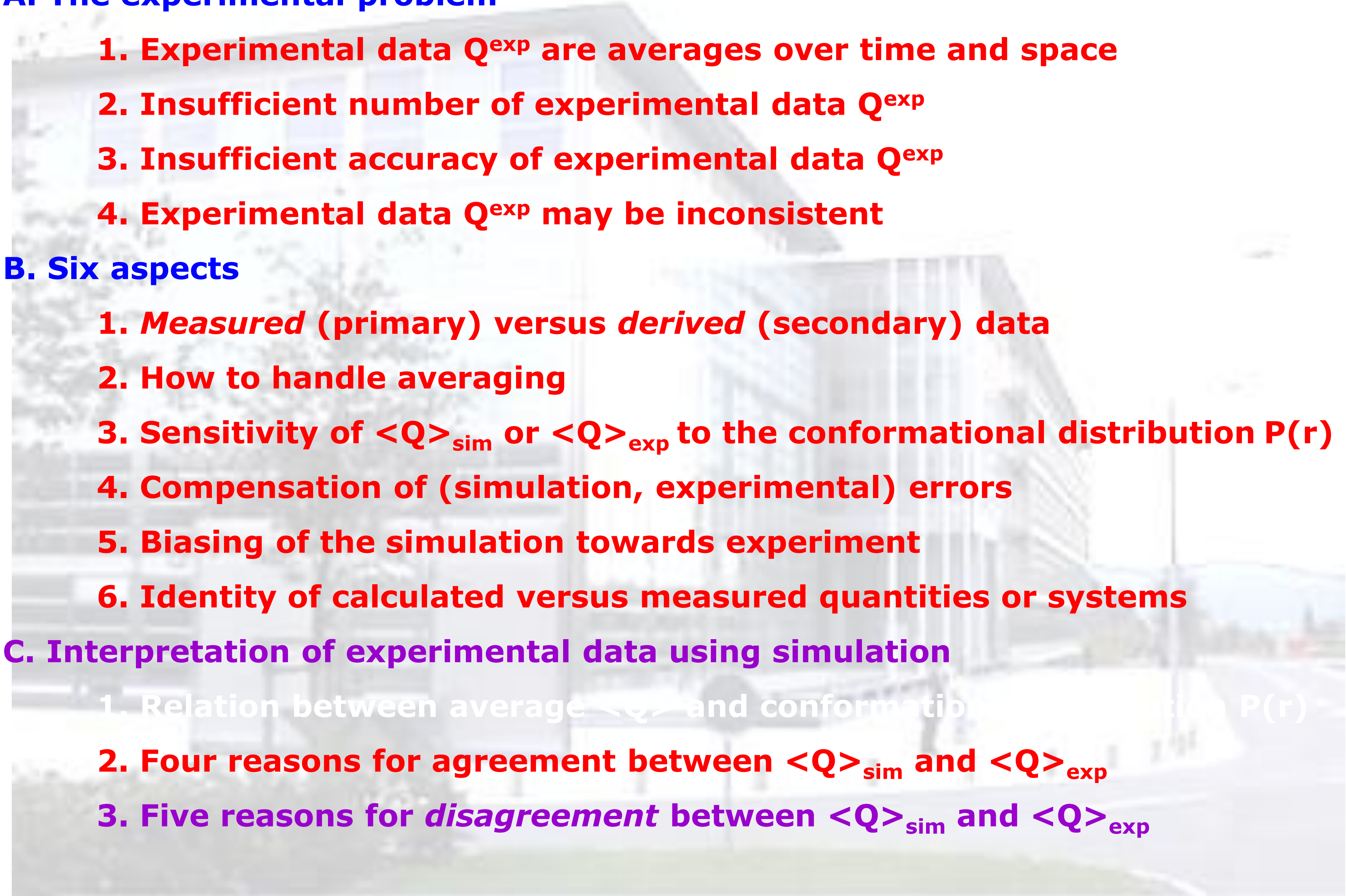
A3 The experimental data of interest, $<Q>_{exp}$, has been used to bias the simulation

A4 $<Q(\mathbf{r})>_{sim}$ ***is sensitive*** to the distribution P(**r**)

Only in the case **A4** can the degree of agreement between $<Q>_{sim}$ and $<Q>_{exp}$ be used to validate the simulation and/or to interpret the experimental results

*W.F. van Gunsteren, J. Dolenc, & A.E. Mark, Curr. Opin. Struct. Biol., 18 (**2008**) 149-153*

On comparing molecular modelling results with experimental data



A. The experimental problem

1. Experimental data Q^{exp} are averages over time and space
2. Insufficient number of experimental data Q^{exp}
3. Insufficient accuracy of experimental data Q^{exp}
4. Experimental data Q^{exp} may be inconsistent

B. Six aspects

1. *Measured* (primary) versus *derived* (secondary) data
2. How to handle averaging
3. Sensitivity of $<Q>_{sim}$ or $<Q>_{exp}$ to the conformational distribution P(r)
4. Compensation of (simulation, experimental) errors
5. Biasing of the simulation towards experiment
6. Identity of calculated versus measured quantities or systems

C. Interpretation of experimental data using simulation

1. Relation between average $<Q>$ and conformational distribution P(r)
2. Four reasons for agreement between $<Q>_{sim}$ and $<Q>_{exp}$
3. Five reasons for *disagreement* between $<Q>_{sim}$ and $<Q>_{exp}$

Interpretation of Experimental Data using Simulation

Reasons for ***disagreement*** between $\langle Q \rangle_{sim}$ and $\langle Q \rangle_{exp}$

Failure to observe a correlation between the simulation and experiment can be due to many reasons:

D1 The simulation is insufficiently accurate: i.e.

a) relevant degrees of freedom were omitted;

b) the force field was insufficiently accurate;

c) approximations made when solving the equations of motion were too crude;

d) inappropriate thermodynamic or spatial boundary conditions were used.

D2 The measured $<Q>_{exp}$ is inaccurate

D3 $<Q>_{sim}$ and $<Q>_{exp}$ are averaged differently with respect to time or spatial extent

D4 Related but different quantities are compared, e.g. atom-positional fluctuations versus crystallographic B factors

D5 Different systems are compared (e.g. crystal versus solution), or systems studied under different thermodynamic conditions (e.g. temperature, pressure, pH, ionic strength, etc.)

*W.F. van Gunsteren, J. Dolenc, & A.E. Mark, Curr. Opin. Struct. Biol., 18 (**2008**) 149-153*

Interpretation of Experimental Data using Simulation

Reasons for disagreement between $\langle Q \rangle_{sim}$ and $\langle Q \rangle_{exp}$

Failure to observe a correlation between the simulation and experiment can be due to many reasons:

D1 The simulation is insufficiently accurate: i.e.

a) relevant degrees of freedom were omitted;

b) the force field was insufficiently accurate;

c) approximations made when solving the equations of motion were too crude;

d) inappropriate thermodynamic or spatial boundary conditions were used.

