Infraestrutura Nacional de **Computação Distribuída**

FMUP FACULDADE DE MEDICINA UNIVERSIDADE DO PORTO

NOVA UNIVERSIDADE NOVA DE LISBOA · NOVA SCHOOL OF SCIENCE & TECHNOLOGY · UCIBIO · i4HB Institute for Health and Bioeconomy

Biomolecular Engineering Lab · bioSIM

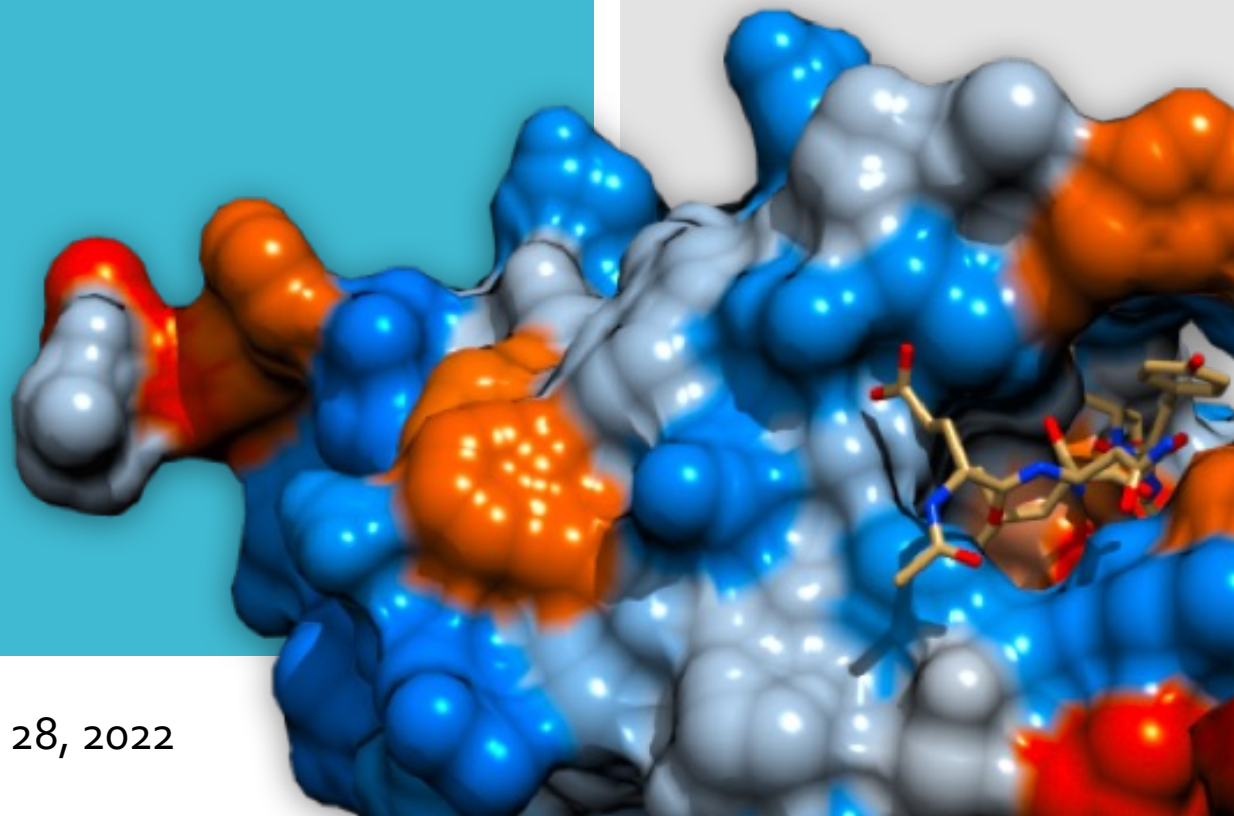
Computational Biosciences Using HPC Systems – Module 5

Molecular Dynamics Simulations

Sérgio F. Sousa & Arménio Barbosa

Email: sergiofsousa@med.up.pt
UCIBIO/REQUIMTE, BioSIM – Departamento de Biomedicina
Faculdade de Medicina da Universidade do Porto, Portugal

Email: aj.Barbosa@fct.unl.pt
UCIBIO-i4HB, Biomolecular Engineering Lab – Departamento de Química
School of Science and Technology, NOVA University Lisbon, Portugal



www.biosim.pt

https://sites.fct.unl.pt/biomolecular_eng/

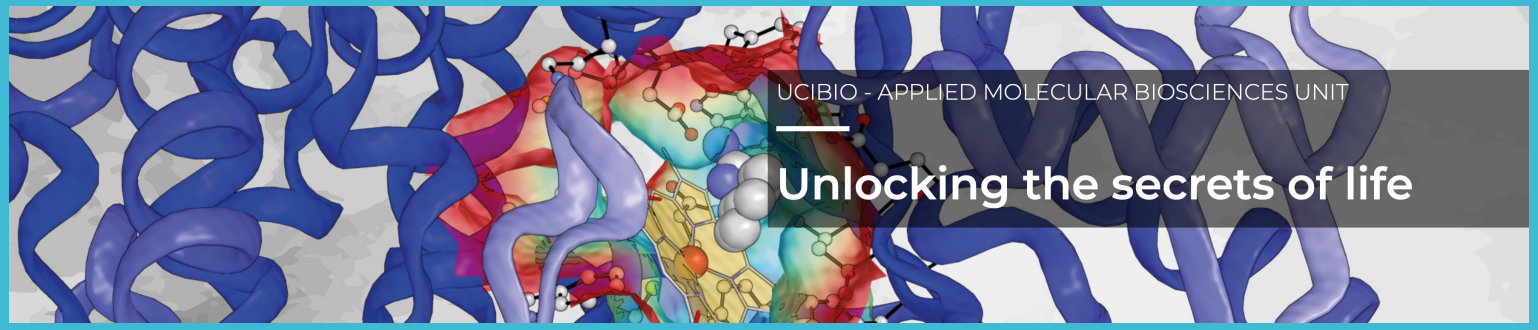
September 28, 2022

UCIBIO
Applied Molecular Biosciences Unit
@REQUIMTE

UCIBIO is formed by the collaborative efforts of researchers from the **University of Porto** and **University NOVA of Lisbon**.

UCIBIO´s research activities occur mainly at the campus of Faculty of Sciences and Technology of the NOVA University of Lisbon (**FCT-NOVA**), and at the Faculty of Sciences (**FCUP**), the Faculty of Pharmacy (**FFUP**), the Faculty of Medicine (**FMUP**) and Instituto de Ciências Biomédicas Abel Salazar (**ICBAS-UP**) from the University of Porto.

UCIBIO combines key expertise in Chemistry and Biological Sciences with an ambitious strategic plan to maximize its national and international impact in terms of scientific productivity, advanced training and translation to society. In the national context, UCIBIO's key strength lies on its **broad scope of fundamental and applied research, standing at the interface of Chemistry, Biology and Engineering** to address pertinent questions at **atomic, molecular, sub-cellular and cellular levels, including cell-to-cell interactions and population evolutionary dynamics**.



BioSIM Research Group

Location

BioSIM – Biomolecular SIMulations Research Group
UCIBIO/REQUIMTE - Departamento de Biomedicina
Faculdade de Medicina da Universidade do Porto
Portugal
www.biosim.pt

U.PORTO
FMUP FACULDADE DE MEDICINA UNIVERSIDADE DO PORTO

BioSIM
Research Group

Research

BioSIM – Biomolecular SIMulations Research Group

Our research team bridges the gap between theory and experiment applying and developing state-of-the-art computational tools focusing on ***Enzymatic Catalysis, Drug Discovery and Molecular Recognition***

For that we combine: ***QM/MM Methods, Quantum Mechanics, Molecular Dynamics, Docking, Virtual Screening,*** *and* ***Free Energy Perturbation methods****, always in close linking with experiment.*

Several ***Software Applications and Databases*** *have also been developed and made available to the scientific community.*

Website: www.biosim.pt

Biomolecular Engineering Lab

Location

Biomolecular Engineering Lab
UCIBIO-i4HB,– Departamento de Química
School of Science and Technology,
NOVA University Lisbon, Caparica, Portugal
https://sites.fct.unl.pt/biomolecular_eng/

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

Biomolecular Engineering Lab

Research

Biomolecular Engineering Lab

*Composed by a **multidisciplinary team** with expertise in **Applied and Computational Chemistry, Biotechnology, Biomedical Engineering, Chemical Engineering and Physics of Materials**.*

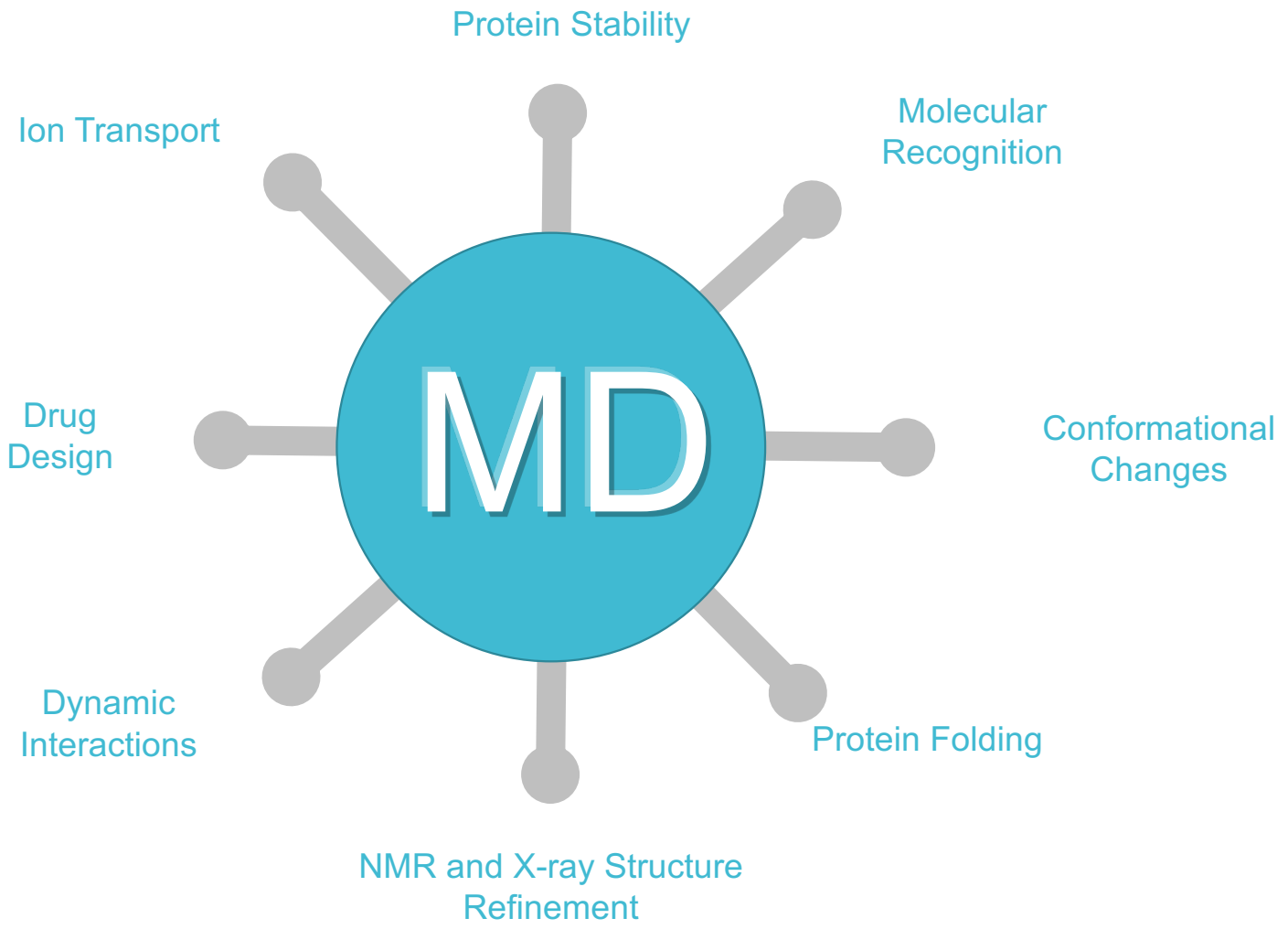
*The group is dedicated to **MINIMAL BIOMIMETIC SYSTEMS**, combining designed molecular recognition agents with functional materials, for **Bioseparation, Biocatalysis, Sensing & Diagnostics, and Nanomedicine**.*

*The main modeling tools to acheive these goals are: **Molecular Dynamics, Docking, Virtual Screening, Protein Structure Prediction, Biomaterials Simulations.***

*We have outstanding benefit for the **interconnected in silico and experimental** in impactful research.*

Website: https://sites.fct.unl.pt/biomolecular_eng/home

Molecular Dynamics (MD)?



