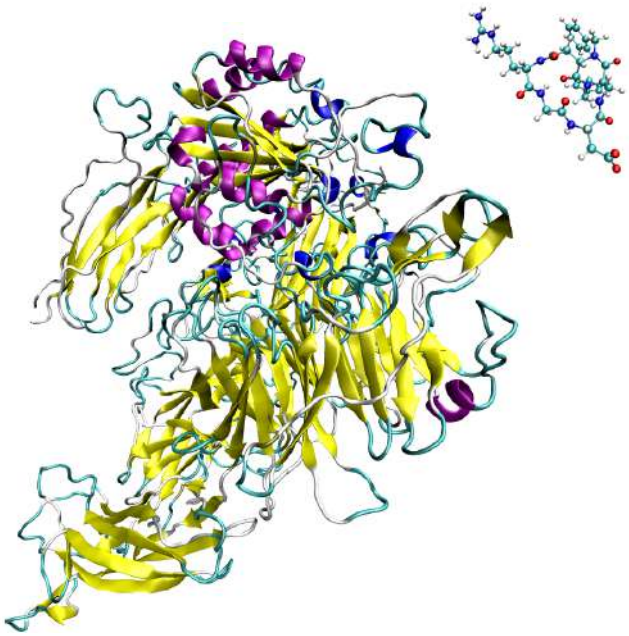
CENTER FOR
COMPLEXITY
& BIOSYSTEMS
University of Milan

UNIVERSITÀ DEGLI STUDI DI MILANO
DIPARTIMENTO DI FISICA

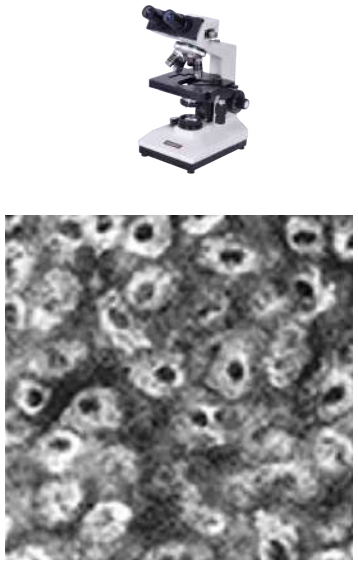
Maximum-entropy modelling of biomolecules