D2 The measured $<Q>_{exp}$ is inaccurate

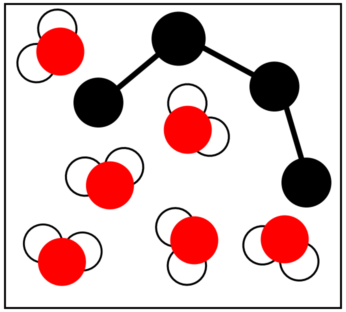
D3 $<Q>_{sim}$ and $<Q>_{exp}$ are averaged differently with respect to time or spatial extent

D4 Related but different quantities are compared, e.g. atom-positional fluctuations versus crystallographic B factors

D5 Different systems are compared (e.g. crystal versus solution), or systems studied under different thermodynamic conditions (e.g. temperature, pressure, pH, ionic strength, etc.)

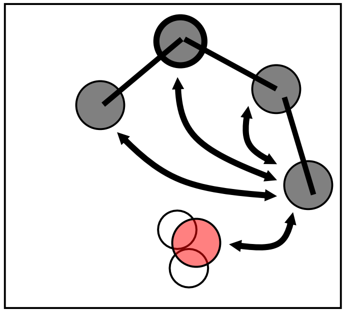
*W.F. van Gunsteren, J. Dolenc, & A.E. Mark, Curr. Opin. Struct. Biol., 18 (**2008**) 149-153*

Definition of a model for molecular simulation



Degrees of freedom: atoms are the elementary particles

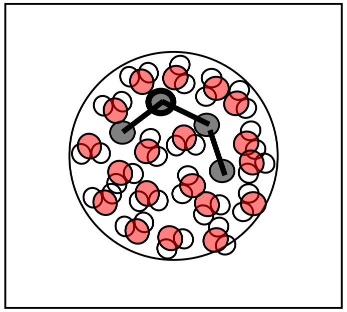
Forces or interactions between atoms



Force field = physico-chemical knowledge

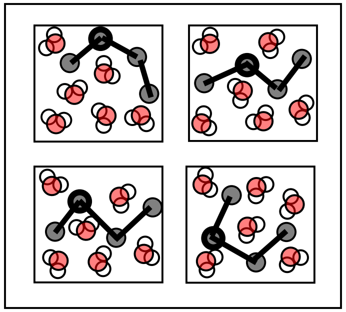
MOLECULAR MODEL

Boundary conditions



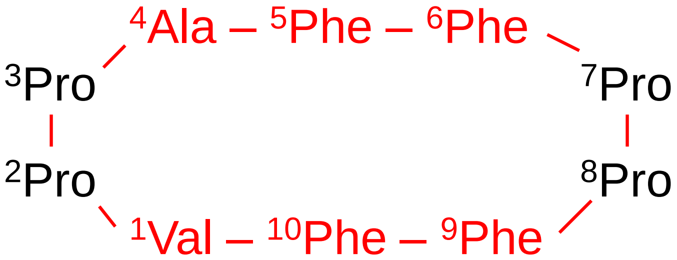
system
temperature
pressure
walls
external forces

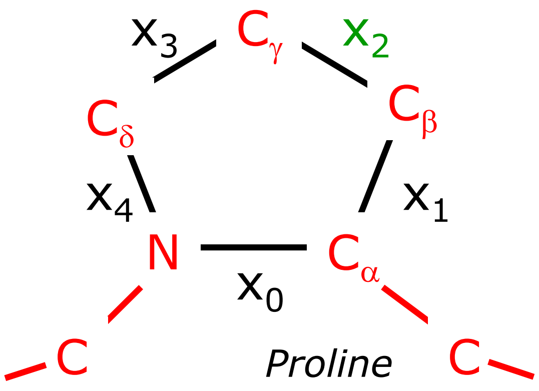
Methods to generate configurations of atoms: Newton



Conformational Dynamics of Proline Residues in Antamanide:

Effect of explicit solvent versus continuum: missing degrees of freedom





Experiment : NMR [**^{13}C relaxation** / **E COSY**] ***R.R. Ernst***

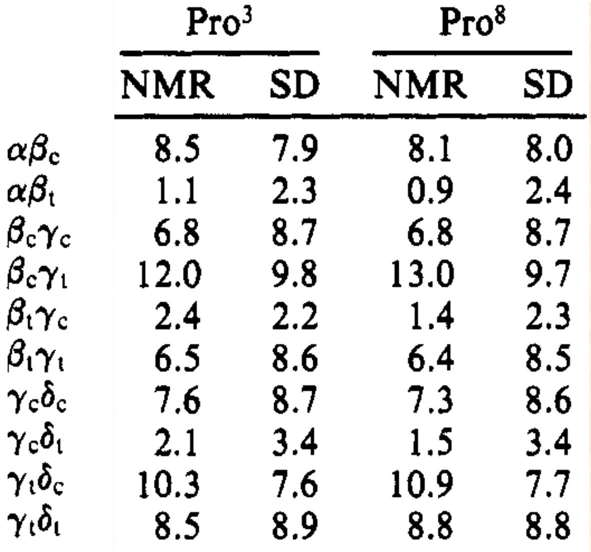
3/8 Pro: rigid

2/7 Pro: 2 conformers (time constant ~ 30ps)

Simulation: stochastic dynamics (500ps) SD

Comparison of $^3J_{HH}$ coupling constants (in Hz) from NMR

Rms deviation simulation-experiment 1.5 Hz



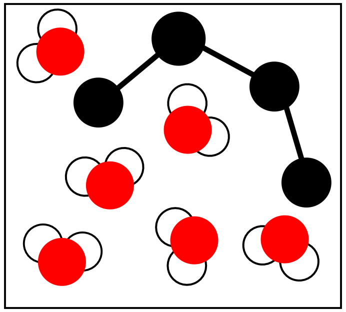
	Pro^3		Pro^8	
	NMR	SD	NMR	SD
$\alpha\beta_c$	8.5	7.9	8.1	8.0
$\alpha\beta_t$	1.1	2.3	0.9	2.4
$\beta_c\gamma_c$	6.8	8.7	6.8	8.7
$\beta_c\gamma_t$	12.0	9.8	13.0	9.7
$\beta_t\gamma_c$	2.4	2.2	1.4	2.3
$\beta_t\gamma_t$	6.5	8.6	6.4	8.5
$\gamma_c\delta_c$	7.6	8.7	7.3	8.6
$\gamma_c\delta_t$	2.1	3.4	1.5	3.4
$\gamma_t\delta_c$	10.3	7.6	10.9	7.7
$\gamma_t\delta_t$	8.5	8.9	8.8	8.8

Dynamics	**GROMOS force field change**	**Friction coefficient *ps^{-1}***	**Residence time *ps***
Experiment			**≈ 30**
SD mean solvent	**-**	**19**	**3**
SD mean solvent	**-**	**1000**	**25**
SD mean solvent	**torsion x kT up**	**19**	**25**
MD explicit solvent	**-**	**-**	**24**

Solvent degrees of freedom are essential for dynamics

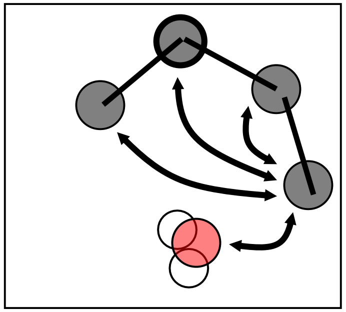
*R.M. Brunne et al., JACS, 115 (**1993**) 4764-4768*
*J.W. Peng et al., J. Biomol. NMR 8 (**1996**) 453-476*