Types of MD Simulations

LEVEL USED

QUANTUM MOLECULAR DYNAMICS

CLASSICAL MOLECULAR DYNAMICS

HYBRID MOLECULAR DYNAMICS

Types of MD Simulations

Most common type of Molecular Dynamics simulations presently in use for biomolecules

Based in the use of Molecular Mechanics and in the notion of Force Field

Outline

I. Molecular Mechanics and Force Fields

II. Principles of Molecular Dynamics Simulations

III. Calculation of Useful Properties from MD

1

Molecular Mechanics and Force Fields

Molecular Mechanics (MM)

Based in the Principles of Classical Physics

Instead of trying to solve the electronic Schrödinger equation, MM methods bypass it, writing the energy of the system as a parametric function of the nuclear coordinates, in a formulation that follows the ideas of the Newtonian Mechanics

The Atom is the Smallest Particle Considered

MM methods neglect both electrons and the quantum aspects of the nuclear motion.

A Simplified Scheme of Interactions is Adopted

A “ball and spring” model is normally employed. Atoms are described as charged spheres of different sizes, whereas the bonds are described as springs with a different stiffness.

Molecular Mechanics (MM)

Limitations – Bond Forming and Breaking Events

The neglect of the concept of electron forecloses any direct study of processes involving the formation or breaking of chemical bonds. Chemical Reactions cannot be directly studied by typical MM

The Energy as a Sum of Different Contributions

The energy of the system is split into a sum of contributions from different processes, including the stretching of bonds, the opening and closing of angles, rotations around simple bonds, etc - an obvious approximation

$$E_{MM} = E_{Stretching} + E_{Bending} + E_{Torsional} + E_{Electrostatic} + E_{VDW}$$

Molecular Mechanics (MM)

The Terms in the Energy Expression

Each energy term represents a physically different contribution. Different mathematical formulation can exist for each term, differing in terms of accuracy and time required for calculation.

$$E_{MM} = E_{Stretching} + E_{Bending} + E_{Torsional} + E_{Electrostatic} + E_{VDW}$$

Other terms can also be found, particularly in more sophisticated force fields. Popular examples include cross terms, which reflect the coupling between the different fundamental terms, and improper torsions and out-of-plane bending terms.

The Force Field

A Match made in Heaven – The Force Field

A MM method is characterized not only by its energy expression, but also by the corresponding parameters, the two of which form a single entity termed force field.

The Parameters are typically derived from experimental data or from calculations with higher level methods (e.g. HF, DFT) for small molecules.

The accuracy of the parameterization protocol is of paramount importance to the reliability of the force field.

The Bond Stretching Term

Describing Chemical Bonds

The bond stretching term accounts for the energy change arising from the increase and shortening of bonds within molecules. This change can be accurately described by a Morse potential, with the form:

$$V_l = D_e \left\{1 - \exp\left[- a\left(l - l_0\right)\right]\right\}^2$$

In this equation, D_e corresponds to the depth of the potential energy minimum, l is the distance between the two atoms, with the subscript zero indicating the reference value of the bond.

The Bond Stretching Term

Describing Chemical Bonds: The Morse Potential

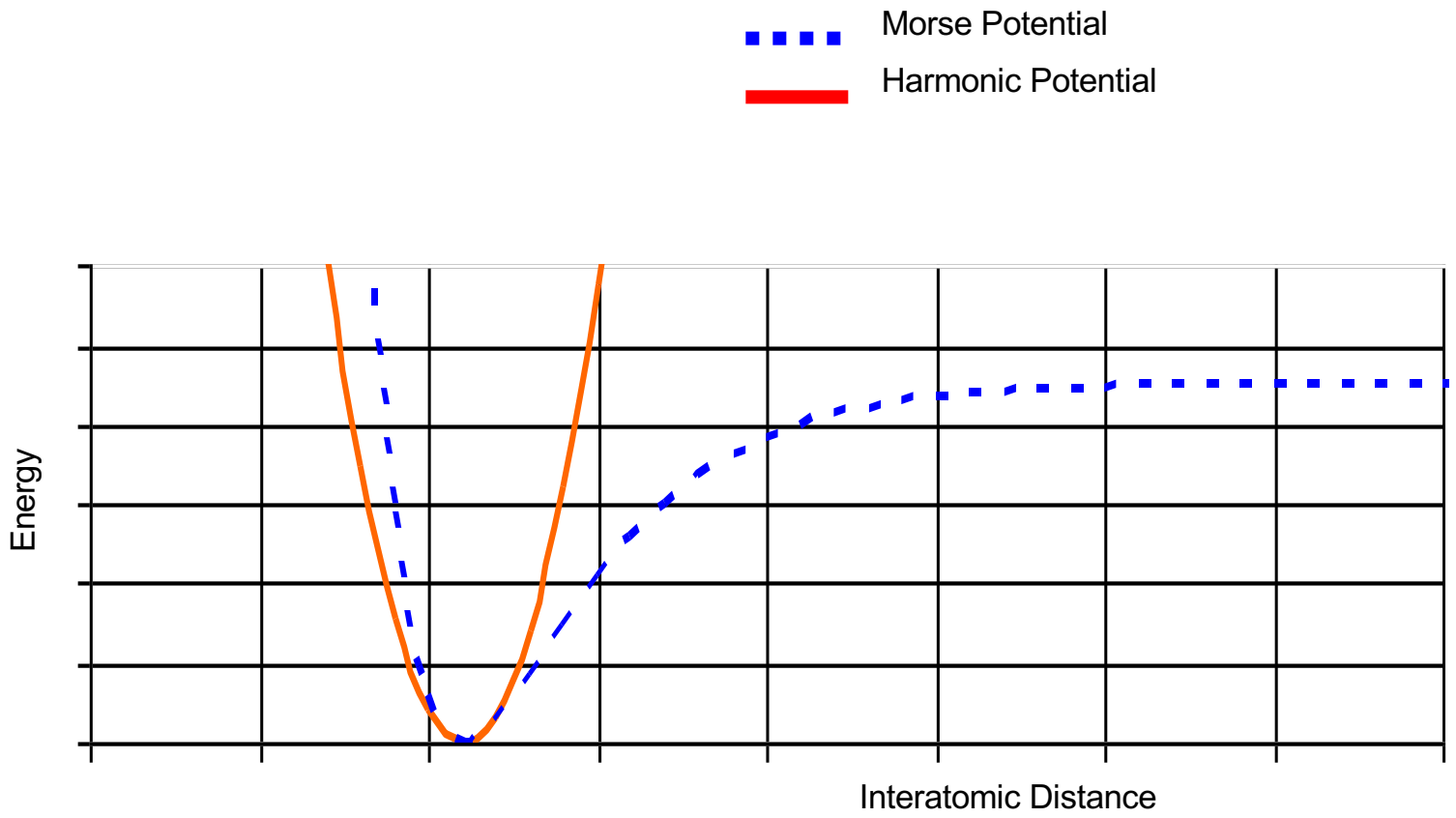
In this equation *a* is a quantity given by:

$$a = \omega \sqrt{\frac{\mu}{2D_e}}$$

while ω is the frequency of the bond vibration, and μ is the reduced mass. The frequency of the bond vibration ω can be further related to *k*, the stretching constant of the bond vibration by:

$$\omega = \sqrt{\frac{k}{\mu}}$$

The Bond Stretching Term



The Bond Stretching Term

Describing Chemical Bonds: The Morse Potential

In spite of the accurate representation allowed by such potential, the Morse potential is not normally included in the standard force-fields.

The corresponding expression is difficult to compute efficiently, and requires three parameters for each individual bond.

$$V_l = D_e\{1-\exp[-a(l-l_0)]\}^2$$

How to make things easier?

Use a simpler potential – The Harmonic Potential

The Bond Stretching Term

Describing Chemical Bonds: The Harmonic Potential

The vast majority of the available force fields consider simply a harmonic description of the bond, by applying the Hooke's law form.

$$V_l = \frac{1}{2} k_l (l - l_0)^2$$

In this equation k_l is the force constant of the correspondent vibration, and l_0 the equilibrium value.

The Bond Stretching Term

Describing Chemical Bonds: The Harmonic Potential

Despite the implicit differences in both expressions, both formulations allow a very similar description around the equilibrium bond length l_0, i.e. at the bottom of the potential well. However, they significantly differ in regions away from this value.