Congress of the Department of Physics
29th June 2017

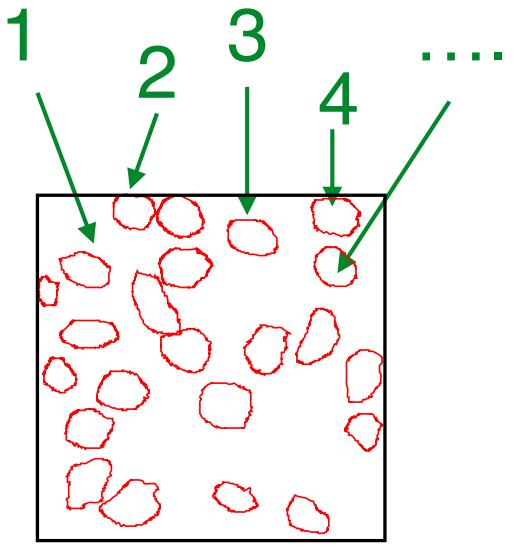


Y. Zhan
L. Giorgetti
F. Vasile
M. Civera
L. Belvisi
D. Potenza
C. Camilloni
GT

The problem

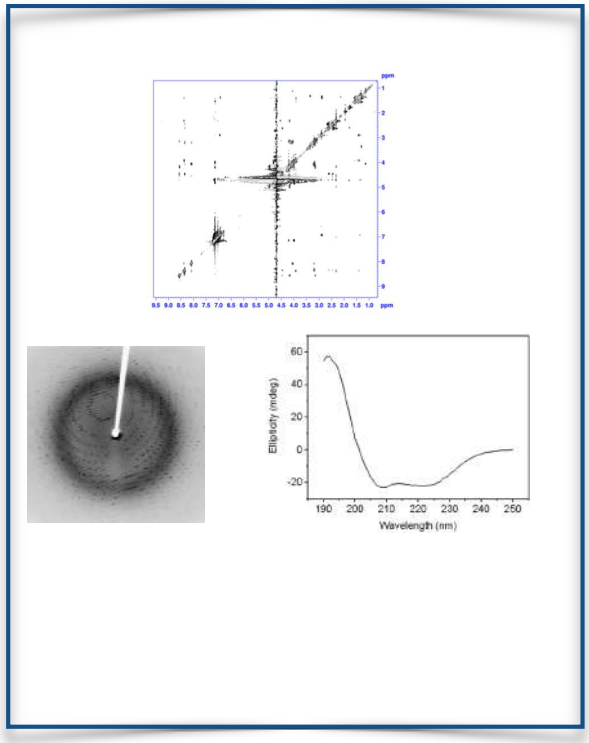


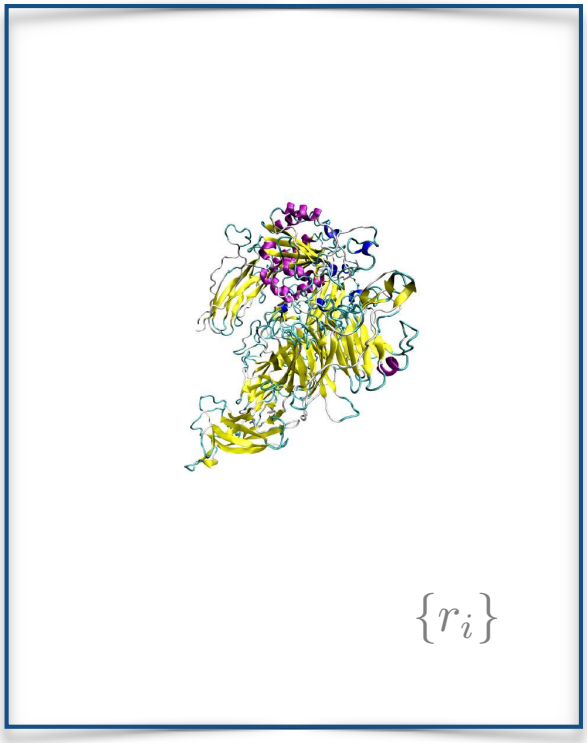
Raw data



Model

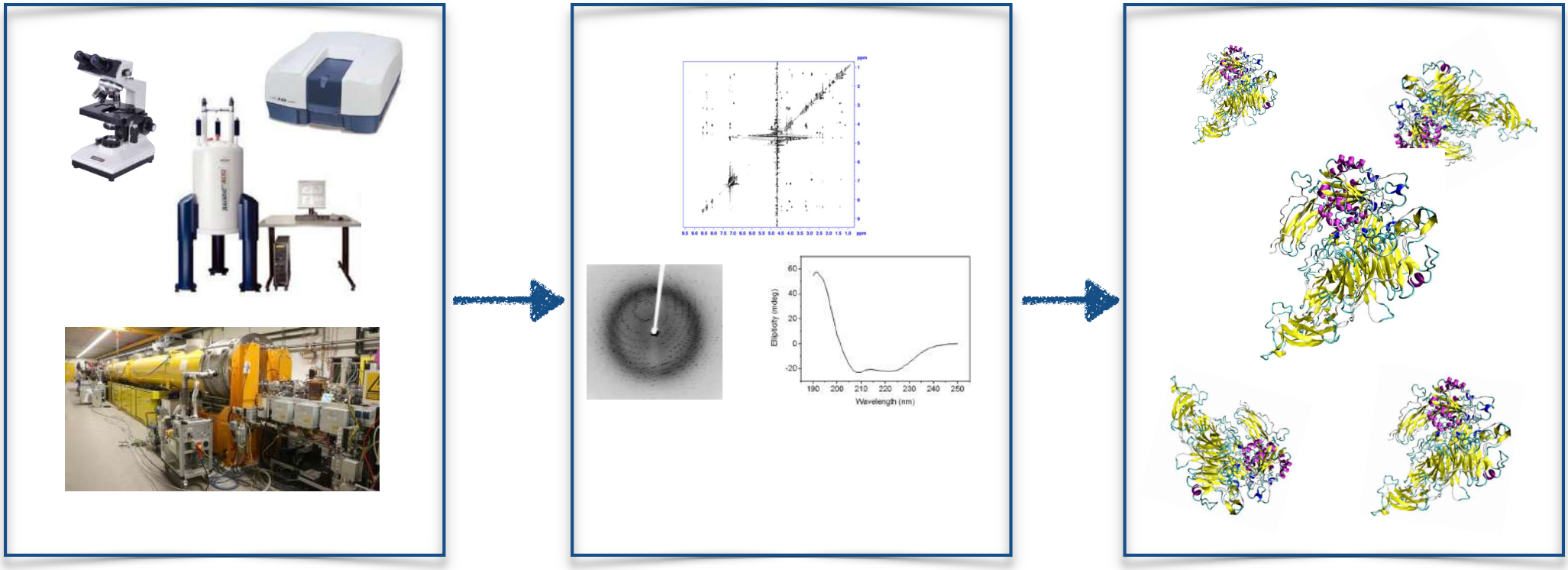
The problem





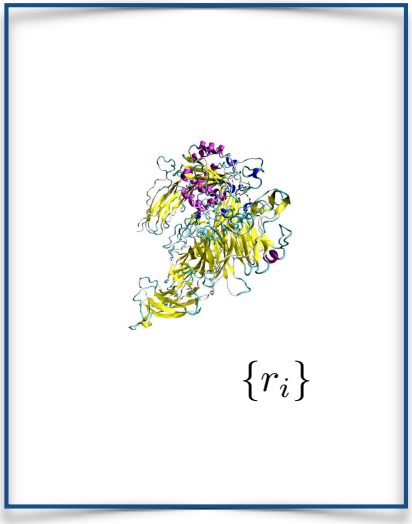
$\{r_i\}$

The problem



- Experimental data are (equilibrium) average over 10^{23} molecules
- Often biomolecules are flexible

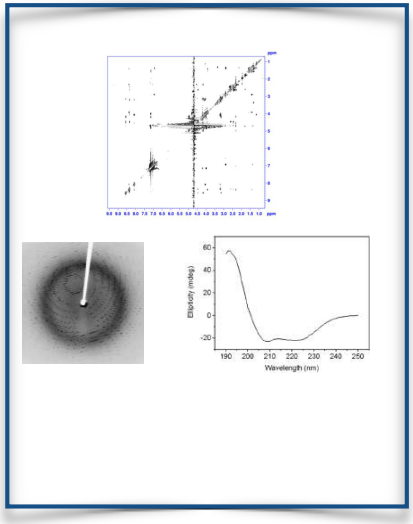
The problem

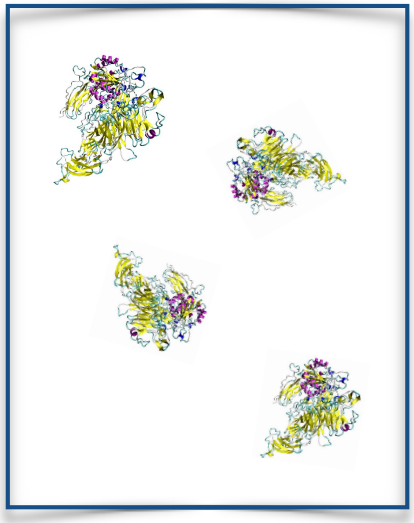


$\{r_i\}$

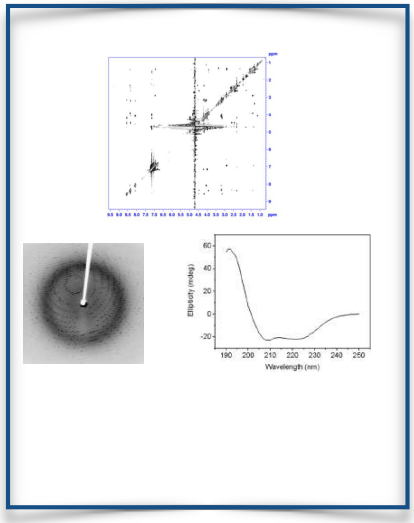
$f(\{r_i\})$

"forward model"