Definition of a model for molecular simulation



Degrees of freedom:
atoms are the
elementary particles

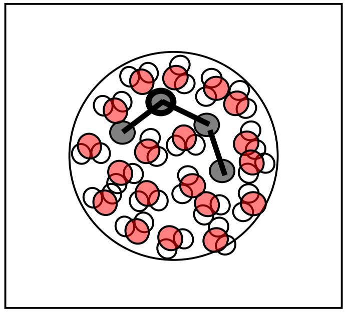
Forces or
interactions
between atoms



Force field =
physico-chemical
knowledge

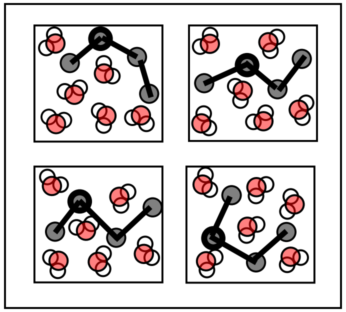
MOLECULAR
MODEL

Boundary conditions



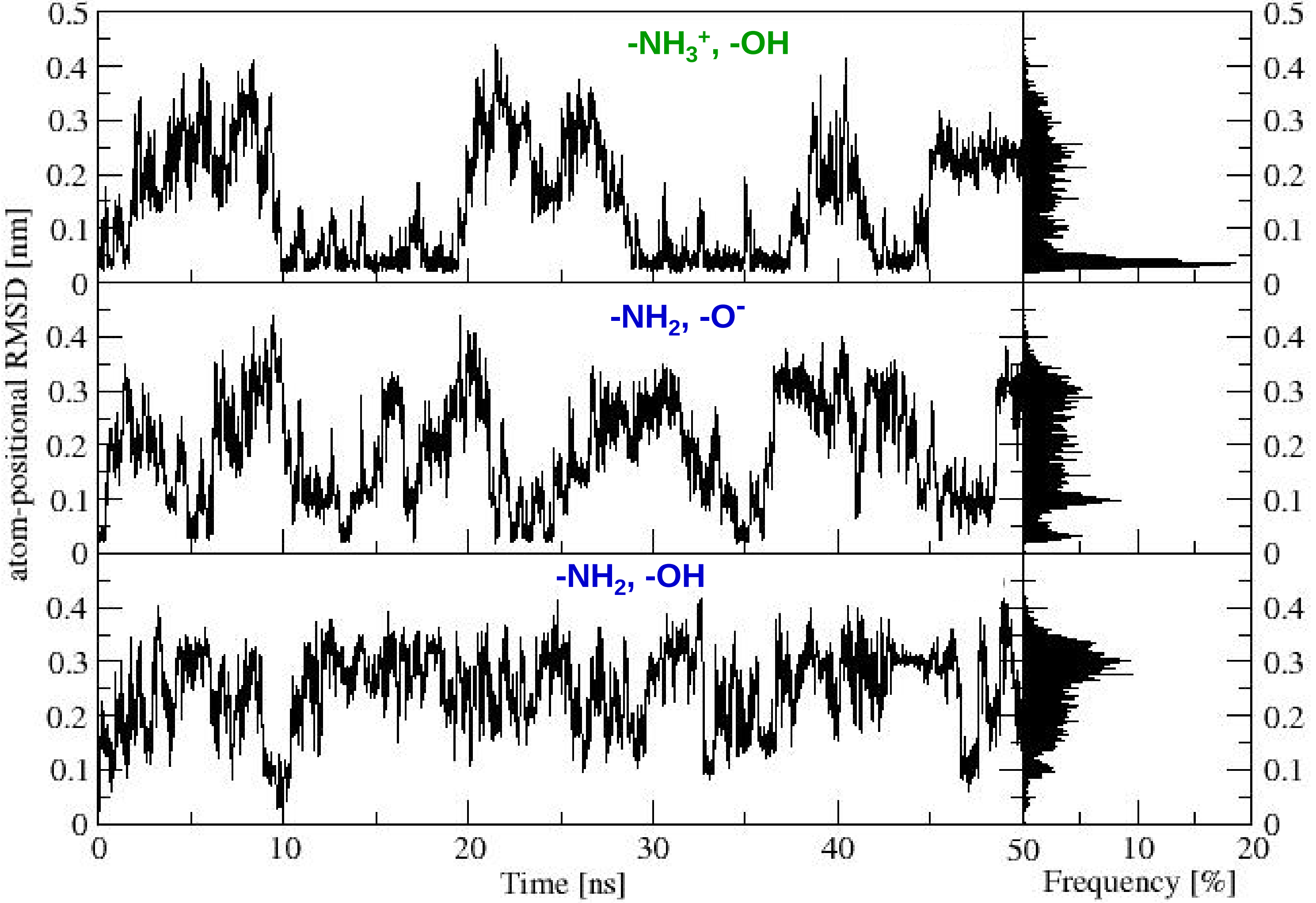
system
temperature
pressure
walls
external forces

Methods to generate
configurations of
atoms: Newton



pH Dependence of the folding equilibrium of a β-peptide in methanol solvent

Backbone atom-positional RMSD from the helical fold



Low pH:
Helix dominant
Corresponds to experiment

High pH:
Bit of helix present

Intermediate pH:
No helix present

Thermodynamic conditions chosen in a simulation may influence the result, i.c. the folding equilibrium

*P.J. Gee & W.F. van Gunsteren, Proteins 63 (**2006**) 136-143*

Interpretation of Experimental Data using Simulation

Reasons for disagreement between $\langle Q \rangle_{sim}$ and $\langle Q \rangle_{exp}$

Failure to observe a correlation between the simulation and experiment can be due to many reasons:

D1 The simulation is insufficiently accurate: i.e.

a) relevant degrees of freedom were omitted;

b) the force field was insufficiently accurate;

c) approximations made when solving the equations of motion were too crude;

d) inappropriate thermodynamic or spatial boundary conditions were used.

D2 The measured $<Q>_{exp}$ is inaccurate

D3 $<Q>_{sim}$ and $<Q>_{exp}$ are averaged differently with respect to time or spatial extent

D4 Related but different quantities are compared, e.g. atom-positional fluctuations versus crystallographic B factors

D5 Different systems are compared (e.g. crystal versus solution), or systems studied under different thermodynamic conditions (e.g. temperature, pressure, pH, ionic strength, etc.)