$$V_l = \frac{1}{2} k_l (l - l_0)^2$$

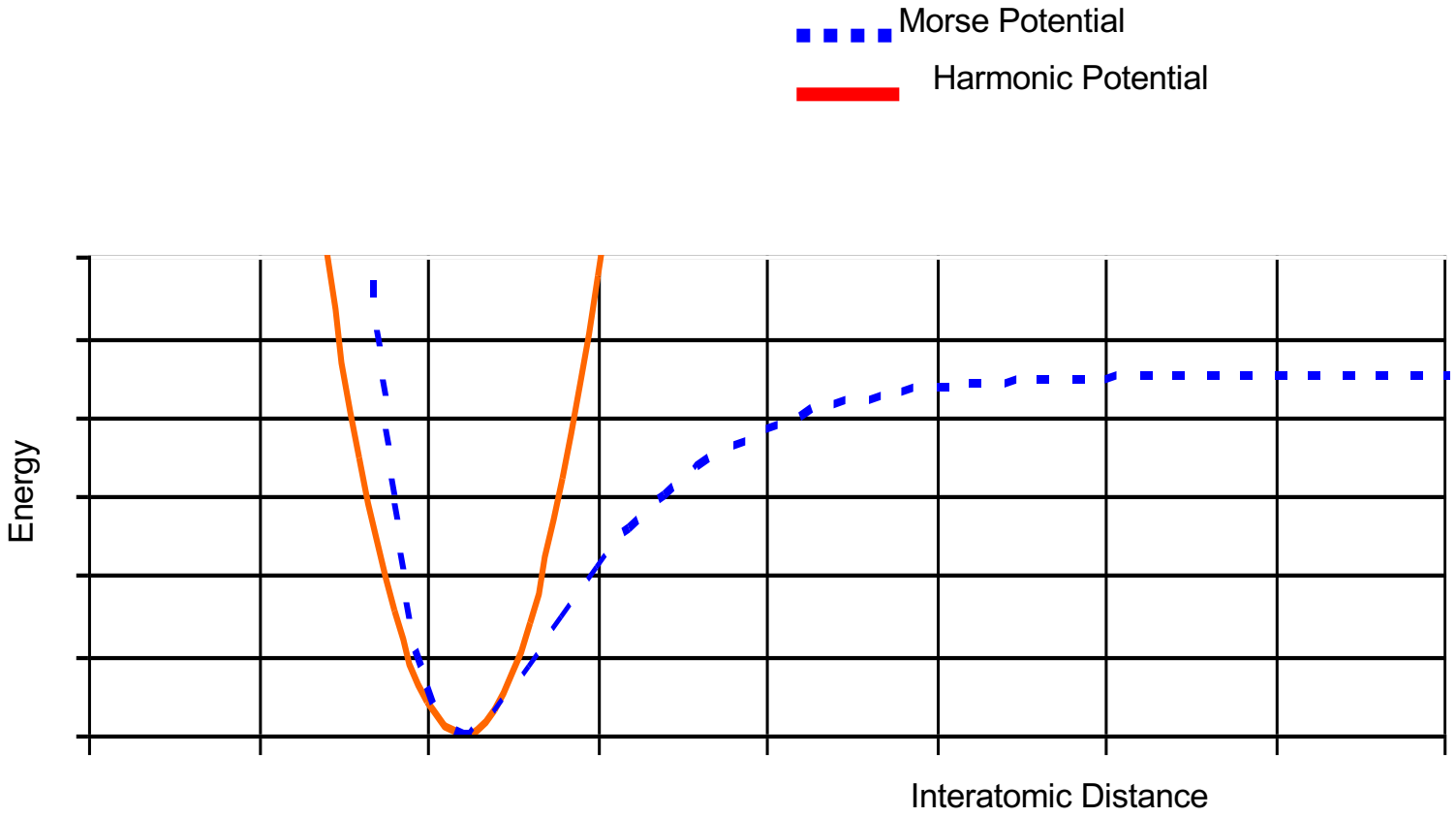
Advantages:

Easier to Compute; Only 2 parameters for bond; Not bad near the minimum;

Disadvantages:

Less Accurate; The energy increases always away from the minimum; No bond dissociation;

The Bond Stretching Term



The Bond Stretching Term

Describing Chemical Bonds: The Harmonic Potential

$$V_l = \frac{1}{2} k_l (l - l_0)^2$$

Can be Improved:

Introducing higher order terms (particularly cubic and quadric) into the harmonic expression.

Normally, however:

For most cases the use of squared terms is enough. All commonly used biomolecular force fields, rely on this basic approximation.

The Angle Bending Term

Describing Angles

This term accounts for the energy variation associated to the increase and decrease of angles from their reference values.

Like in the Bond Stretching term, the harmonic potential is normally employed.

$$V_\theta = \frac{1}{2} k_\theta (\theta - \theta_0)^2$$

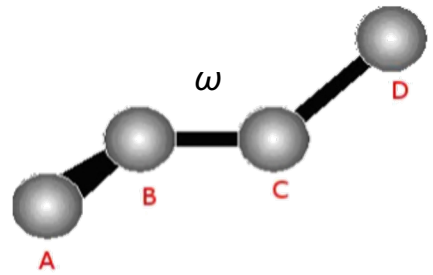
k_θ is the force constant associated to the bending mode, whereas θ and θ_0 are respectively, the angle value and the equilibrium angle value. The accuracy of this term can also be improved by the addition of higher order terms.

The Torsional Term

Describing Dihedrals

In most molecular mechanical force fields, the rotation of molecular fragments that are linked together through covalent bonds is normally described by torsional barriers, computed from a Fourier expansion of a series of cosine functions.

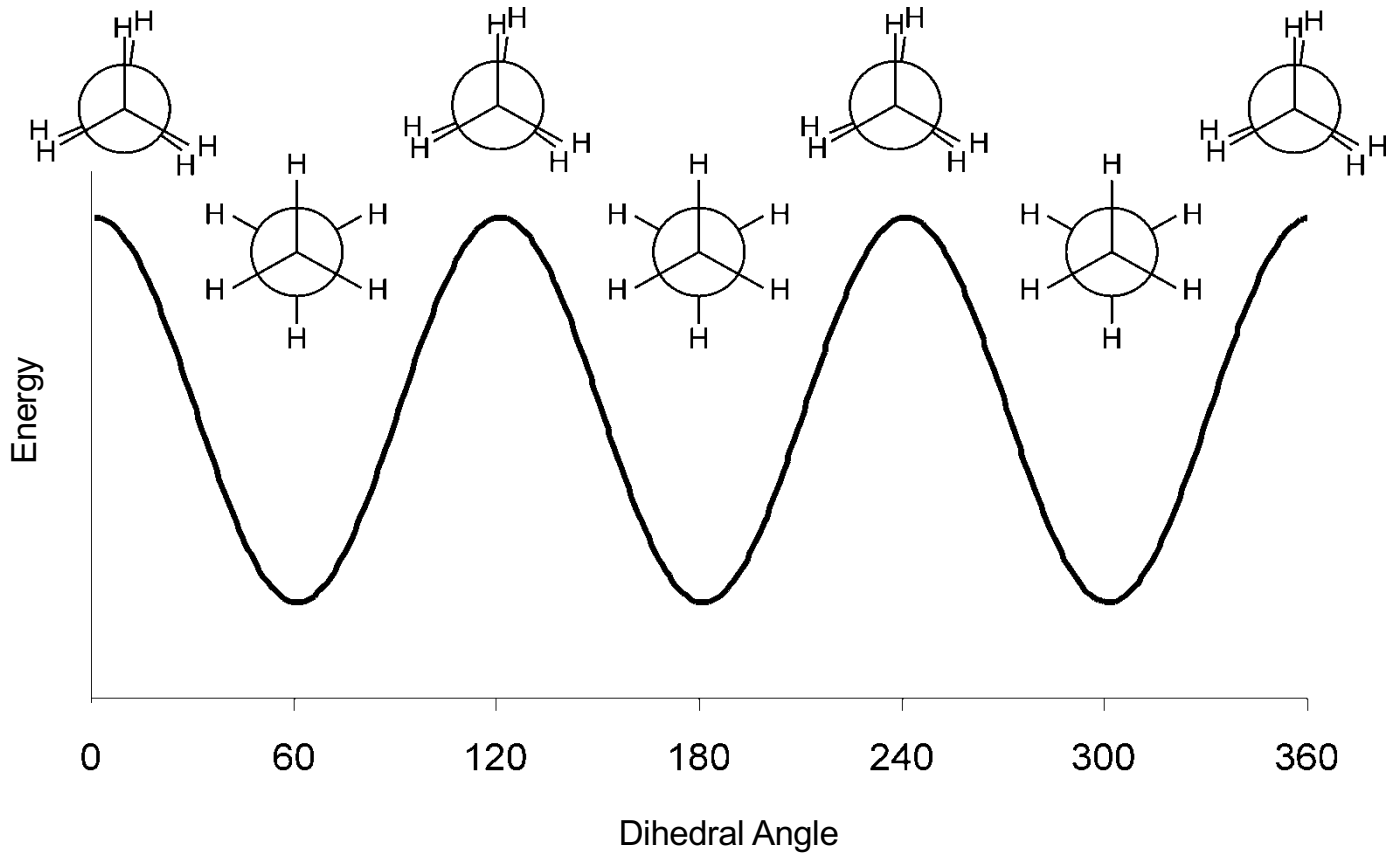
$$V_\omega = \sum_j \frac{1}{2} V_j \left(1 - \cos\left(j\omega\right)\right)$$



In this equation ω is the dihedral angle, i.e. the angle defined between the AB and CD bonds, whereas j is the correspondent multiplicity, i.e. a quantity that gives the number of minimum points in the function as the bond is rotated by 360º. Vj is the correspondent torsional force constant.

The Torsional Term

Describing Dihedrals



The Torsional Term

Describing Dihedrals

When expanded into a Fourier series up to the third term, this equation can take the form:

$$V_\omega = \frac{V_1}{2}(1-\cos\omega)+\frac{V_2}{2}(1-\cos 2\omega)+\frac{V_3}{2}(1-\cos 3\omega)$$

Parameterization

The parameters *V1*, *V2* and *V3* are normally determined by fitting to experimentally determined torsional profiles, or against data determined by quantum mechanical methods.

Difficulties

The parameterization of the torsional terms requires a large number of parameters, even for a modest set of molecules. Typically, parameters for each atomic quartet, according to their atom types, are required.

Molecular Mechanics (MM)

The Terms in the Energy Expression

$$E_{MM} = \underbrace{E_{Stretching} + E_{Bending} + E_{Torsional}}_{\text{Bonded Terms}} + \underbrace{E_{Electrostatic} + E_{VDW}}_{\text{Non-Bonded Terms}}$$

The Electrostatic Term

Describing Charges

The electrostatic energy term describes the non-bonding interactions arising from the presence of atomic charges, i.e. from the internal distribution of electrons.

From this, positively and negatively charged regions in the molecules result.

Most Common: Coulomb Expression

The Electrostatic Term

Describing Charges: The Coulomb Expression

Partial Atomic charges are assigned to each individual atom, with the electrostatic interactions between different molecules or different parts of a same molecule being calculated from the sum of the interactions between pairs of atoms, computed from the Coulomb equation, as indicated below:

$$V_{ele} = \sum_{ij} \frac{q_i q_j}{\varepsilon r_{ij}}$$

The electrostatic energy is then a function of the atomic charges (q_i and q_j), the interatomic distance (r_{ij}), and of the dielectric constant (ε) which accounts for the effect of the surrounding environment.

The Electrostatic Term

Describing Charges: The Coulomb Expression

$$V_{ele} = \sum_{ij} \frac{q_i q_j}{\varepsilon r_{ij}}$$

Parameterization

The main difference in the calculation of the electrostatic energy term between the several available force fields lies in the way the atomic charges are calculated.

Mulliken charges - Natural Bond Orbital (NBO) charges - Merz-Kollman charges - Restrained ElectroStatic Potential (RESP) charges - DelRe charges - Gasteiger charges - Pullman charges

The van der Waals Term

Describing the Non-Electrostatic Interactions

The van der Waals energy term describes the non-electrostatic attractions and repulsions between atoms that are not directly coordinated.

These interactions are attractive at small distances, but fall rapidly to zero when the interacting atoms are separated by more than a few atoms.

For very short distances, i.e. shorter than the sum of the van der Waals radii of two interacting atoms, this interaction is highly repulsive due to the overlap of the electronic density.