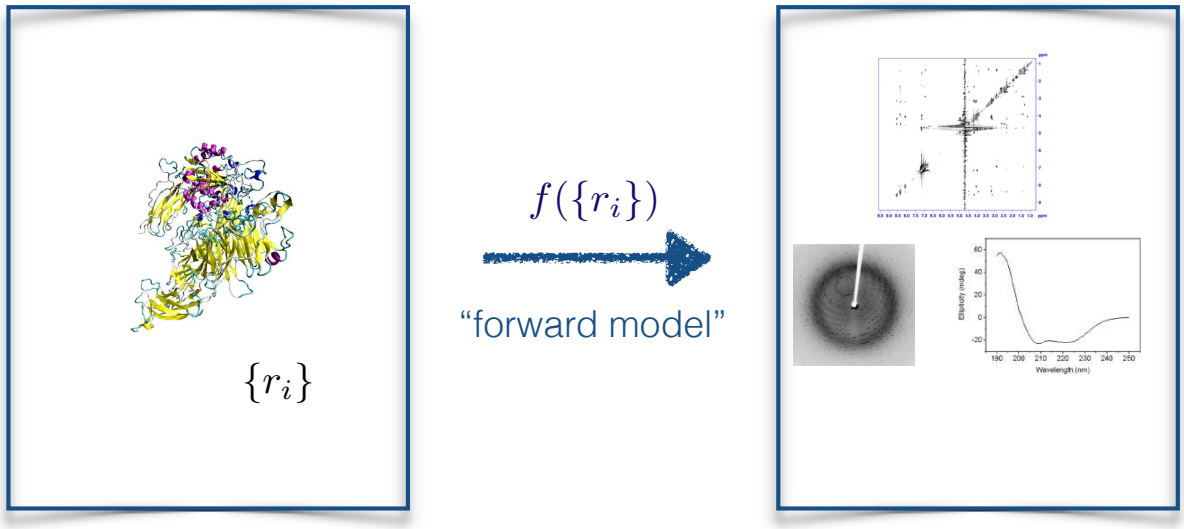


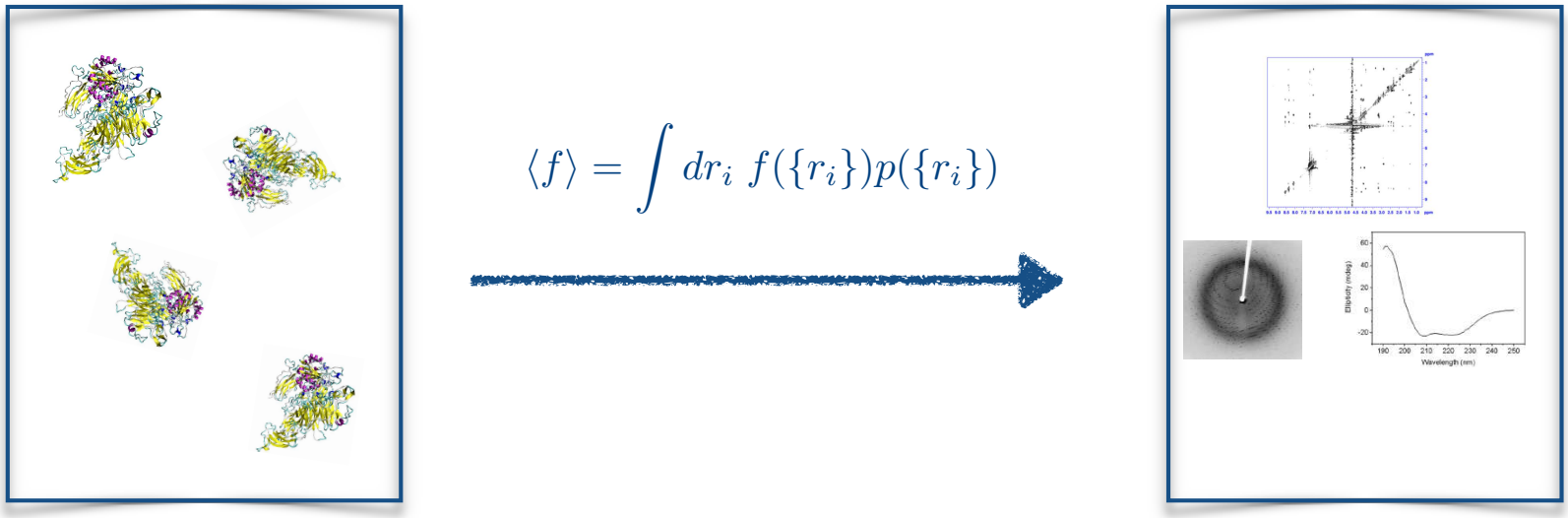


$$\langle f \rangle = \int dr_i \; f(\{r_i\}) p(\{r_i\})$$



The problem

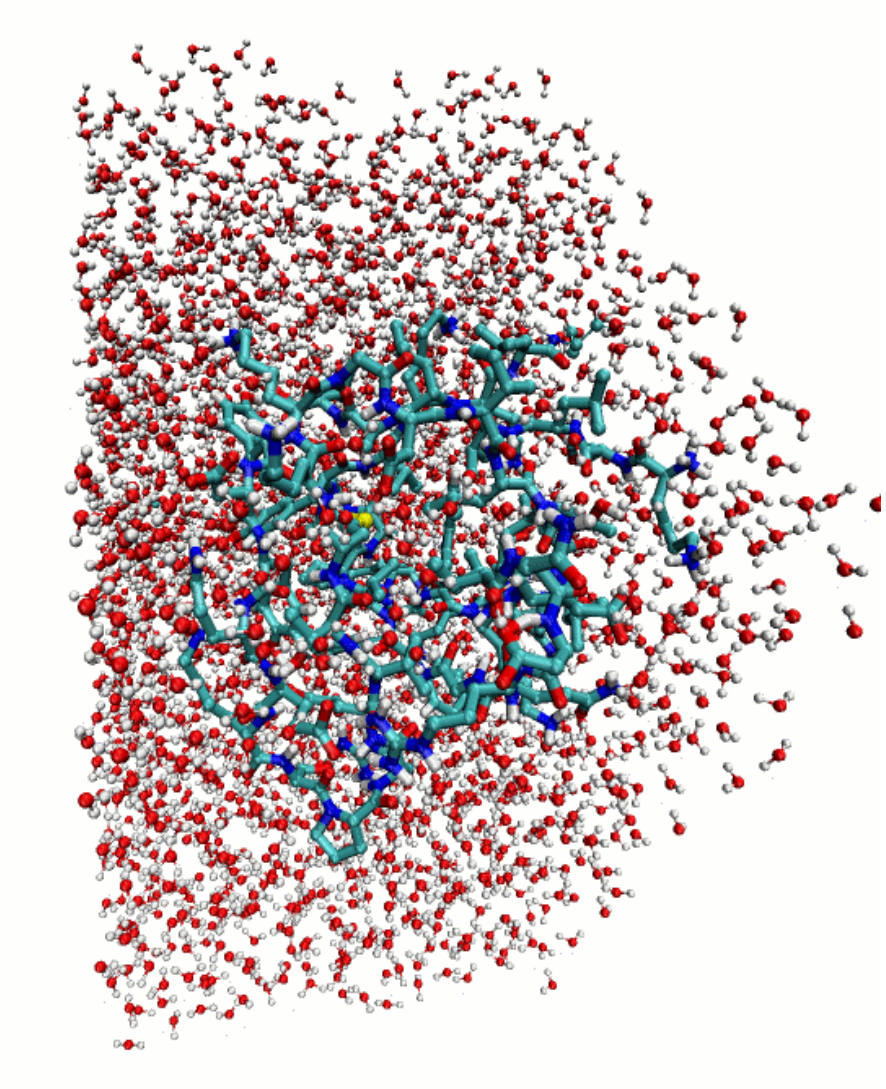




?

Inverse problem

We use simplified models

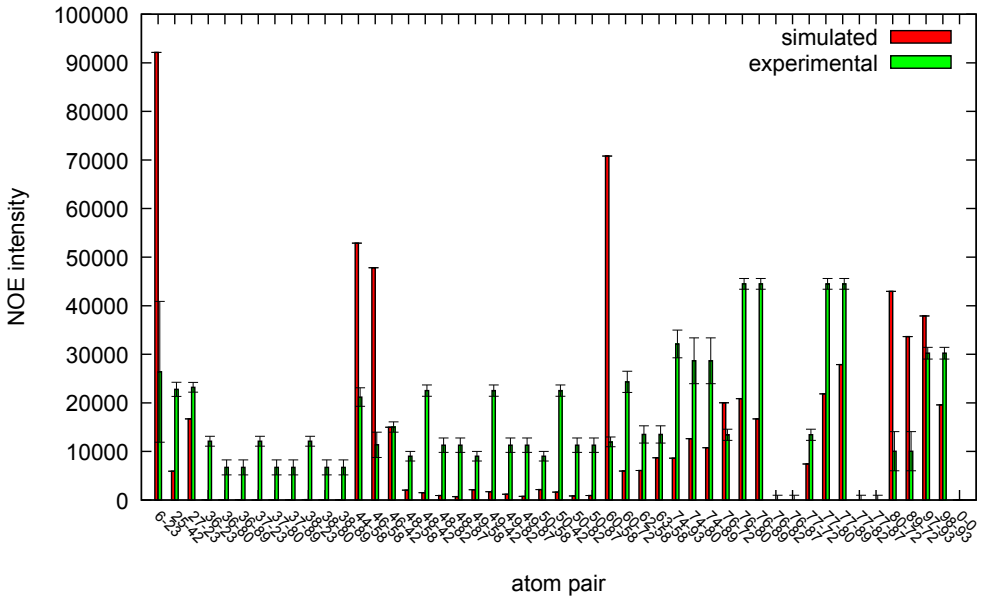


Amber03

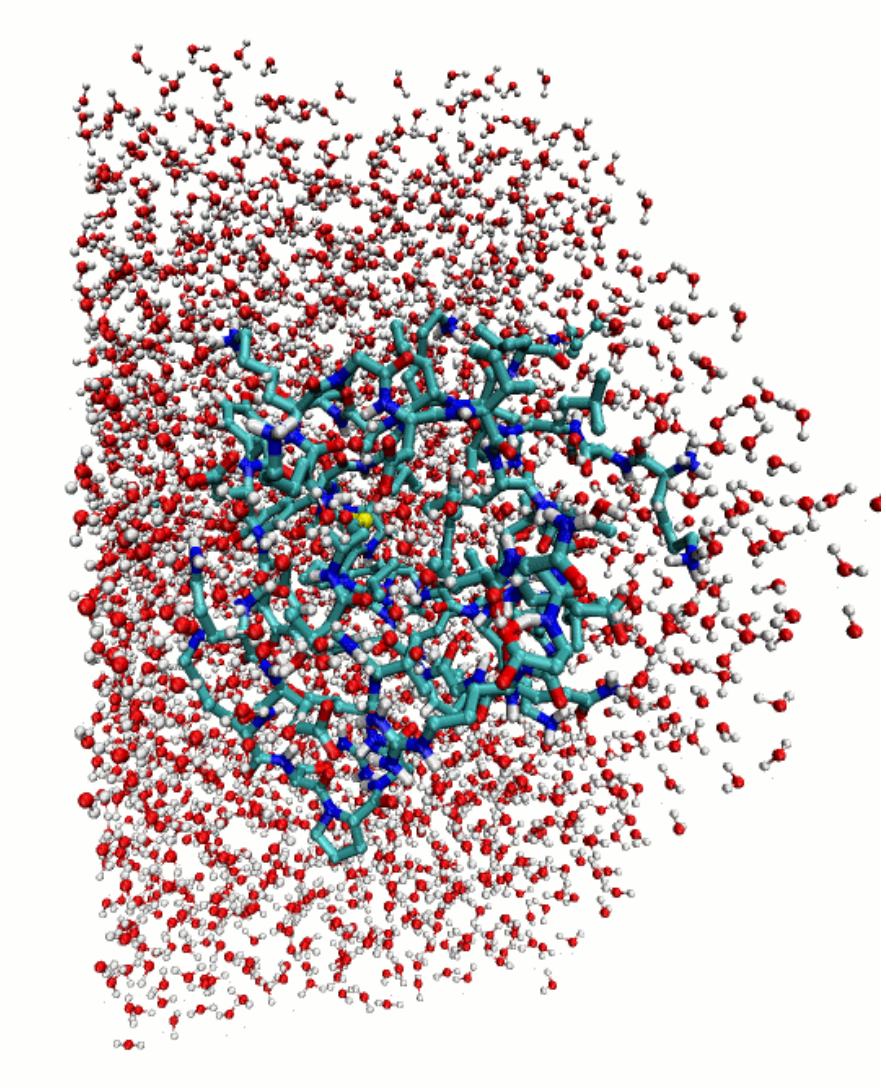
Recipe:

- define your degrees of freedom
- define an effective potential $U(r)$
- most experimental data are at equilibrium

$$p(r) = \frac{1}{Z} \exp\left[-\frac{U(r)}{kT}\right]$$



We use simplified models

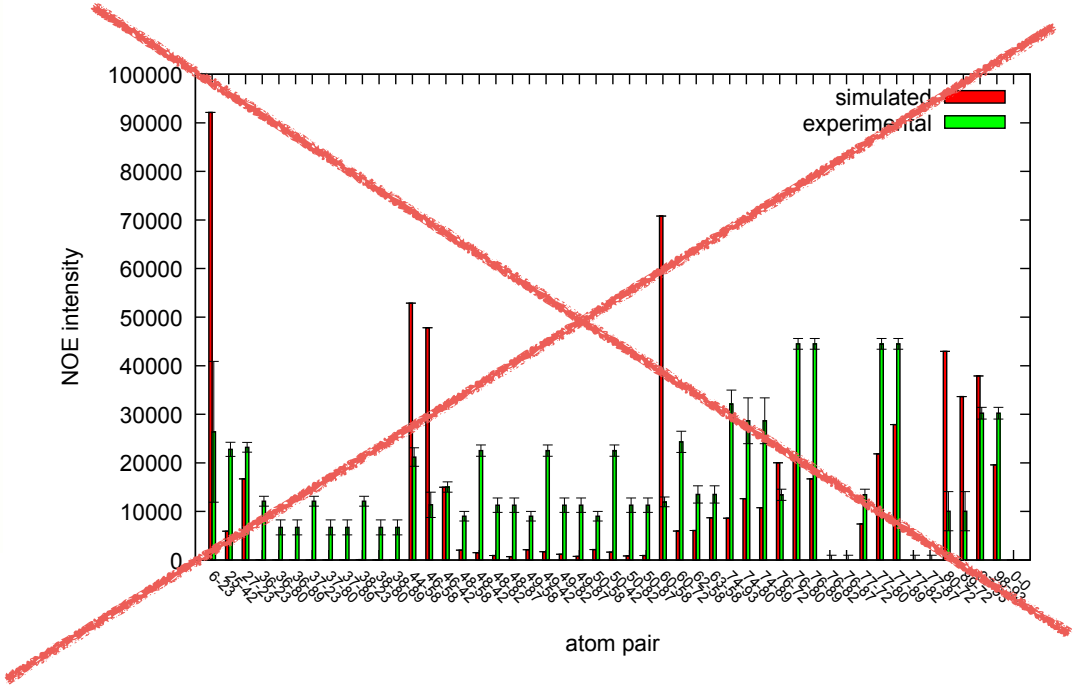


Amber03

Recipe:

- define your degrees of freedom
- define an effective potential $U(r)$
- most experimental data are at equilibrium

$$p(r) = \frac{1}{Z} \exp\left[-\frac{U(r)}{kT}\right]$$



Change strategy: from data to model

Information Theory and Statistical Mechanics

E. T. JAYNES
Department of Physics, Stanford University, Stanford, California
(Received September 4, 1956; revised manuscript received March 4, 1957)

Information theory provides a constructive criterion for setting up probability distributions on the basis of partial knowledge, and leads to a type of statistical inference which is called the maximum-entropy estimate. It is the least biased estimate possible on the given information; i.e., it is maximally noncommittal with regard to missing information. If one considers statistical mechanics as a form of statistical inference rather than as a physical theory, it is found that the usual computational rules, starting with the determination of the partition function, are an immediate consequence of the maximum-entropy principle. In the resulting "subjective statistical mechanics," the usual rules are thus justified independently of any physical argument, and in particular independently of experimental verification; whether or not the results agree with experiment, they still represent the best estimates that could have been made on the basis of the information available.

It is concluded that statistical mechanics need not be regarded as a physical theory dependent for its validity on the truth of additional assumptions not contained in the laws of mechanics (such as ergodicity, metric transitivity, equal *a priori* probabilities, etc.). Furthermore, it is possible to maintain a sharp distinction between its physical and statistical aspects. The former consists only of the correct enumeration of the states of a system and their properties; the latter is a straightforward example of statistical inference.

1. INTRODUCTION

THE recent appearance of a very comprehensive survey[1] of past attempts to justify the methods of statistical mechanics in terms of mechanics, classical or quantum, has helped greatly, and at a very opportune time, to emphasize the unsolved problems in this field. Although the subject has been under development for many years, we still do not have a complete and satisfactory theory, in the sense that there is no line of argument proceeding from the laws of microscopic mechanics to macroscopic phenomena, that is generally regarded by physicists as convincing in all respects. Such an argument should (a) be free from objection on mathematical grounds, (b) involve no additional arbi-

[1] D. ter Haar, Revs. Modern Phys. **27**, 289 (1955).

model $\longrightarrow p(r_i)$

Change strategy: from data to model

Information Theory and Statistical Mechanics

E. T. JAYNES
Department of Physics, Stanford University, Stanford, California
(Received September 4, 1956; revised manuscript received March 4, 1957)

Information theory provides a constructive criterion for setting up probability distributions on the basis of partial knowledge, and leads to a type of statistical inference which is called the maximum-entropy estimate. It is the least biased estimate possible on the given information; i.e., it is maximally noncommittal with regard to missing information. If one considers statistical mechanics as a form of statistical inference rather than as a physical theory, it is found that the usual computational rules, starting with the determination of the partition function, are an immediate consequence of the maximum-entropy principle. In the resulting "subjective statistical mechanics," the usual rules are thus justified independently of any physical argument, and in particular independently of experimental verification; whether or not the results agree with experiment, they still represent the best estimates that could have been made on the basis of the information available.

It is concluded that statistical mechanics need not be regarded as a physical theory dependent for its validity on the truth of additional assumptions not contained in the laws of mechanics (such as ergodicity, metric transitivity, equal *a priori* probabilities, etc.). Furthermore, it is possible to maintain a sharp distinction between its physical and statistical aspects. The former consists only of the correct enumeration of the states of a system and their properties; the latter is a straightforward example of statistical inference.

1. INTRODUCTION

THE recent appearance of a very comprehensive survey[1] of past attempts to justify the methods of statistical mechanics in terms of mechanics, classical or quantum, has helped greatly, and at a very opportune time, to emphasize the unsolved problems in this field.

[1] D. ter Haar, Revs. Modern Phys. **27**, 289 (1955).

Although the subject has been under development for many years, we still do not have a complete and satisfactory theory, in the sense that there is no line of argument proceeding from the laws of microscopic mechanics to macroscopic phenomena, that is generally regarded by physicists as convincing in all respects. Such an argument should (a) be free from objection on mathematical grounds, (b) involve no additional arbi-

model $\longrightarrow p(r_i)$

$$S[p] = -\int dr_i p(r_i) \log p(r_i)$$

Change strategy: from data to model

- The distribution $p(r_i)$ which maximises the entropy $S[p] = -\int dr_i p(r_i) \log p(r_i)$ under to the constrains $\langle f_m \rangle = f_m^{exp}$ is given by

$$\frac{\delta}{\delta p(r_i)} \left(-\int dr_i p(r_i) \log p(r_i) - \sum_m \lambda_m \left[\int dr_i f_m(r_i) p(r_i) - f_m^{exp} \right] - \mu \left[\int dr_i p(r_i) - 1 \right] \right) = 0$$

that is,

$$p(r_i) = e^{-1-\mu} \exp \left[-\sum_m \lambda_m f_m(r_i) \right]$$

- If we assume that $p(r_i)$ is an equilibrium distribution, then

$$p(r_i) = \frac{1}{Z} \exp \left[-\frac{U(r_i)}{kT} \right]$$

- Consequently,

$$U(r_i) = -kT \sum_m \lambda_m f_m(r_i)$$

where

$$\frac{d}{d\lambda_m} \log Z = f_m^{exp}$$

Change strategy: from data to model

- The distribution $p(r_i)$ which minimises the entropy $S[p] = -\int dr_i p(r_i) \log p(r_i)$ under to the constrains $\langle f_m \rangle = f_m^{exp}$ is given by

$$\frac{\delta}{\delta p(r_i)} \left(-\int dr_i p(r_i) \log p(r_i) - \sum_m \lambda_m \left[\int dr_i f_m(r_i) p(r_i) - f_m^{exp} \right] - \mu \left[\int dr_i p(r_i) - 1 \right] \right) = 0$$

that is,

$$p(r_i) = e^{-1-\mu} \exp \left[-\sum_m \lambda_m f_m(r_i) \right]$$

- If we assume that $p(r_i)$ is an equilibrium distribution, then

$$p(r_i) = \frac{1}{Z} \exp \left[-\frac{U(r_i)}{kT} \right]$$

- Consequently,

$$U(r_i) = -kT \sum_m \lambda_m f_m(r_i)$$

where

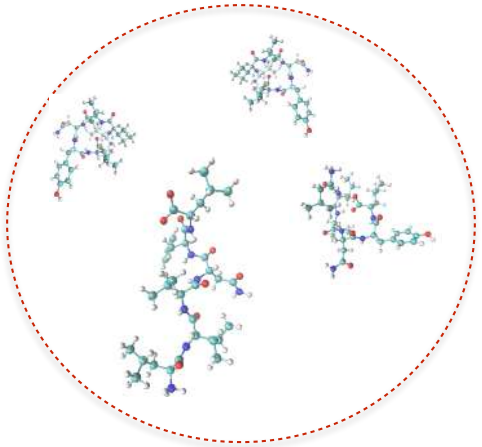
$$\frac{d}{d\lambda_m} \log Z = f_m^{exp}$$