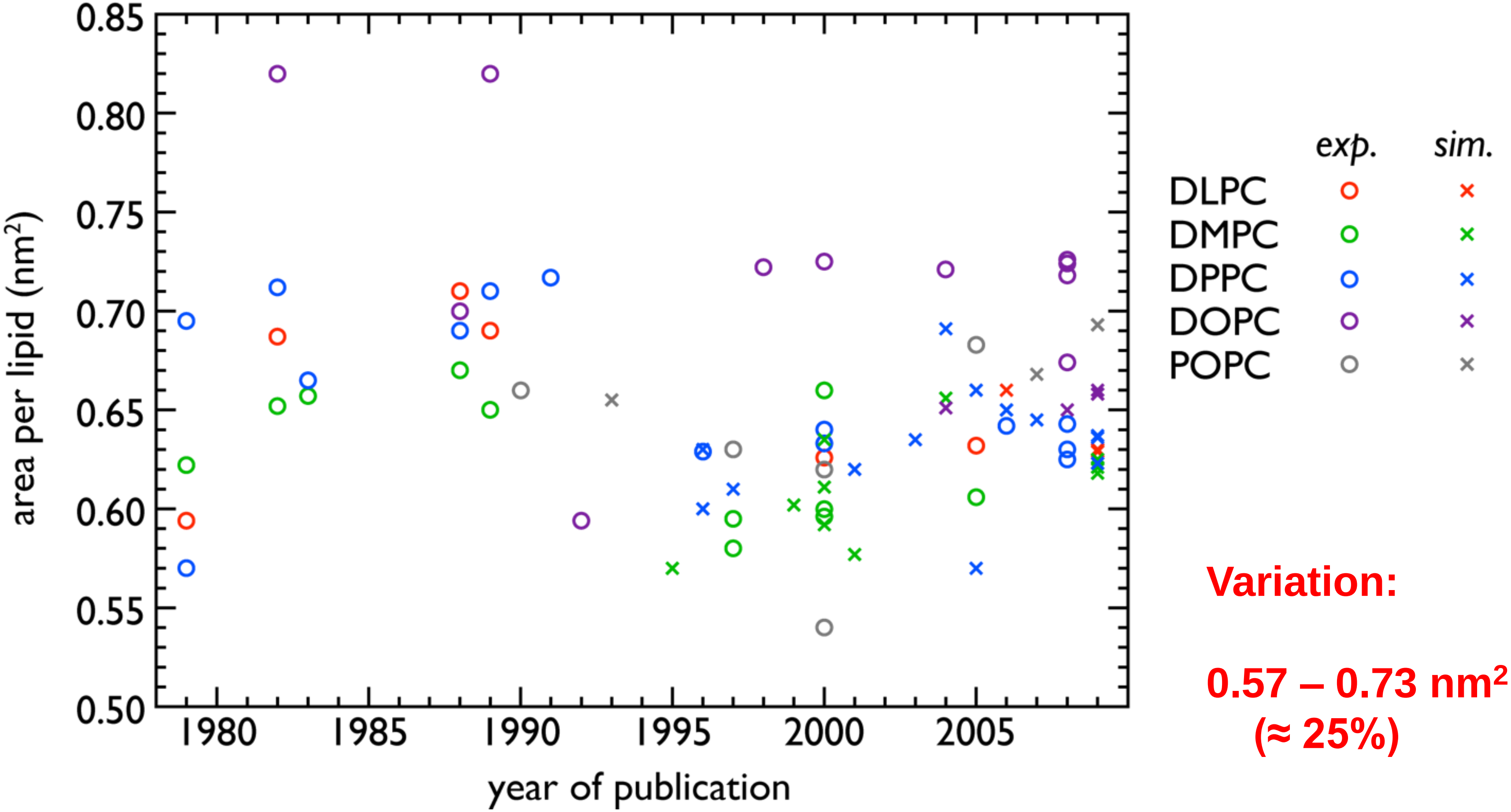
*W.F. van Gunsteren, J. Dolenc, & A.E. Mark, Curr. Opin. Struct. Biol., 18 (**2008**) 149-153*

Modelling a specific membrane: area per lipid

Area per lipid:

*D. Poger & A.E. Mark, JCTC 6 (**2010**) 325-226*

1. Most commonly used experimental quantity to validate a lipid model.
2. Difficult to measure directly.
3. Often inferred from NMR relaxation data.
4. Depends on measurement conditions.
5. Few research groups generate data.



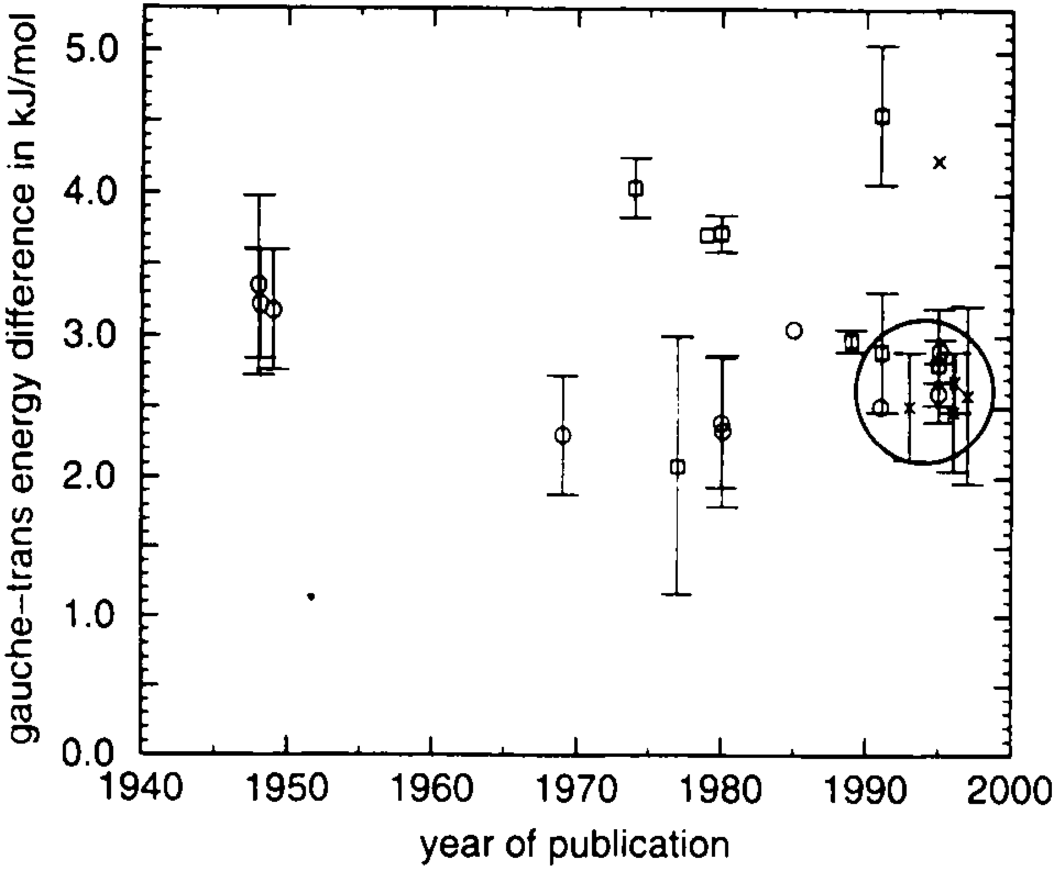
Variation:

0.57 – 0.73 nm^2 (≈ 25%)

Experimental data vary with time

Variation over time in the experimental data regarding the trans-gauche energy difference in aliphatic chains

This quantity will influence the structure and mobility of lipid chains



Variation:
2.0 to
3.2 – 4.5 kJ/mol

Experimental data vary with time

*Schuler & van Gunsteren, Mol. Simulation 25 (**2000**) 301-319*

Test of Force Field and NMR Data for Hen Egg White Lysozyme

Experimental Data

(Smith et. al., 1991, 1993; Buck et. al., 1995; Schwalbe et. al., 2001, both Oxford)