The Van der Waals Term

Describing the Non-Electrostatic Interactions

The van der Waals energy term is normally approximated by a Lennard-Jones 12-6 potential.

$$V_{vdw} = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

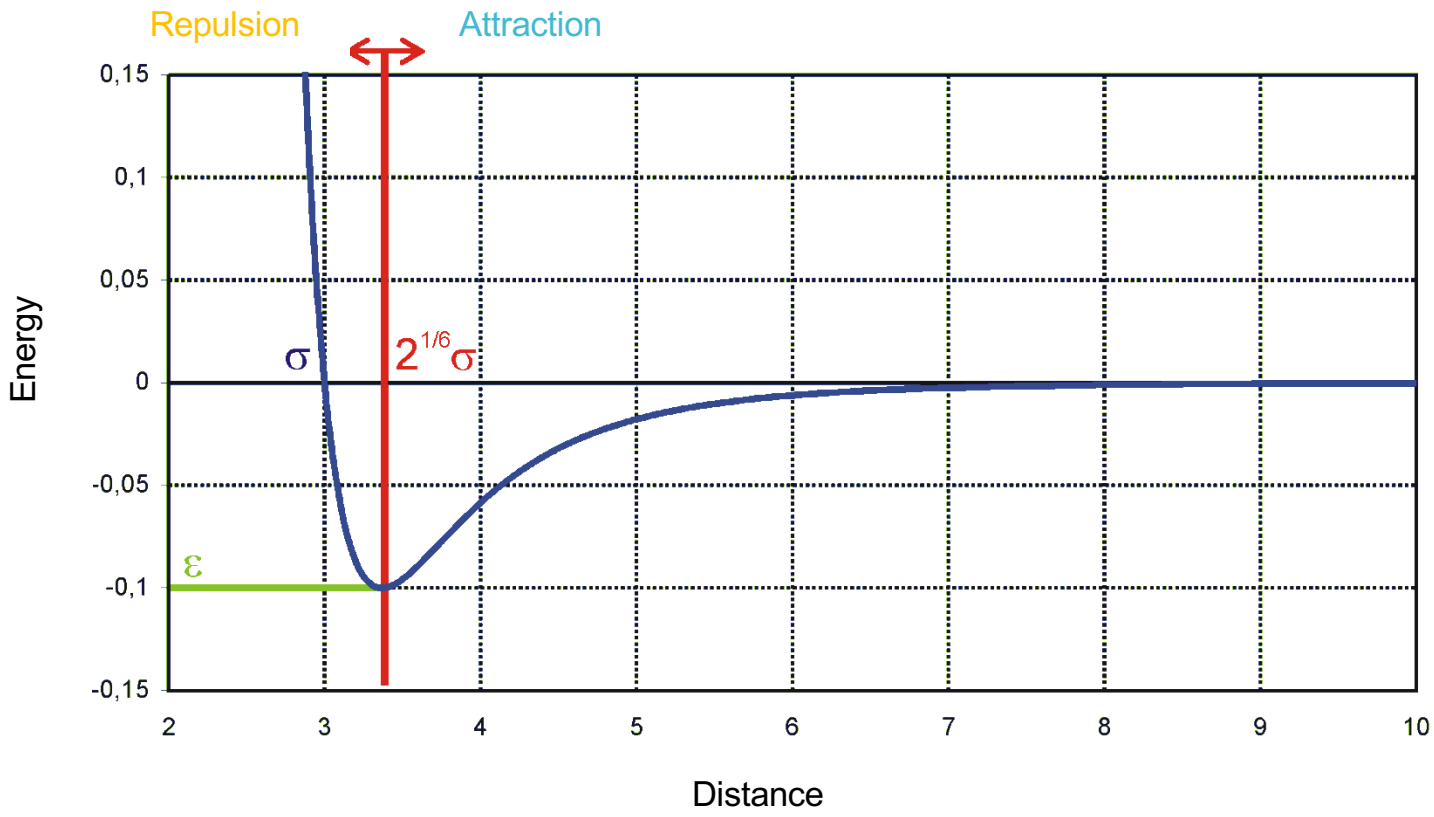
ε is the depth of the potential well

σ is the (finite) distance at which the inter-particle potential is zero

r is the distance between the particles.

The van der Waals Term

The Lennard Jones 12-6 Potential



The Force Field

A Match made in Heaven – The Force Field

A MM method is characterized not only by its energy expression, but also by the corresponding parameters, the two of which form a single entity termed force field.

Biomolecular Force Fields

Most Common Biomolecular Force Fields

Different classes of molecules require different energy expressions for optimal description, and of course different MM parameters

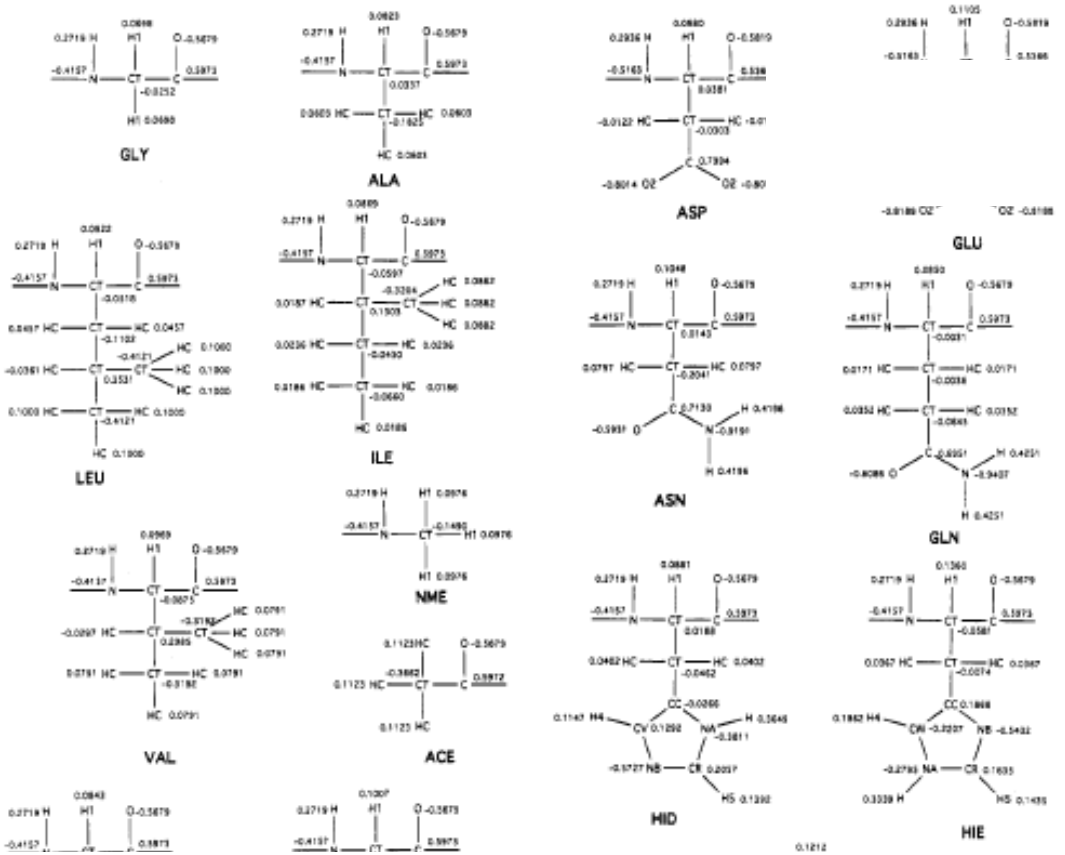
To obtain high accuracy calculations a careful parameterization of an extremely diverse and complete set of reference molecules is required.

This is, in practice, an impossible mission. Currently available general force fields had to sacrifice accuracy for a wider applicability.

Improved quality is normally achieved by developing specialized force fields, ensuring accurate calculations to be performed, albeit in a much more limited class of compounds.

Biomolecular Force Fields

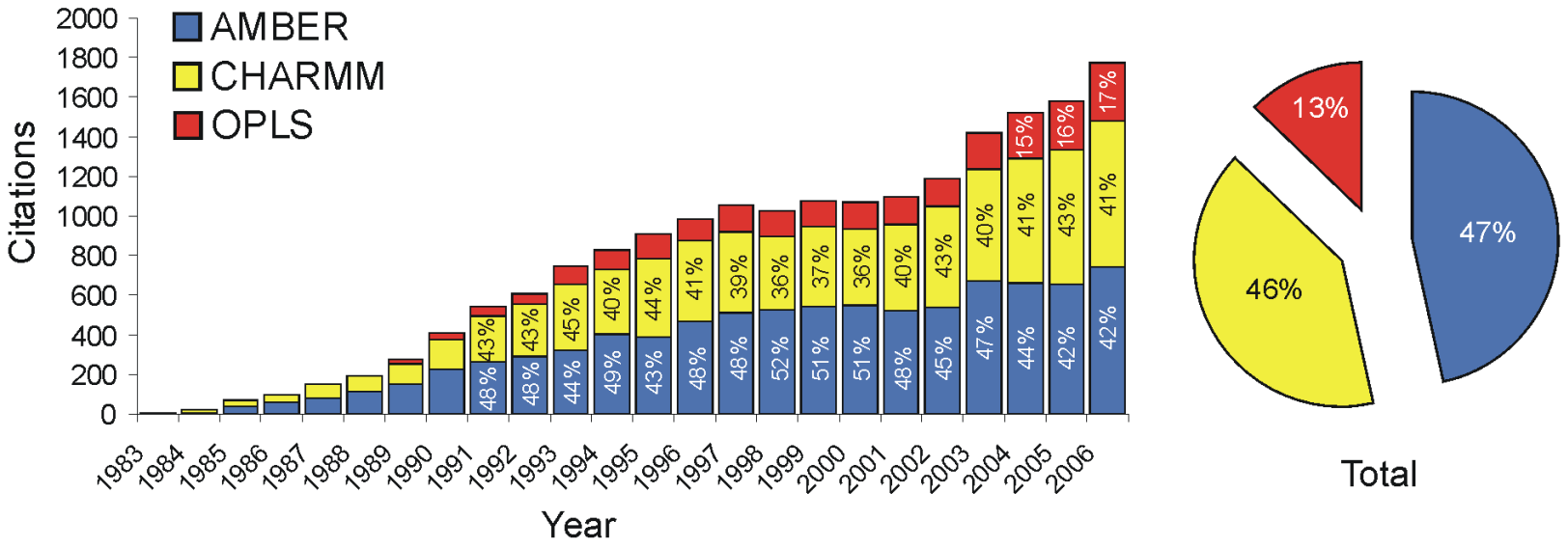
Biomolecular Force Fields for Proteins



Biomolecular Force Fields

Most Common Biomolecular Force Fields

AMBER, CHARMM, GROMOS and OPLS are the most popular molecular force fields devised to describe proteins.



Biomolecular Force Fields

Most Common Biomolecular Force Fields

AMBER, CHARMM, GROMOS and OPLS are the most popular molecular force fields devised to describe proteins.

Common Characteristics:

All four use energy expressions with harmonic terms for bonds and angles, Fourier series for torsions, and pairwise van der Waals and Coulombic interactions between atoms that are separated by three or more bonds.

Differences:

They are parameterized in conceptually different ways.