- The potential has the same functional form of the forward model

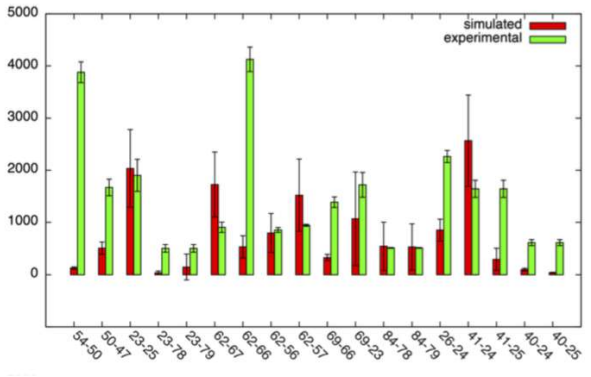
- The Lagrange multipliers are hard to find

Computational implementation

$U_0(r_i)$ MD simulation



back-calculate data



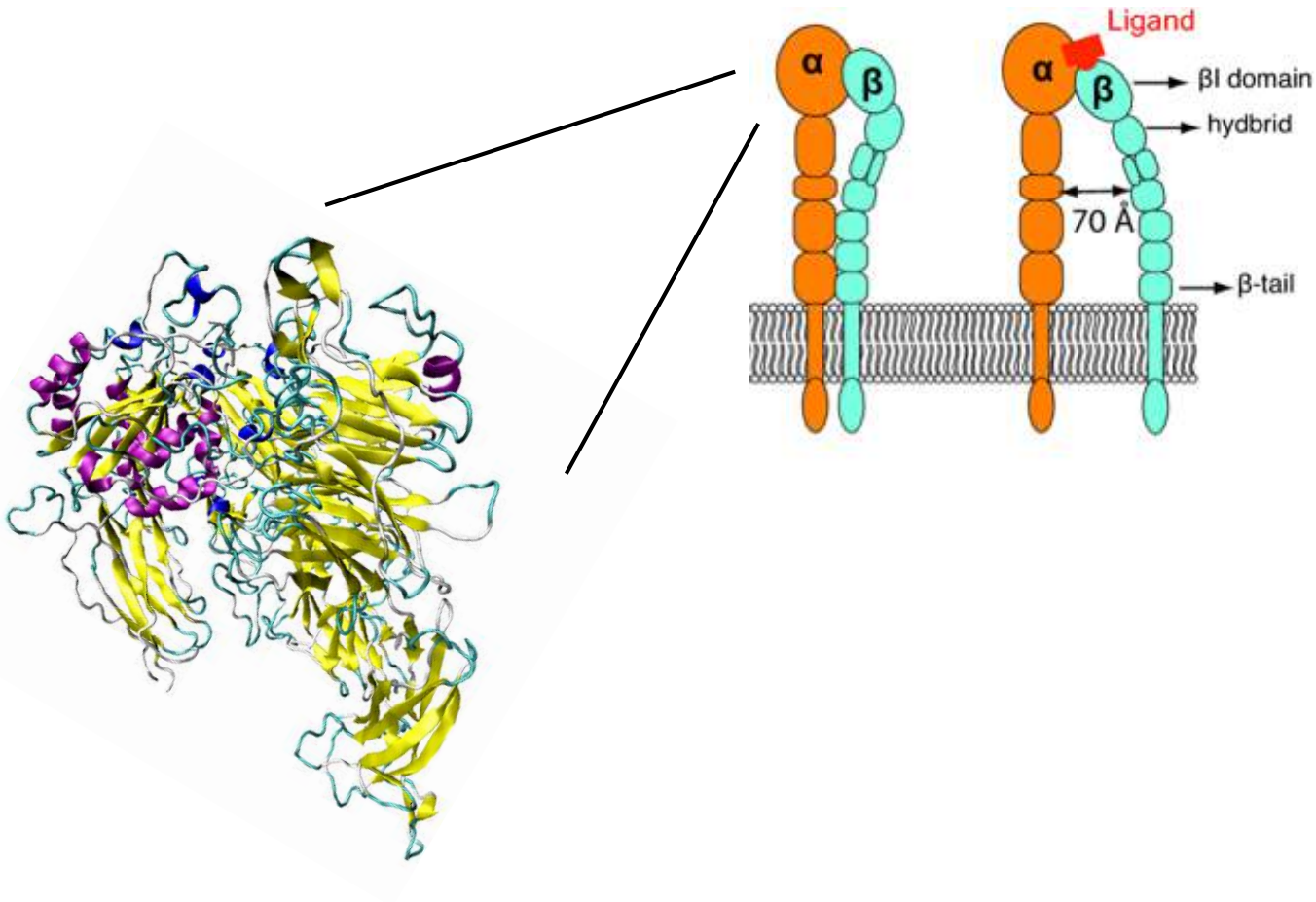
re-weight sampled conformations

minimize χ^2

MD simulation

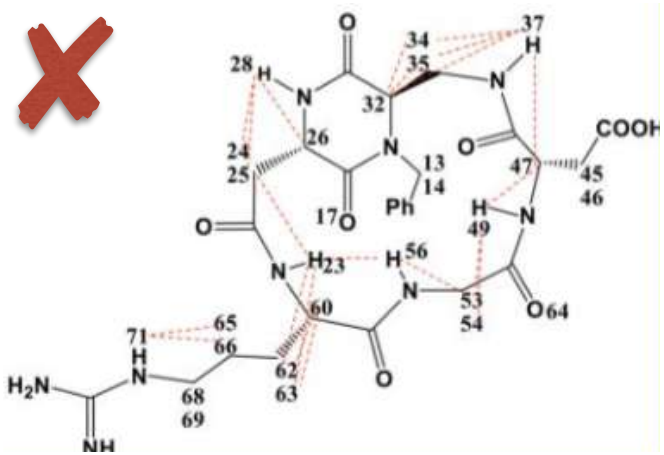
$U_0(r_i) - kT \sum_m \lambda_m f_m(r_i)$

An application: Integrin ligands

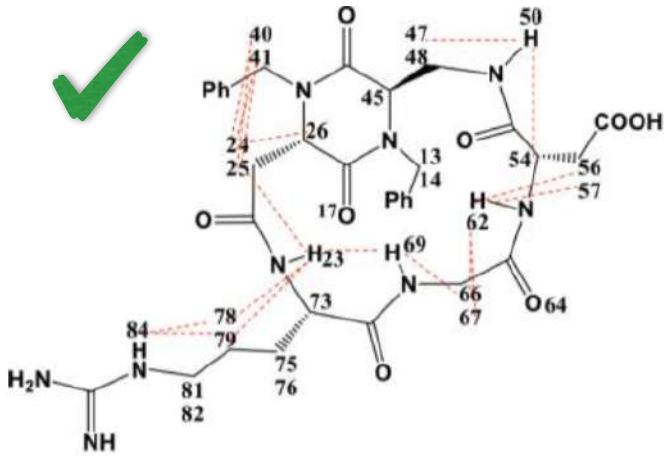


An application: Integrin ligands

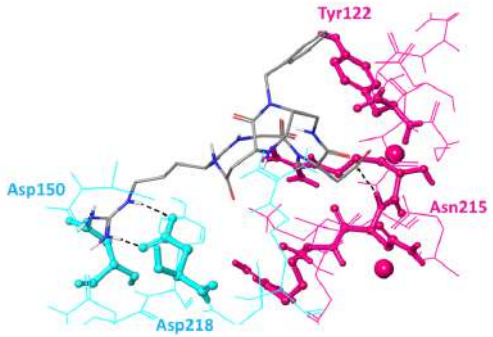
de Ressurreicao *et al. Chem Eur. J.* (2009)