1158 NOE's derived inter-proton distances (set1 1993)

1525 NOE's derived inter-proton distances (set2 2001)

95 $^3J_{HN\alpha}$-coupling constants

100 $^3J_{\alpha\beta}$-coupling constants

124 backbone and 28 side-chain order parameters

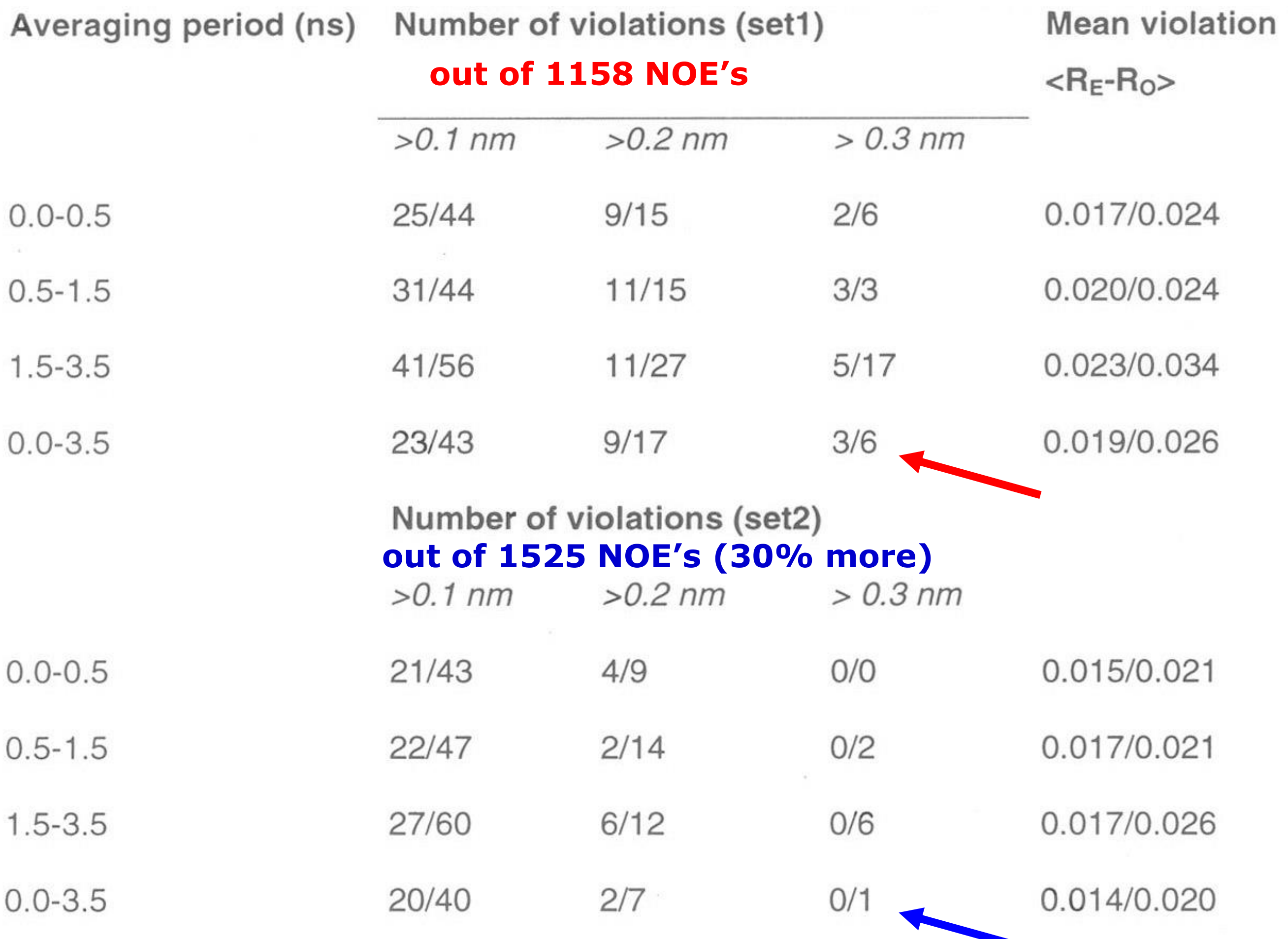
X-ray coordinates (PDB 1aki, 1.5 Å)

NMR coordinates (PDB 1e8l, set of 50 structures)

*Soares et al., J. Biomol. NMR 30 (**2004**) 407-422*
*van Gunsteren et al., Angew. Chemie Intl. Ed. 45 (**2006**) 4064-4092*

NOE distance bound violations in HEWL

NOE bound violations computed from MD trajectories (43A1(1996)/45A3(2001)) **against *two sets of experimental NOE distance bounds*** from Smith *et. al.* (set1, 1993) and from Schwalbe *et. al.* (set2, 2001)



Averaging period (ns)	Number of violations (set1) out of 1158 NOE's			Mean violation $<R_E-R_O>$
	>0.1 nm	*>0.2 nm*	*> 0.3 nm*	
0.0-0.5	25/44	9/15	2/6	0.017/0.024
0.5-1.5	31/44	11/15	3/3	0.020/0.024
1.5-3.5	41/56	11/27	5/17	0.023/0.034
0.0-3.5	23/43	9/17	3/6	0.019/0.026
	Number of violations (set2) out of 1525 NOE's (30% more)			
	>0.1 nm	*>0.2 nm*	*> 0.3 nm*	
0.0-0.5	21/43	4/9	0/0	0.015/0.021
0.5-1.5	22/47	2/14	0/2	0.017/0.021
1.5-3.5	27/60	6/12	0/6	0.017/0.026
0.0-3.5	20/40	2/7	0/1	0.014/0.020