Note:

Individual parameters from different force fields should not be compared, as the parameterization scheme varies from force-field to force-field

AMBER

Assisted Model Building and Energy Refinement

$$U(\vec{R}) = \sum_{bonds} K_b (b - b_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 +$$

$$\sum_{dihedrals} \frac{V_n}{2}(1 + \cos(n\phi - \delta)) + \sum_{nonbond} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon_1 R_{ij}} \right]$$

where K_b, K_θ, and V_n are the bond, angle, and dihedral angle force constants

b, θ, and ϕ are the bond length, bond angle, and dihedral angle, with the subscript zero representing the equilibrium values for the individual terms

γ is the phase angle and takes values of either 0º or 180º

A_{ij} and B_{ij} represent respectively the van der Waals and London dispersion terms,

q_i and q_j represent the partial atomic charges

ε is the dielectric constant that takes into account the effect of the medium

CHARMM

The Chemistry at HARvard Macromolecular Mechanics

$$U(\vec{R}) = \sum_{bonds} K_b (b - b_0)^2 + \sum_{UB} K_{UB} (S - S_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 +$$
$$\sum_{dihedrals} K_\chi (1 + \cos(n\chi - \delta)) + \sum_{impropers} K_{imp} (\varphi - \varphi_0)^2 +$$
$$\sum_{nonbond} \varepsilon \left[\left(\frac{R_{\min_{ij}}}{r_{ij}} \right)^{12} - \left(\frac{R_{\min_{ij}}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\varepsilon_1 r_{ij}}$$

where K_b, K_{UB}, K_θ, K_χ, and K_{imp} are the bond, Urey-Bradley, angle, dihedral angle, and improper dihedral angle force constants,

b, S, θ, χ and φ are the bond length, Urey-Bradley 1,3-distance, bond angle, dihedral angle, and improper torsion angle, respectively, with the subscript zero representing the equilibrium values for the individual terms.

ε is the Lennard-Jones well depth and R_{min} is the distance at the Lennard-Jones minimum

q_i is the partial atomic charge, ε_1 is the effective dielectric constant, and r_{ij} is the distance between atoms i and j.

OPLS-AA

Optimized Potential for Liquid Simulations

$$U(\vec{R}) = \sum_{bonds} K_r (r - r_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 +$$

$$\sum_{dihedrals} \left(\frac{V_1^i}{2}(1 + \cos(\phi_i - f_i 1)) + \frac{V_2^i}{2}(1 - \cos(2\phi_i + f_i 2)) + \frac{V_3^i}{2}(1 + \cos(3\phi_i - f_i 3)) \right) +$$

$$\sum_{nonbond} \left[4\varepsilon_{ij} \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j e^2}{r_{ij}}$$

where K_r, and K_θ are the bond, and angle force constants;

r and θ are the bond length and bond angle, with the subscript zero representing the equilibrium values for the individual terms;

$ø$ is the dihedral angle, $V1$, $V2$, and $V3$ are the coefficients in the Fourier series for each torsion, and f_i1, f_i2, and f_i3 are the correspondent phase angles;

ε is the Lennard-Jones well depth, σ is the finite separation at which the Lennard-Jones potential is zero, qi is the partial atomic charge, and r_{ij} is the distance between atoms i and j.

Biomolecular Force Fields

Which one is better?

2

Molecular Dynamics Principles

General Principles

From Molecular Mechanics to Molecular Dynamics

The MM Energy of the System can be given by:

$$U(\vec{R}) = \sum_{bonds} K_b (b - b_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 +$$

$$\sum_{dihedrals} \frac{V_n}{2}(1 + \cos(n\phi - \delta)) + \sum_{nonbond} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon_1 R_{ij}} \right]$$

The Force can be easily obtained from an MM Energy:

$$-\frac{dU}{dr} = F(r)$$

General Principles

From Molecular Mechanics to Molecular Dynamics

Now consider Newton's Second Law:

$$F = ma$$

For particle i with position x_i the positions and velocities (i.e. the trajectory) can be obtained

$$\frac{d^2 x_i}{dt^2} = \frac{F_{x_i}}{m_i}$$

By integrating Newton's equations of motion, one gets a trajectory describing the positions, velocities, and accelerations of the several particles that constitute the system, as they vary with time.

General Principles

A Deterministic Method

Given an initial set of positions and velocities, the subsequent time evolution is *in principle* completely determined.

Once the positions and velocities of each particle are known, the state of the system can be predicted at any time, both in the future and in the past.

Requires a Numeric Integration

The potential energy of the system is a function of the atomic position (3N) of all the particles within the system.

No analytic solution exists for the Newton's equations of motion of such a system, which must therefore be solved numerically

A finite time step (δt) must be used.

General Principles

Integration Algorithms Used

Several different integration algorithms available

- Verlet algorithm
- Leap-frog algorithm
- Velocity Verlet algorithm
- Beeman's algorithm

Steps in System Preparation for MD

For a typical Biomolecular System (e.g. Protein)

Example illustrating only some of the most common choices and conditions

MODEL PREPARATION

Structure obtained from the Protein Data Bank

Add Hydrogens

Check existence of MM Parameters

ADDITION OF COUNTER-IONS

To neutralize the excess of positive or negative charges in the system

Cl^- or Na^+ are added

SOLVATION OF THE SYSTEM

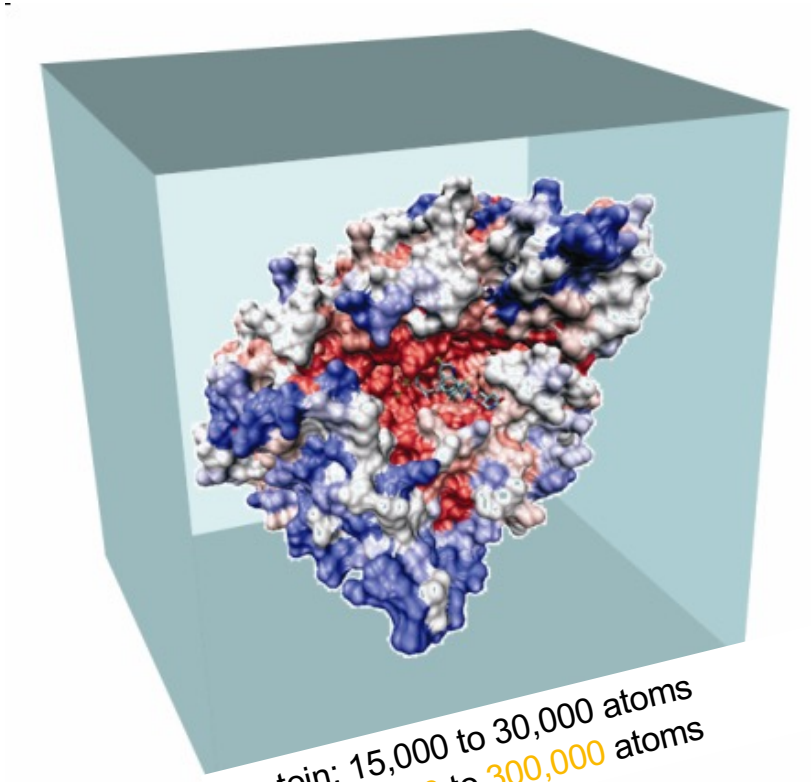
Addition of a Box of Explicit Waters involving all the system

TIP3P Type more common

Steps in System Preparation for MD

For a typical Biomolecular System (e.g. Protein)

Example illustrating only some of the most common choices and conditions



Protein: 15,000 to 30,000 atoms
Total: 100,000 to 300,000 atoms

SOLVATION OF THE SYSTEM

Addition of a Box of Explicit Waters involving all the system

TIP3P Type more common

Steps in a MD Simulation

For a typical Biomolecular System (e.g. Protein)

Example illustrating only some of the most common choices and conditions

MM MINIMIZATION

Eliminates bad contacts

Optimize the position of the hydrogen atoms added

MD EQUILIBRATION

NVT Ensemble

Heats up the system (from 0 to 310.15 K)

Density of the System is equilibrated

100 – 200 ps

MD PRODUCTION

NPT Ensemble

P = 1 atm
T = 310.15 K

Part of the Simulation Used for Evaluation

50-500 ns

Times Scales

The Simulation Time vs the Different Motion Time Scales

Different chemical phenomena involve different time scales. Even when considering only proteins it is important to keep in mind that their various characteristic types of motion have some very different time scales

Folding	Vibration of Bonds	Rotation of Side Chains
Subunit Rearrangement	Diffusion of Molecules	Subunit Association or Dissociation

The Length of the Simulation should therefore be adequate to the type of motion under study

Times Scales

The Simulation Time vs the Different Motion Time Scales

Type of Motion	Functionality Examples	Time and Amplitude Scales
Local Motions Atomic Fluctuation Side Chain Motion	Ligand Docking Flexibility Temporal Diffusion Pathways	**Femtoseconds (fs) to picoseconds (ps)** (10^{-15}-10^{-12} s) Less than 1 Å
Medium-Scale Motions Loop Motion Terminal-Arm Motion Rigid-Body Motion	Active site Conformation Adaptation Binding Specificity	**Nanoseconds (ns) to microseconds (μs)** (10^{-9}-10^{-6} s) 1-5 Å
Large-Scale Motions Domain Motion Subunit Motion	Hinge-bending Motion Allosteric Transitions	**Microseconds (μs) to milliseconds (ms)** (10^{-6}-10^{-3} s) 5-10 Å
Global Motions Helix-Coil Transition Folding/Unfolding Subunit Association/Dissociation	Hormone Activation Protein Functionality	**Milliseconds (ms) to hours (ms)** (10^{-3}-10^{4} s) More than 10 Å

Times Scales

The Simulation Time vs the Different Motion Time Scales

Different chemical phenomena involve different time scales. Even when considering only proteins it is important to keep in mind that their various characteristic types of motion have some very different time scales

Folding	Vibration of Bonds	Rotation of Side Chains
Subunit Rearrangement	Diffusion of Molecules	Subunit Association or Dissociation

It is important to retain that the different types of motion are interdependent and coupled to one another, although for some practical applications some types may be regarded as independent.

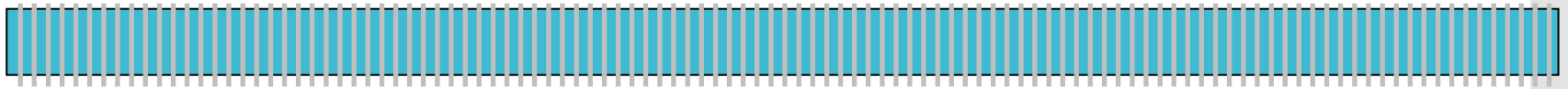
Technical Aspects in MD Simulations

I. Selecting the Time Step

How often are the equations of motion integrated?
Which Time Step (δt) should be used?

Too Small Time Steps

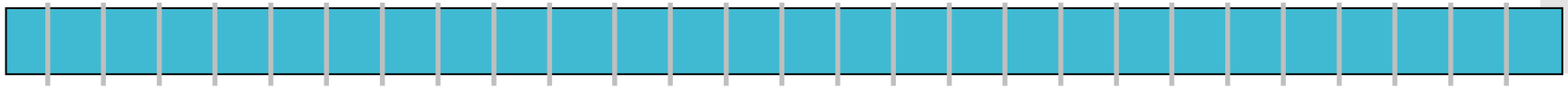
A small time step implies that much more computational time will be required to simulate a given MD run, limiting in practice the proportion of the phase space that can be covered in a given CPU time interval.



Simulation Time

Too Large Time Steps

Too large time steps, tend to result in instabilities in the integration algorithm (due to brisk changes in the forces), which can ultimately result in program failure due to numerical overflow.