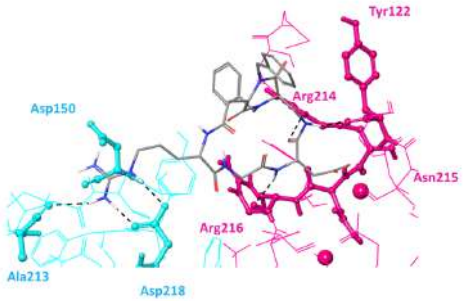


DKP2



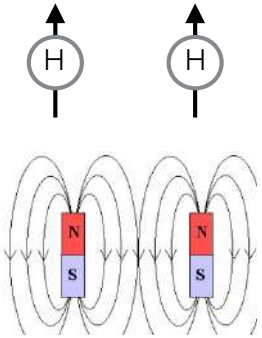
DKP5





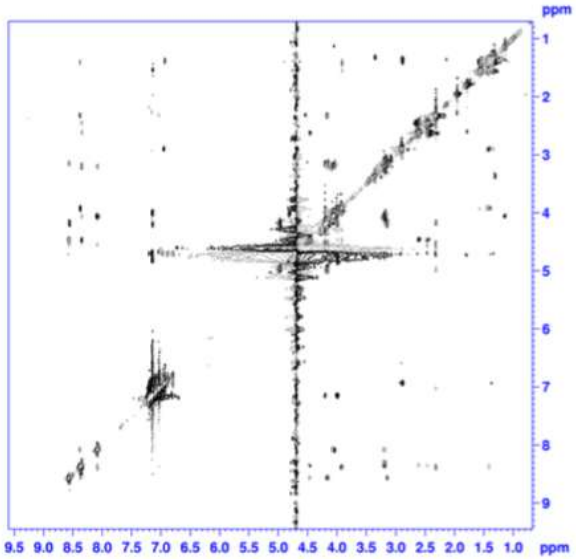
NMR provides raw conformational data

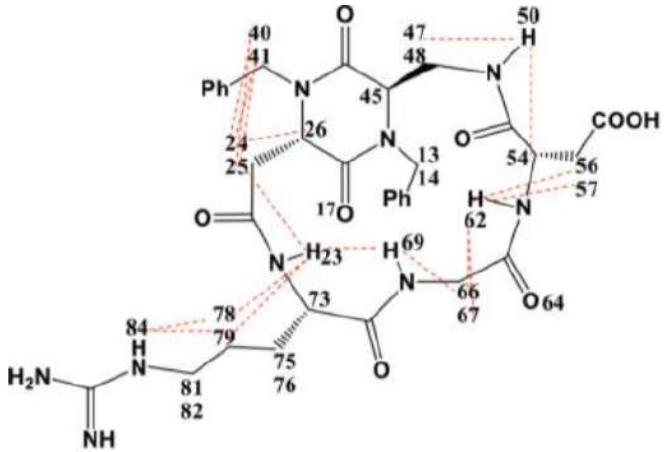
Nuclear Overhauser Effect (NOE)

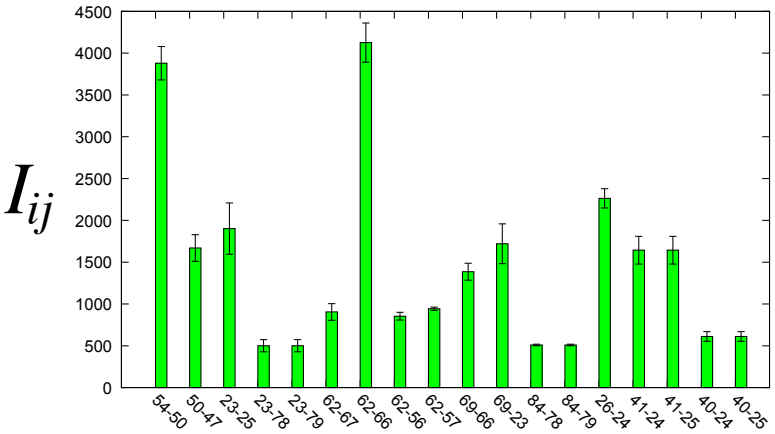


dipolar coupling

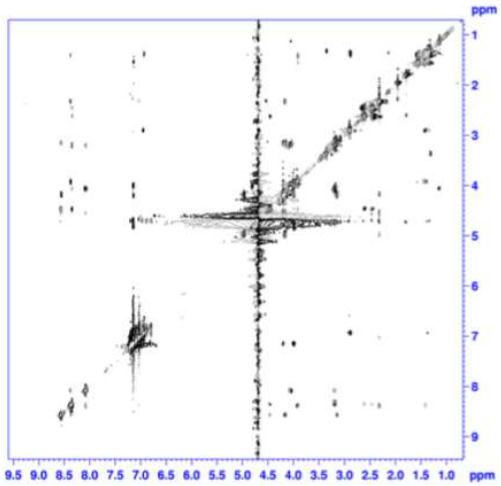
$$I_{ij} = I_0 \left\langle \frac{1}{d_{ij}^6} \right\rangle$$







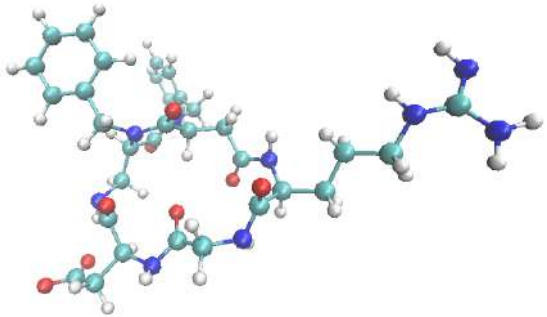
The standard interpretation of NOE intensities



raw NOE intensities

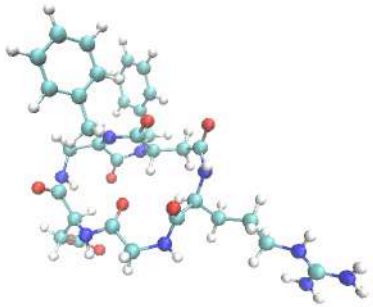
$$\left\langle \frac{1}{d_{ij}^6} \right\rangle \approx \frac{1}{\langle d_{ij} \rangle^6}$$

"mean-field approximation"



"mean-field conformation"

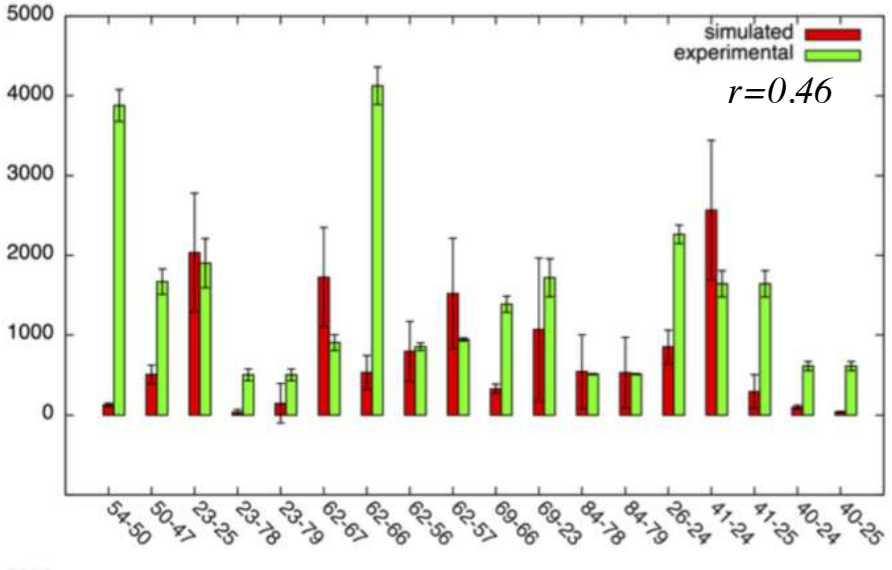
Molecular dynamics simulations…



forces from GAFF

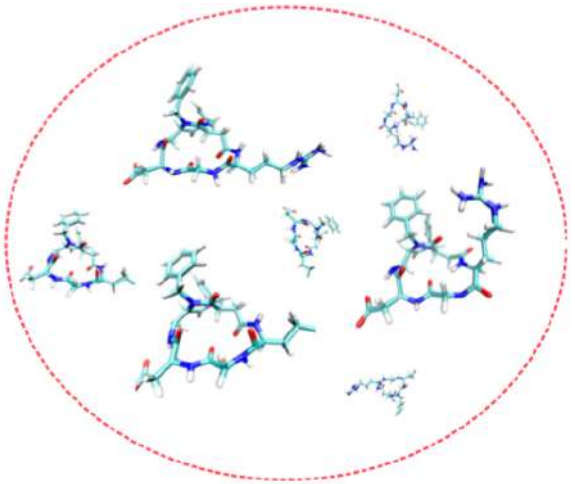
(Wang et al. 2006)