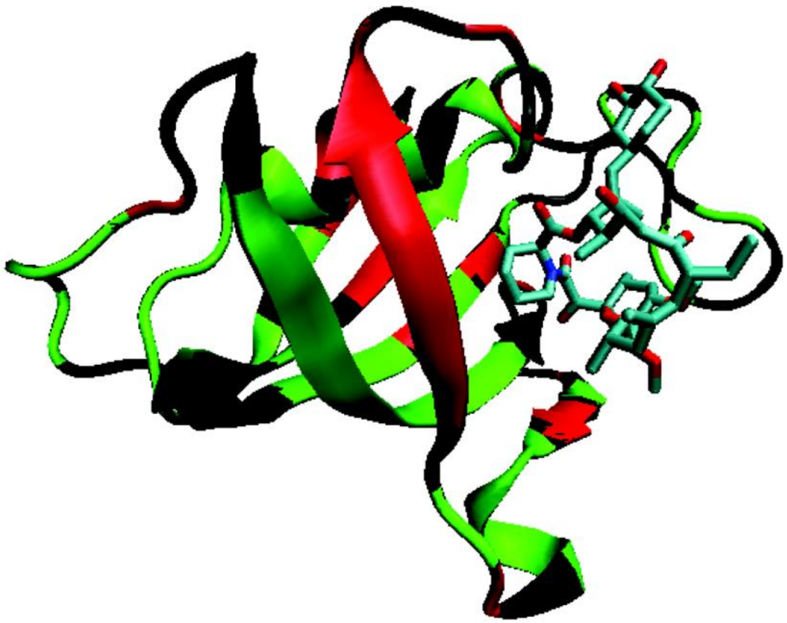
1993 set

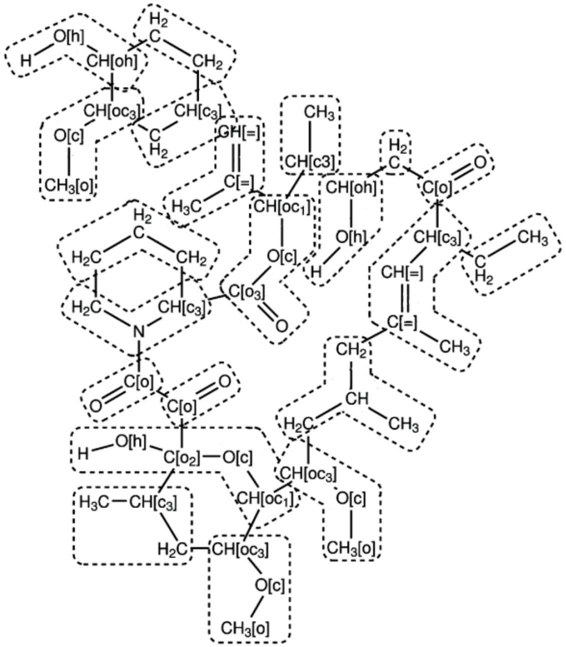
2001 set

Over time (1993 → 2001) the experimental data converged towards simulated ones

FKBP (107 residues) + ascomycin

inconsistent experimental data





The protein is coloured according to whether or not there is a range of χ_1 dihedral angle values corresponding to the experimental ^3J-coupling data (±1 Hz variation, distribution analysis):

black: no data (no ^3J-couplings : 38; one ^3J-coupling : 7 residues)

green: there is a single, continuous range of angle values that satisfies all of the experimental data (39 out of 62 residues: 63%)

red: there is no such solution (23 out of 62 residues: 37 %): ***inconsistent ?***

yes, for 5 residues

103 $^3J_{N\text{-}H\beta}$ and 94 $^3J_{H\alpha\text{-}H\beta}$-values

*Allison & van Gunsteren, ChemPhysChem 10 (**2009**) 3213-3228*

Interpretation of Experimental Data using Simulation

Reasons for disagreement between $\langle Q \rangle_{sim}$ and $\langle Q \rangle_{exp}$

Failure to observe a correlation between the simulation and experiment can be due to many reasons:

D1 The simulation is insufficiently accurate: i.e.

a) relevant degrees of freedom were omitted;

b) the force field was insufficiently accurate;

c) approximations made when solving the equations of motion were too crude;

d) inappropriate thermodynamic or spatial boundary conditions were used.

D2 The measured $<Q>_{exp}$ is inaccurate

D3 $<Q>_{sim}$ and $<Q>_{exp}$ are averaged differently with respect to time or spatial extent

D4 Related but different quantities are compared, e.g. differently defined free energies of folding

D5 Different systems are compared (e.g. crystal versus solution), or systems studied under different thermodynamic conditions (e.g. temperature, pressure, pH, ionic strength, etc.)

*W.F. van Gunsteren, J. Dolenc, & A.E. Mark, Curr. Opin. Struct. Biol., 18 (**2008**) 149-153*

Reasons for disagreement between <Q>$_{sim}$ and <Q>$_{exp}$

Related but different quantities are compared

$$\langle Q \rangle_{sim} \Leftrightarrow \langle Q' \rangle_{exp}$$

1. $Q'(\vec{r})$ is a **free energy change**

$$\Delta G_{\text{folding}} = G_{\text{fold}} - G_{\text{denatured}}$$

of folding or renaturation,

as derived from experiment by ***changing the thermodynamic conditions***:

- *temperature change* ⟵
- pH change
- ionic strength or *co-solvent change* ⟵

Different solute stabilities
or
free energies of folding Q'(r)

2. $Q(\vec{r})$ is a **free energy change**

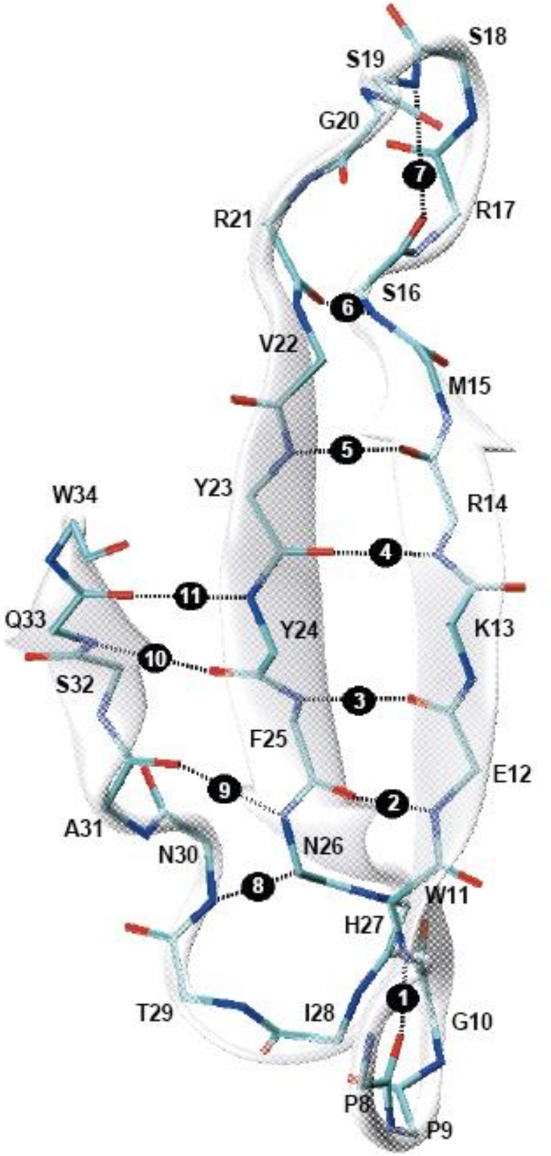
$$\Delta G_{\text{folding}} = G_{\text{fold}} - G_{\text{unfolded}}$$

of folding ***as derived from*** one ***simulation*** at one given thermodynamic state point by ***counting the ratio of folded versus unfolded conformations***

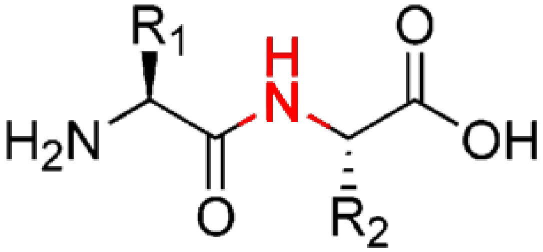
Comparison has only limited value: $\Delta G_{folding}\left(Q'\right) \neq \Delta G_{folding}\left(Q\right)$

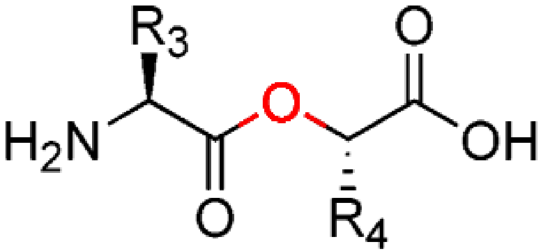
The WW domain of the PIN1 protein:

Contribution of backbone hydrogen bonding to β-sheet stability



11 backbone-backbone hydrogen bonds, distributed over three β-strands, *each NH individually replaced by Oxygen*





$R_1 = R_3$ and $R_2 = R_4$

Deechongkit et al., JACS 126 (2004) 16762

20 single residue mutants synthesised and their stability measured by experiments:

1. Thermal denaturation
2. Chaotrope denaturation (Gd-Cl)

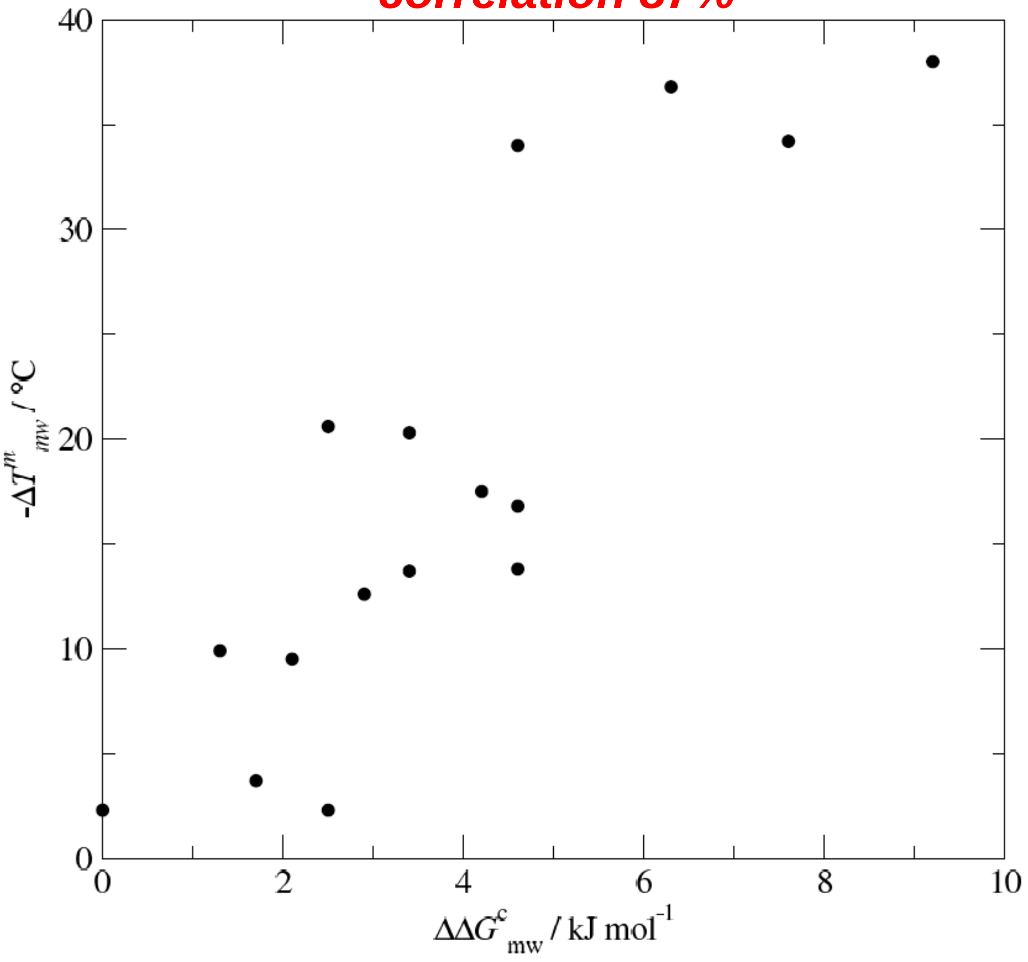
The WW domain of the PIN1 protein:

Contribution of backbone hydrogen bonding to β-sheet stability

Deechongkit et al., JACS 126 (2004) 16762

Temperature versus chaotrope denaturation

correlation 87%



Dominant fold or structure of the 16 mutants verified by:

1. far-UV CD spectroscopy
2. fluorescence spectroscopy
3. 1D ^{1}H NMR spectroscopy
4. ligand-binding assay

R14ρ, F25φ, N26ν, Q33θ
only stable upon addition of TMAO (trimethylamine N-oxide), no melting temperature reported

Different quantities reflecting fold stability need not show high correlation

Interpretation of Experimental Data using Simulation

Reasons for disagreement between $\langle Q \rangle_{sim}$ and $\langle Q \rangle_{exp}$

Failure to observe a correlation between the simulation and experiment can be due to many reasons:

D1 The simulation is insufficiently accurate: i.e.

a) relevant degrees of freedom were omitted;

b) the force field was insufficiently accurate;

c) approximations made when solving the equations of motion were too crude;

d) inappropriate thermodynamic or spatial boundary conditions were used.

D2 The measured $<Q>_{exp}$ is inaccurate

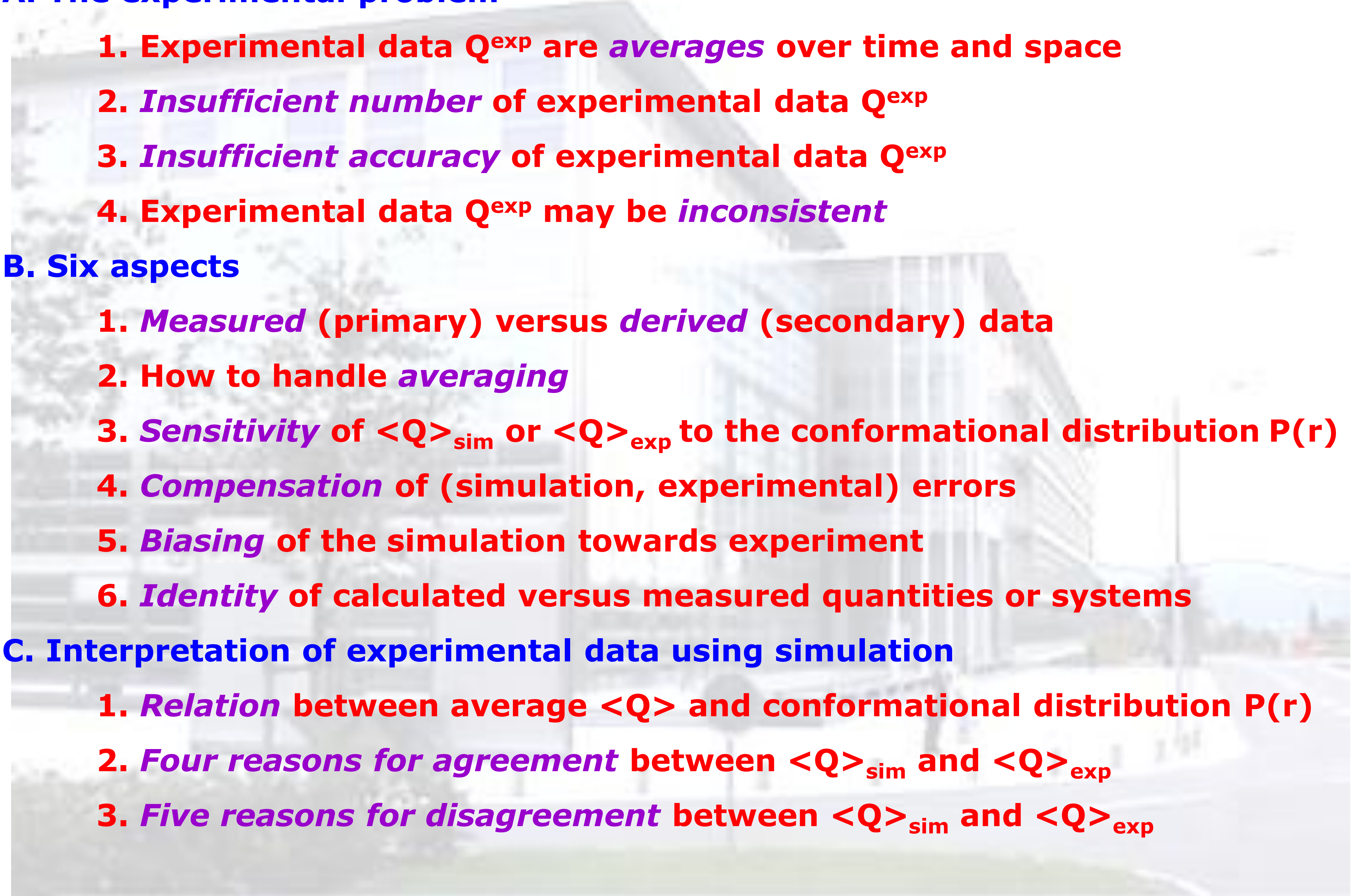
D3 $<Q>_{sim}$ and $<Q>_{exp}$ are averaged differently with respect to time or spatial extent