Simulation Time

Technical Aspects in MD Simulations

Choice of the time step to use in an MD simulation will result from a balance between economy and accuracy.

General Rule:

A common guideline when simulating flexible molecules is to choose a time step at least one order of magnitude smaller than the time of the shortest period of motion.

Fastest Motion?

Typical Biomolecular Systems (Proteins)

Bond Stretching of Bonds Involving H atoms

Period: 10 fs (e.g. for the case of C-H bond)

Recommended Time Step: 1 fs

Technical Aspects in MD Simulations

I. Selecting the Time Step

High-frequency motions such as the C-H vibrations pose a significant restriction on the length of time step used. However, they are seldom of interest and have a marginal effect on the global behaviour of the system.

Freeze the higher-frequency vibrations by constraining the correspondent bonds to their equilibrium values, without affecting the remaining degrees of freedom. e.g. all bonds involving hydrogen atoms can be frozen

Fastest Motion?

Typical Biomolecular Systems (Proteins)

Bond Stretching of Bonds Involving Heavy Atoms

Period: 2-5 times higher

Permited Time Step: 2 fs

Technical Aspects in MD Simulations

I. Selecting the Time Step

In practice, this constrains on the type of motions considered in MD simulations can be done using specific algorithms

e.g: SHAKE, LINCS, SETTLE

The property under evaluation must be independent of the degrees of freedom associated to these higher-frequency vibrations.

Technical Aspects in MD Simulations

II. Cut-off for Non-bonded Interactions

In a typical MD simulation most of the time is spend (> 90%) calculating the non-bonded interactions terms: the electrostatic and van der Waals energy terms.

In principle, such interactions would have to be calculated for each given atom in relation to every other atom in the system, i.e. for all atom pairs (except those that are covalently bonded or separated by less than three bonds).

Many Pairs of Atoms

COMPUTATIONALLY EXPENSIVE !!!

Calculate only the Relevant Ones!!!

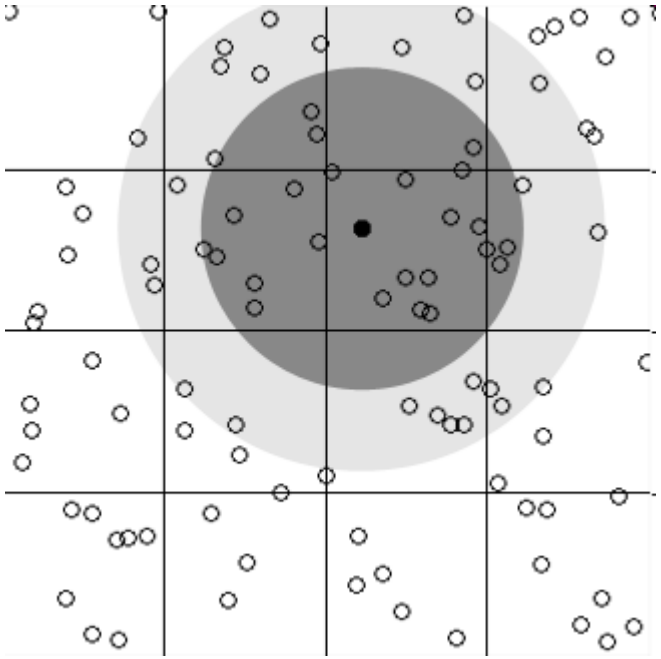
How to Choose?

Technical Aspects in MD Simulations

II. Cut-off for Non-bonded Interactions

Simplest Approach: Use of Cut-Offs

Only the interactions of a given atom with the atoms within a given radius (the cut-off radius) are considered.



Technical Aspects in MD Simulations

II. Cut-off for Non-bonded Interactions

Simplest Approach: Use of Cut-Offs

Only the interactions of a given atom with the atoms within a given radius (the cut-off radius) are considered.

ELECTROSTATIC
Decays slowly at long distances
(with r^{-1})

$$V_{ele} = \sum_{ij} \frac{q_i q_j}{\varepsilon r_{ij}}$$

VAN DER WAALS
Decays very strongly with distance (with r^{-6})

$$V_{vdw} = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

Technical Aspects in MD Simulations

II. Cut-off for Non-bonded Interactions

Simplest Approach: Use of Cut-Offs

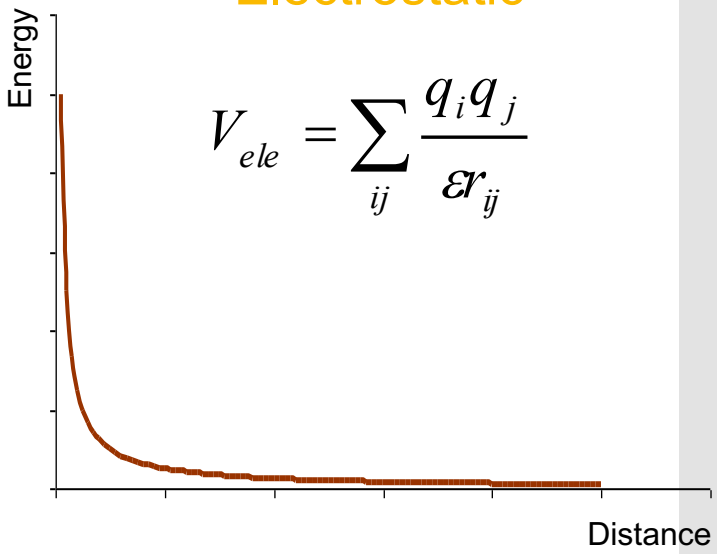
Only the interactions of a given atom with the atoms within a given radius (the cut-off radius) are considered.

ELECTROSTATIC
Decays slowly at long distances
(with r^{-1})

VAN DER WAALS
Decays very strongly with distance (with r^{-6})

Electrostatic

$$V_{ele} = \sum_{ij} \frac{q_i q_j}{\varepsilon r_{ij}}$$



Technical Aspects in MD Simulations

II. Cut-off for Non-bonded Interactions

Simplest Approach: Use of Cut-Offs

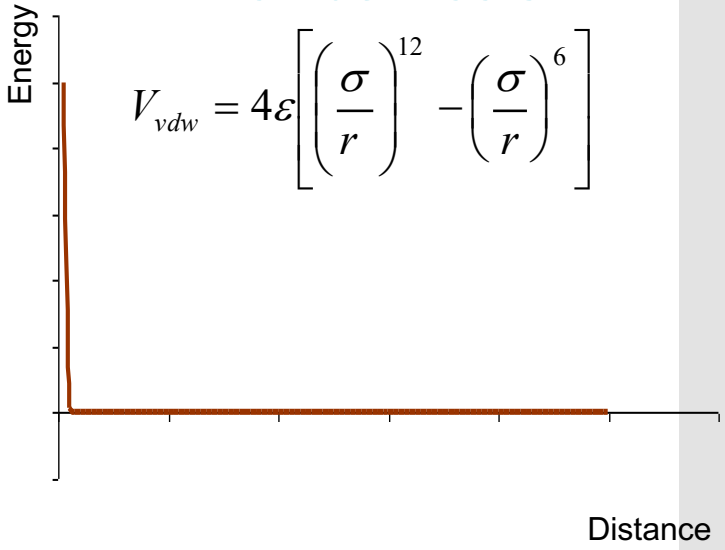
Only the interactions of a given atom with the atoms within a given radius (the cut-off radius) are considered.

ELECTROSTATIC
Decays slowly at long distances
(with r^{-1})

VAN DER WAALS
Decays very strongly with distance (with r^{-6})

Van de Waals

$$V_{vdw} = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$



Technical Aspects in MD Simulations

II. Cut-off for Non-bonded Interactions

Simplest Approach: Use of Cut-Offs

Only the interactions of a given atom with the atoms within a given radius (the cut-off radius) are considered.

ELECTROSTATIC
Decays slowly at long distances
(with r^{-1})

Can be Dangerous!!!

VAN DER WAALS
Decays very strongly with distance (with r^{-6})

Safe Approximation!

Technical Aspects in MD Simulations

II. Cut-off for Non-bonded Interactions

Simplest Approach: Use of Cut-Offs

Typical Cut-offs
10 or 12 Å

ELECTROSTATIC
Decays slowly at long distances
(with r^{-1})

Other more advanced methods have to be included

VAN DER WAALS
Decays very strongly with distance (with r^{-6})

Sometimes with a small correction term

Technical Aspects in MD Simulations

II. Cut-off for Non-bonded Interactions

Methods to handle the Long-Range Electrostatic Interactions

- Ewald summation method
- Reaction field method
- Cell multipole method
- Particle Mesh Ewald Summation method (PME)

Technical Aspects in MD Simulations

As the size of a given spherical system increases, its volume grows as the cube of the radius, whereas its surface grows as the square.

Atoms near the surface (i.e. at the boundary) have fewer neighbours than atoms located inside.

Handling Surface Effects:

In a real macroscopic system the importance of surface effects is typically small in comparison with the chemical phenomena under study (with natural exceptions)

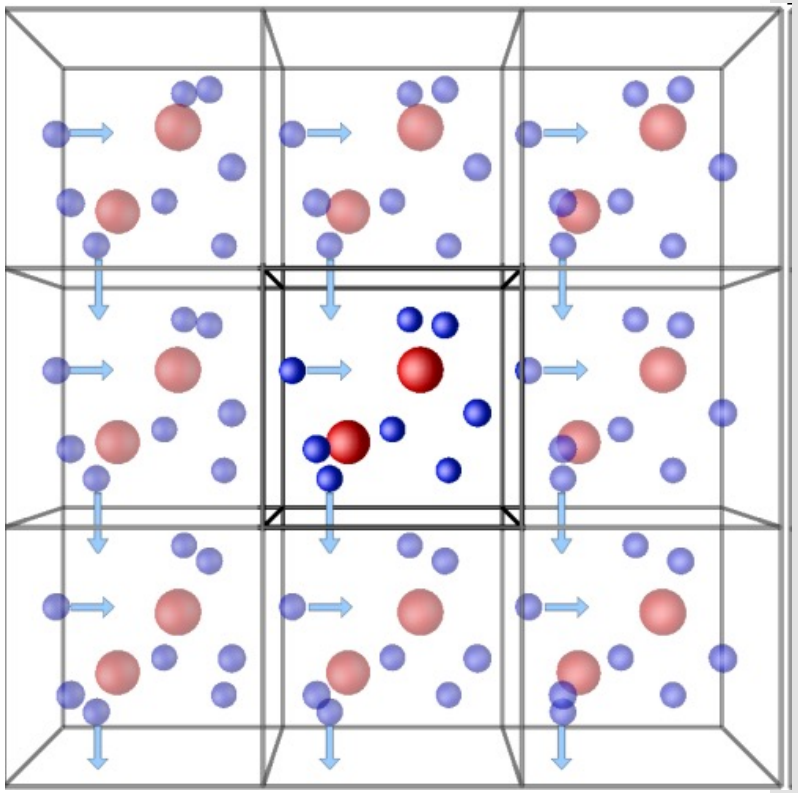
For the average chemical problem considered in MD simulations, the dimension of the model system is inevitability so small that these surface effects cannot be neglected and may actually come to dominate over the systems properties.