$$I_{ij} = I_0 \left\langle \frac{1}{d_{ij}^6} \right\rangle$$

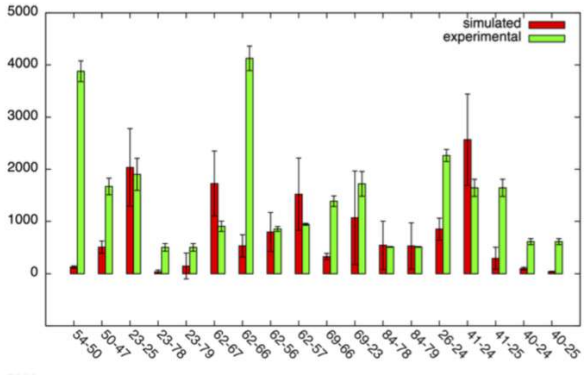


Let's apply the MaxEnt principle

$U_{GAFF}(r)$ MD simulation



back-calculate NOEs



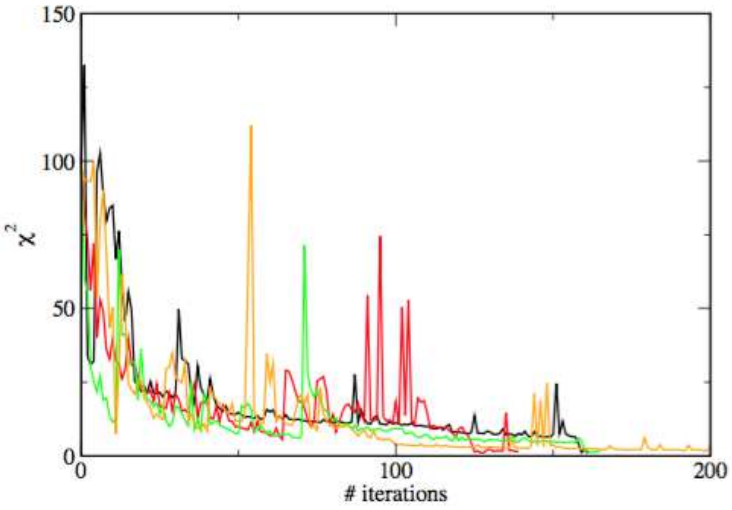
re-weight sampled conformations

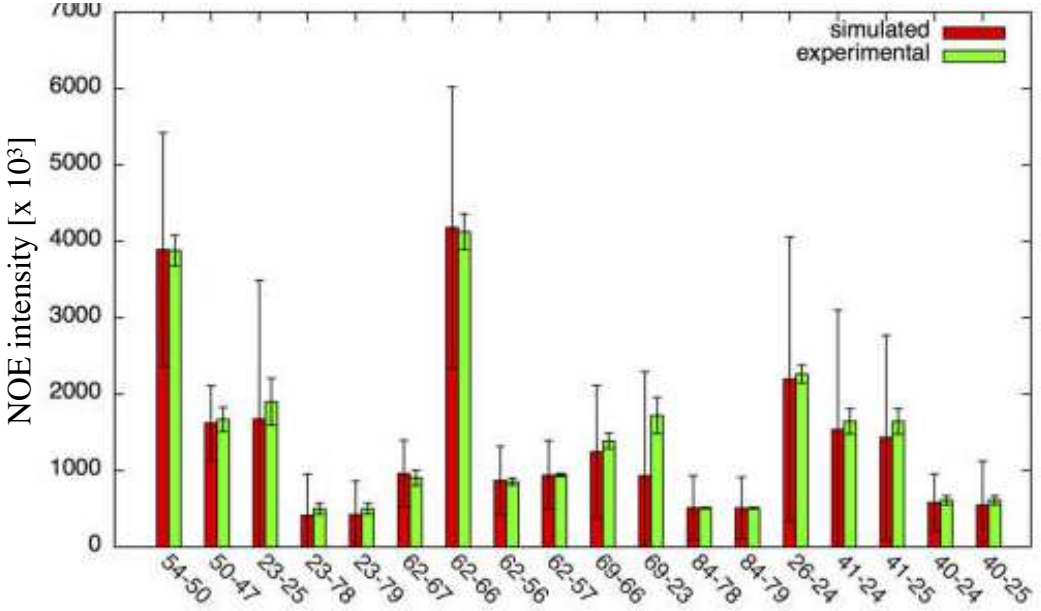
minimize χ^2

MD simulation

$U_{GAFF}(r) + \gamma_k \, \delta U(r)$

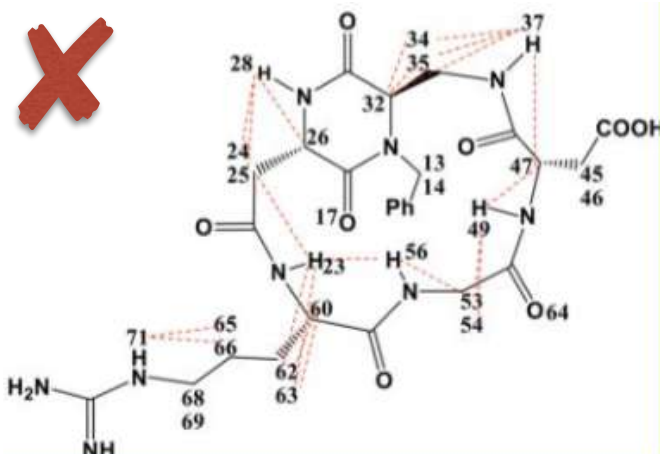
…until convergence….



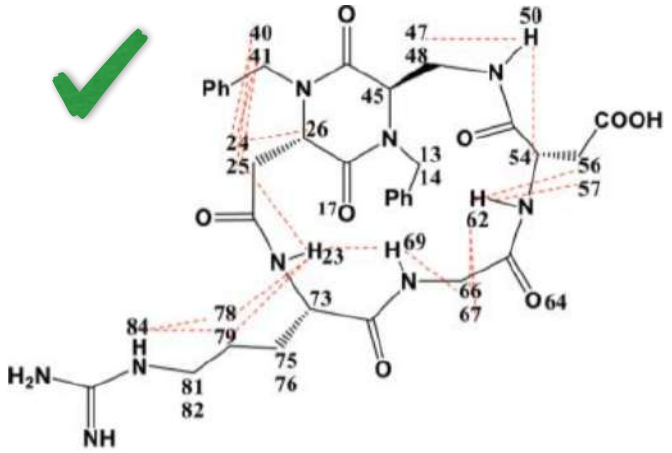


χ^2=1.13

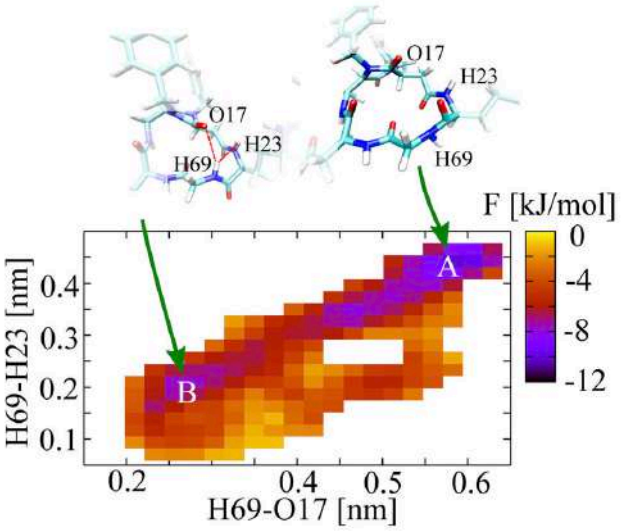
The worse molecule populates more states