D4 Related but different quantities are compared, e.g. atom-positional fluctuations versus crystallographic B factors

D5 Different systems are compared (e.g. crystal versus solution), or systems studied under different thermodynamic conditions (e.g. temperature, pressure, pH, ionic strength, etc.)

*W.F. van Gunsteren, J. Dolenc, & A.E. Mark, Curr. Opin. Struct. Biol., 18 (**2008**) 149-153*

On comparing molecular modelling results with experimental data



A. The experimental problem

1. Experimental data Q^{exp} are *averages* over time and space
2. *Insufficient number* of experimental data Q^{exp}
3. *Insufficient accuracy* of experimental data Q^{exp}
4. Experimental data Q^{exp} may be *inconsistent*

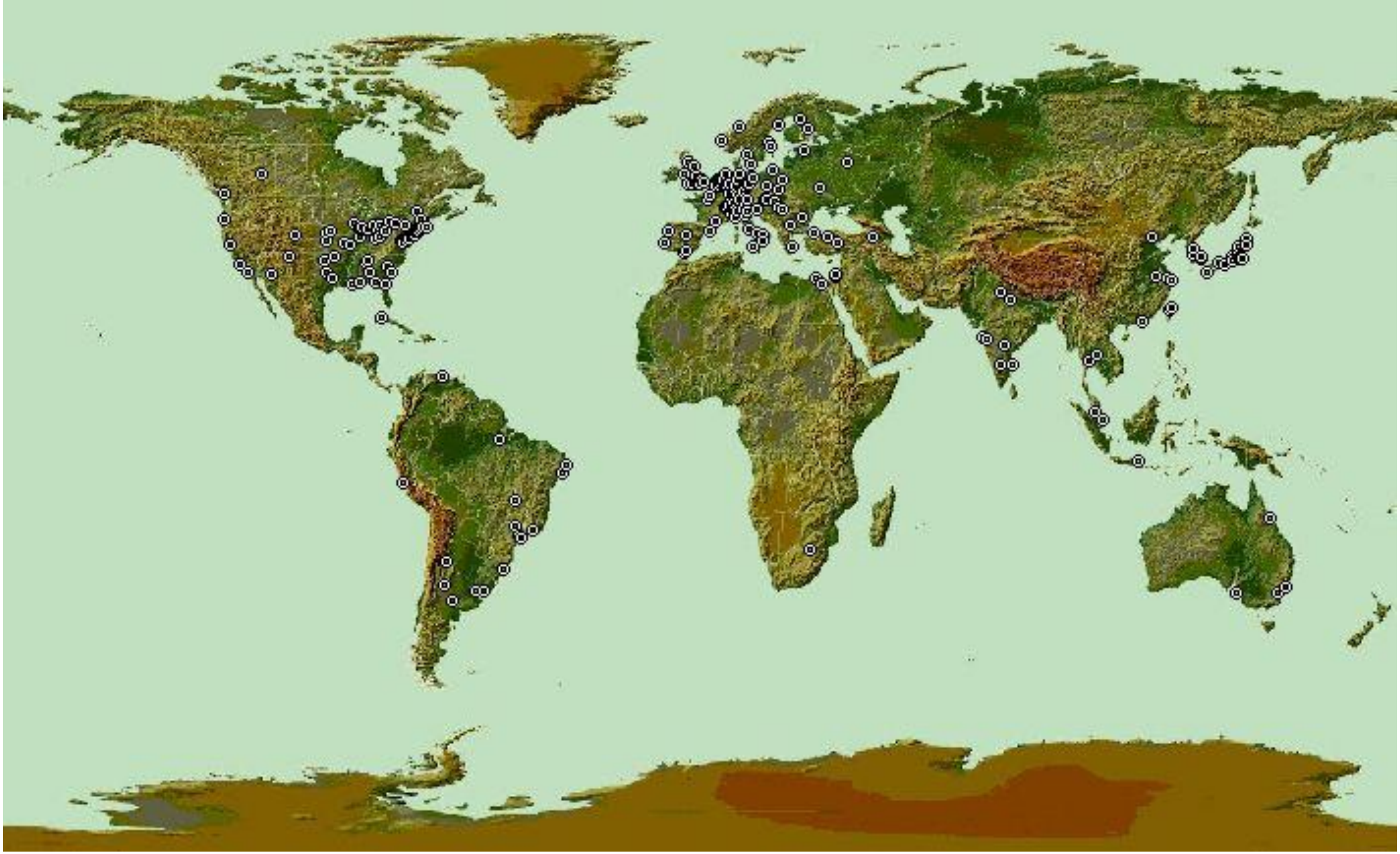
B. Six aspects

1. *Measured* (primary) versus *derived* (secondary) data
2. How to handle *averaging*
3. *Sensitivity* of $\langle Q \rangle_{sim}$ or $\langle Q \rangle_{exp}$ to the conformational distribution P(r)
4. *Compensation* of (simulation, experimental) errors
5. *Biasing* of the simulation towards experiment
6. *Identity* of calculated versus measured quantities or systems

C. Interpretation of experimental data using simulation

1. *Relation* between average $\langle Q \rangle$ and conformational distribution P(r)
2. *Four reasons for agreement* between $\langle Q \rangle_{sim}$ and $\langle Q \rangle_{exp}$
3. *Five reasons for disagreement* between $\langle Q \rangle_{sim}$ and $\langle Q \rangle_{exp}$

Spatial distribution of licences
GROMOS biomolecular simulation software



GROMOS = Groningen Molecular Simulation + GROMOS Force Field

Generally available: http://www.gromos.net