Technical Aspects in MD Simulations

III. Boundary Conditions

Solution: Use of Periodic Boundary Conditions (PBC)

Consists in assuming the model system to be enclosed within a box (a unit cell) that is considered replicated into infinity in all directions, thereby completely filling the surrounding space, and virtually eliminating all surface effects from the model system.



Technical Aspects in MD Simulations

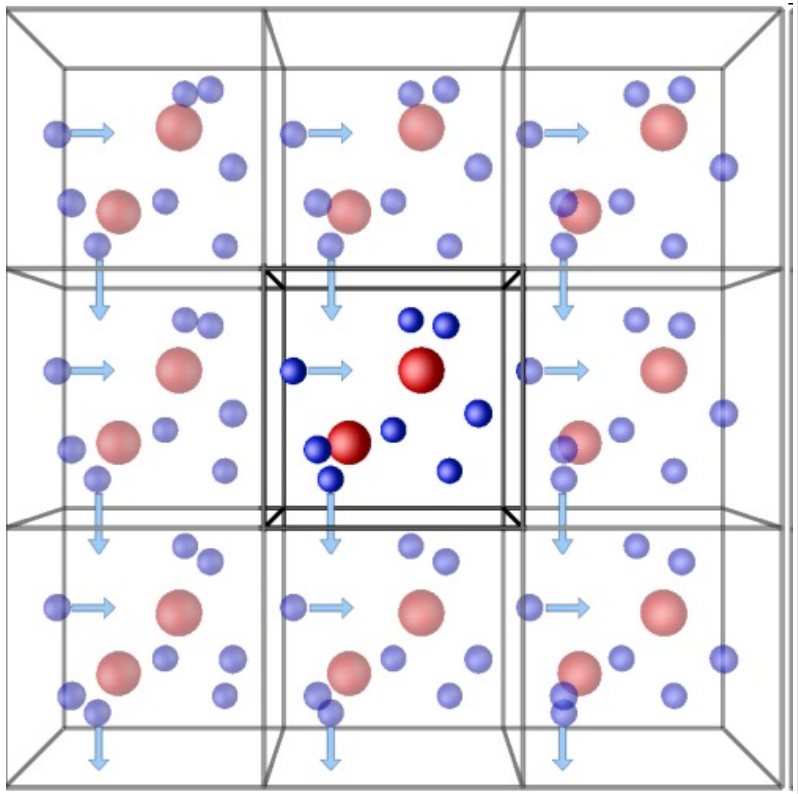
III. Boundary Conditions

Solution: Use of Periodic Boundary Conditions (PBC)

Each particle within an unit cell is considered as interacting not only with all the other particles within that very same unit cell, but also with their images in the nearby unit cells.

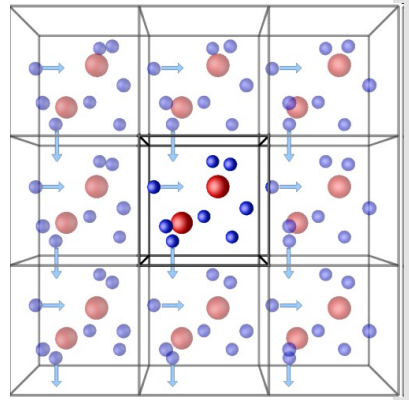
If the trajectory of an individual particle takes it outside the unit cell boundary according to a given direction in space, its image simultaneously enters that same unit cell from an opposite point on the surface of the unit cell.

Both the interactions and the particles can “go through” unit cell boundaries.



Technical Aspects in MD Simulations

III. Boundary Conditions



Solution: Use of Periodic Boundary Conditions (PBC)

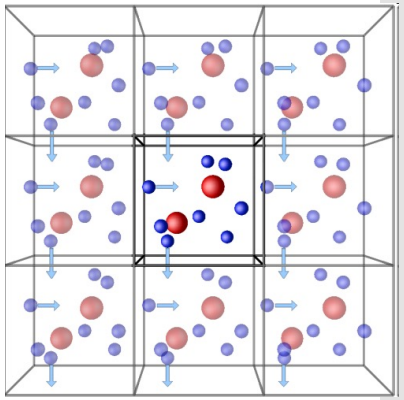
Key Issue: Dimension of the unit cell

Should be at least as large as the largest interaction cut-off length considered in the MD simulation.

The use of smaller unit cells would make some interatomic interactions to be counted twice, i.e. one time within the unit cell and another time with an image outside the cell.

Technical Aspects in MD Simulations

III. Boundary Conditions



Solution: Use of Periodic Boundary Conditions (PBC)

Key Issue: Dimension of the unit cell

When modeling a typical biomolecular system, such as a biomolecule in a solvent, the dimensions of the unit cell should encompass at least the dimension of the biomolecule, plus at least two-times the largest interaction cut-off, as to prevent two solute molecules on different unit cells of interacting between each other.

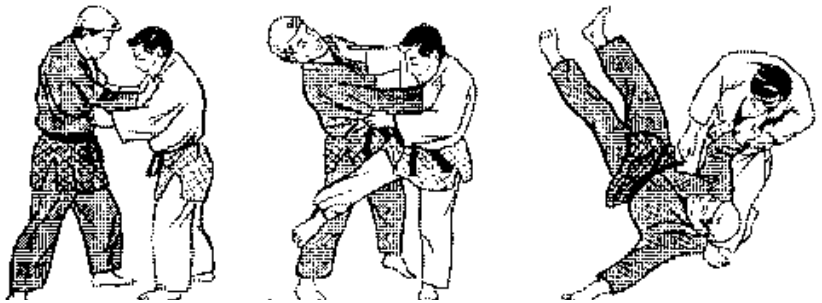
3

Calculation of Properties from MD

Properties from MD Simulations

A Dynamic Portrait at the Atomic Level

Transforms a static X-ray structure from a protein, for example directly taken from the Protein Data Bank, into an ensemble of meaningful structures at a given temperature and pressure (ensemble NPT) that can be illustrative of the plethora of states in equilibrium that characterize a given biomolecular system.



Properties from MD Simulations

A Dynamic Portrait at the Atomic Level

Transforms a static X-ray structure from a protein, for example directly taken from the Protein Data Bank, into an ensemble of meaningful structures at a given temperature and pressure (ensemble NPT) that can be illustrative of the plethora of states in equilibrium that characterize a given biomolecular system.

Properties from MD Simulations

A Dynamic Portrait at the Atomic Level

Transforms a static X-ray structure from a protein, for example directly taken from the Protein Data Bank, into an ensemble of meaningful structures at a given temperature and pressure (ensemble NPT) that can be illustrative of the plethora of states in equilibrium that characterize a given biomolecular system.

Dynamic Events within a Biological System

Transient Phenomena

Solvation and Coordination spheres

Formation and breaking of Hydrogen Bonds

Effect of the Solvent

The Root Mean Square Deviation (RMSd)

The Root Mean Square Deviation (RMSD) is a very useful measure in tridimensional geometry to compare two different conformations of a set of atoms, normally in the form of molecules.

For an identical set of atoms present in two different conformations, an RMSD analysis measures the difference in terms of position of the two sets of atoms. Consider two sets of n points in conformation v and w. The RMSD between the two conformations would be given by:

$$RMSD(v,w) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\left(v_{ix} - w_{ix}\right)^2 + \left(v_{iy} - w_{iy}\right)^2 + \left(v_{iz} - w_{iz}\right)^2\right)}$$

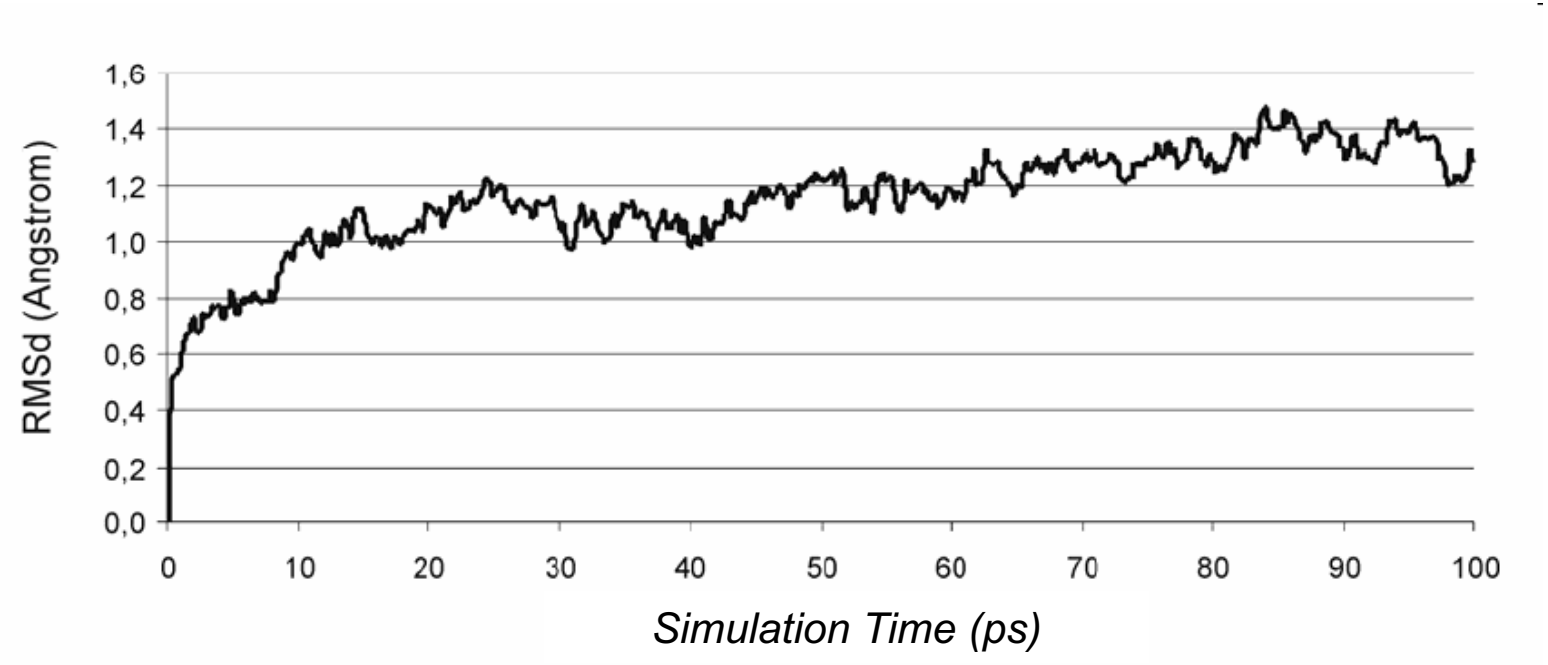
The Root Mean Square Deviation (RMSd)

The Root Mean Square Deviation (RMSD) is a very useful measure in tridimensional geometry to compare two different conformations of a set of atoms, normally in the form of molecules.

For an identical set of atoms present in two different conformations, an RMSD analysis measures the difference in terms of position of the two sets of atoms. Consider two sets of n points in conformation v and w. The RMSD between the two conformations would be given by:

$$RMSD(v,w) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\left(v_{ix} - w_{ix}\right)^2 + \left(v_{iy} - w_{iy}\right)^2 + \left(v_{iz} - w_{iz}\right)^2\right)}$$

The Root Mean Square Deviation (RMSd)



Used to evaluate if the structures generated in the course of a MD simulation are already in equilibrium or not, using typically the initial structure of the system as a reference.