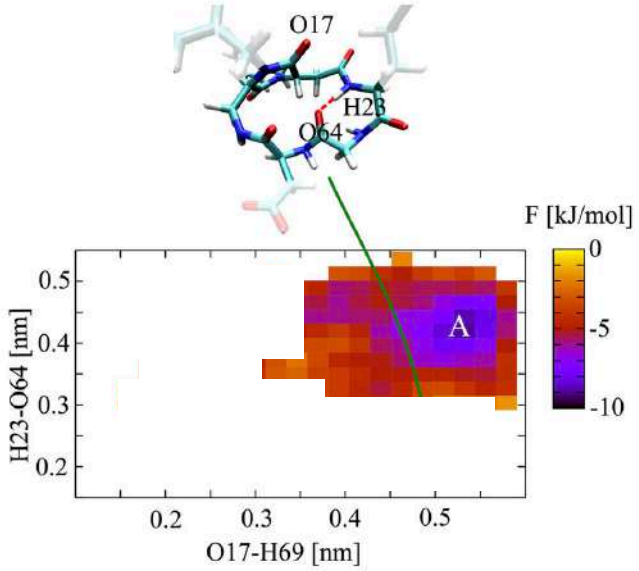


DKP2



DKP5



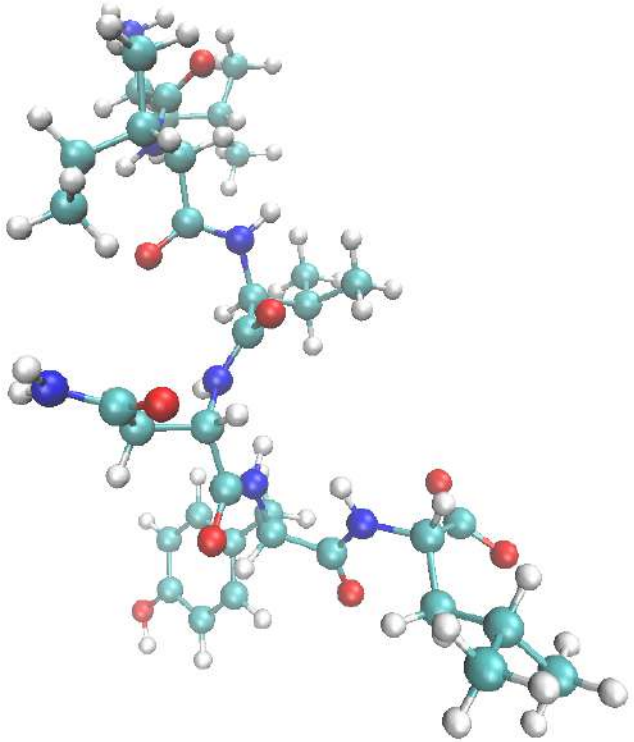


A more complex case: spin diffusion

We assumed $I_{ij} = I_0 \left\langle \frac{1}{d_{ij}^6} \right\rangle$, but this is true only if $\tau_m \to 0$.

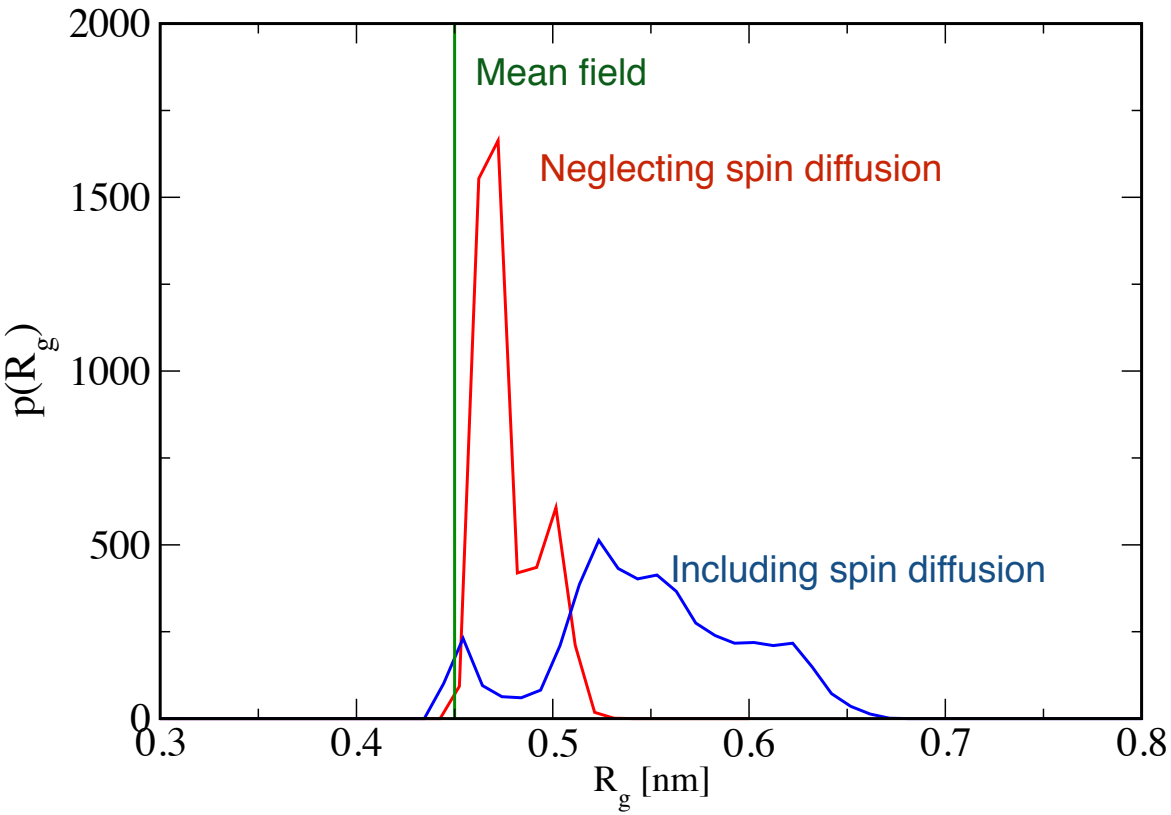
In general, $I_{ij}(\tau_m) = e^{-W\tau_m} I_{ij}(0)$

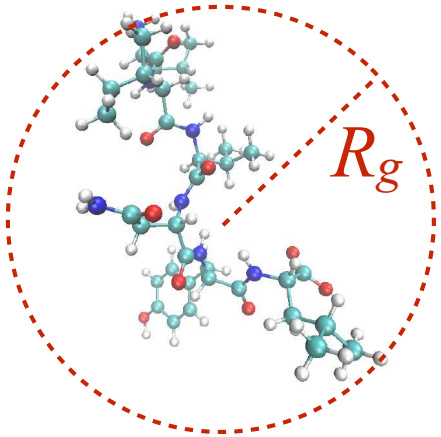
where $W_{ij} = \overline{\langle i|\hat{H}_{rel}|j\rangle^2} \sim \frac{1}{d_{ij}^6}$



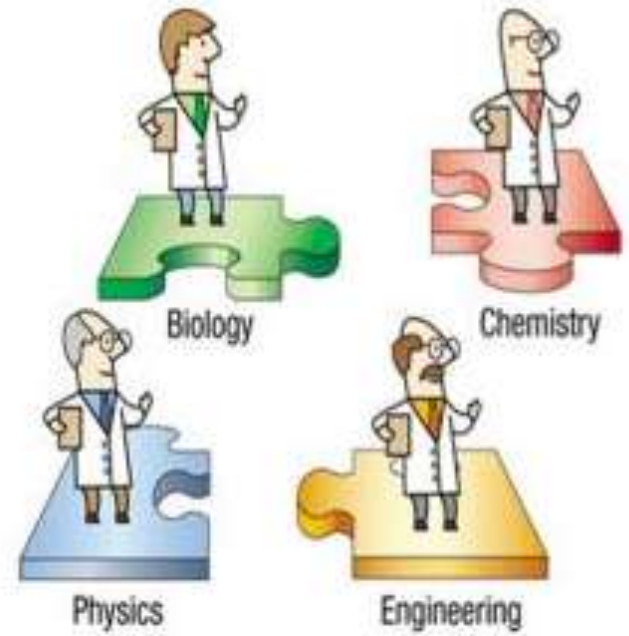
→ The forward model is much heavier

The model changes a lot….





Take home message….



Reitner, Hadjichristidid & Moller, Nature Materials (2003)