The Root Mean Square Fluctuation (RMSF)

Standard Deviation of the Fluctuation of the Atomic Positions over time

Gives a Measure of the Movement of a Subset of the System with Respect to the Average Structure over the whole simulation.

Measure of the Flexibility of a Given Structural Unit (chain, helix,residue, atom).

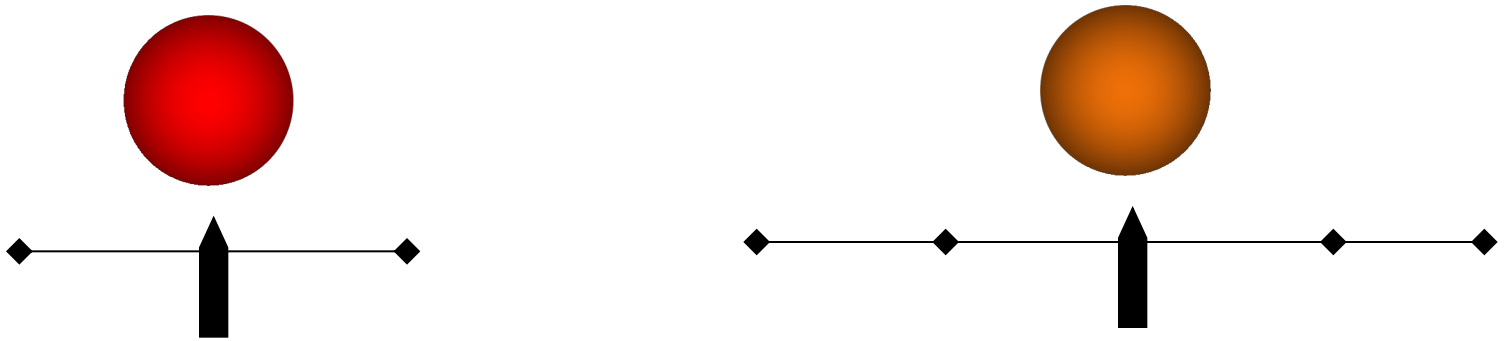
$$RMSF(x) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(x_i - \tilde{x}_i\right)^2}$$

The Root Mean Square Fluctuation (RMSF)

Standard Deviation of the Fluctuation of the Atomic Positions over time

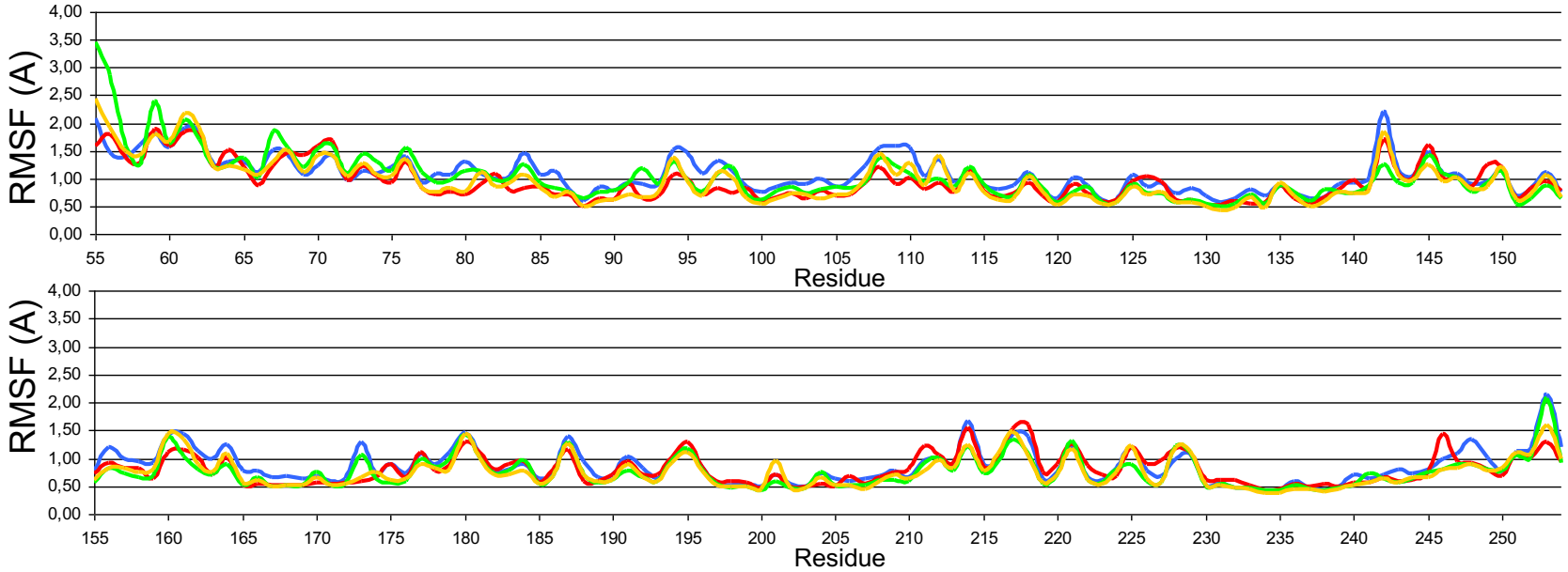
Gives a Measure of the Movement of a Subset of the System with Respect to the Average Structure over the whole simulation.

Measure of the Flexibility of a Given Structural Unit (chain, helix, residue, atom).

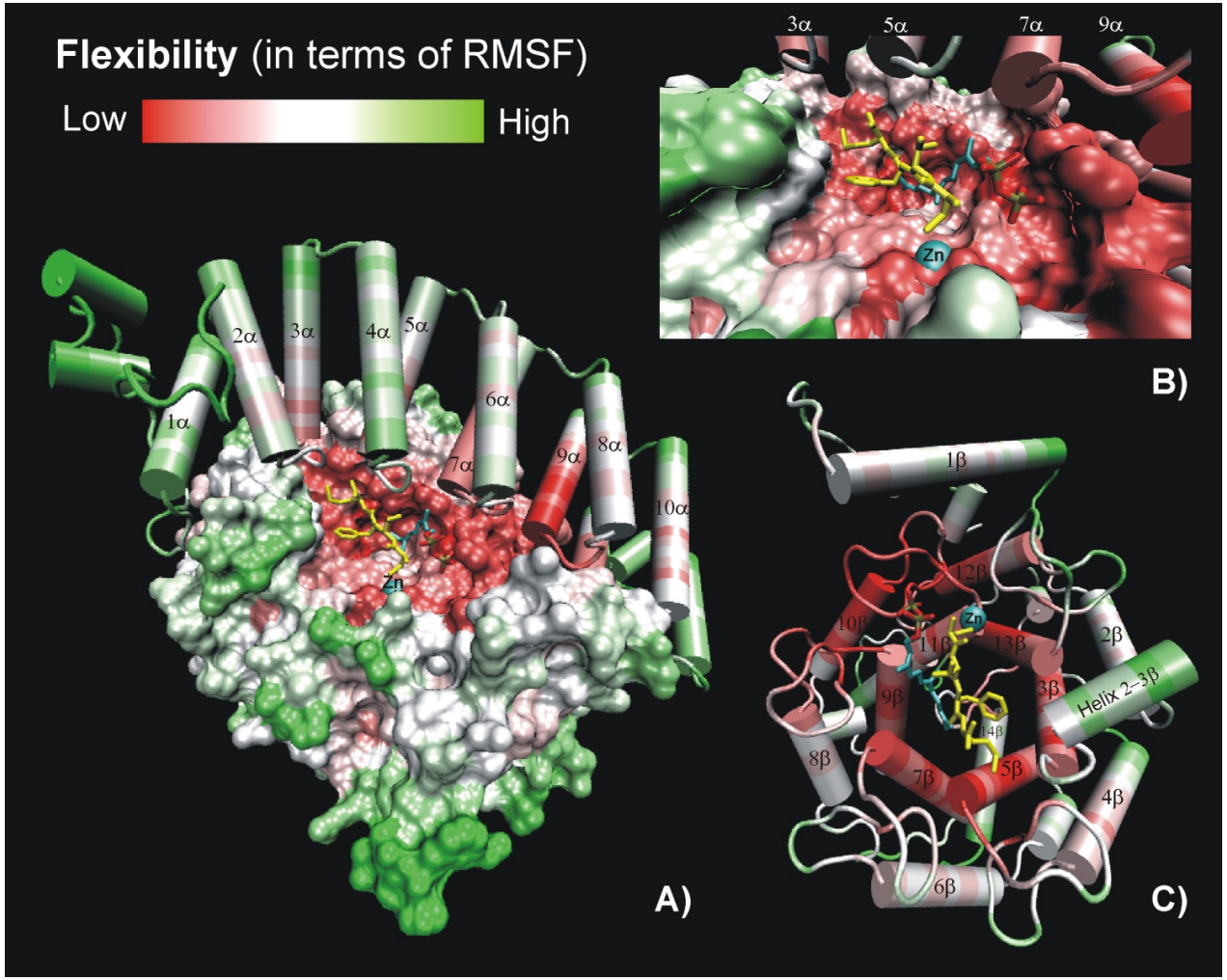


The Root Mean Square Fluctuation (RMSF)

How Does the Flexibility of a Given Amino Acid Residue changes from one state to the other in a protein?

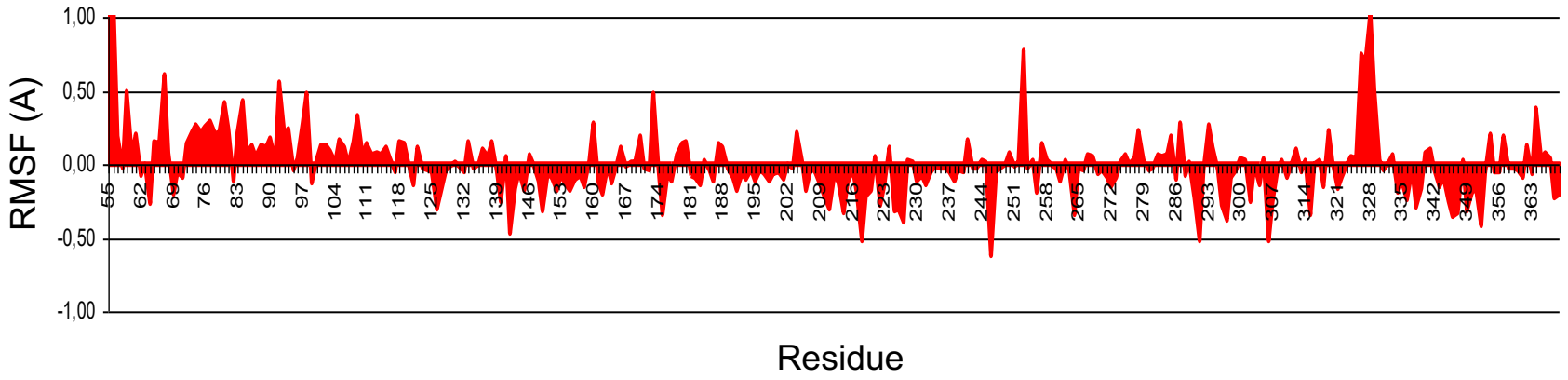


The Root Mean Square Fluctuation (RMSF)



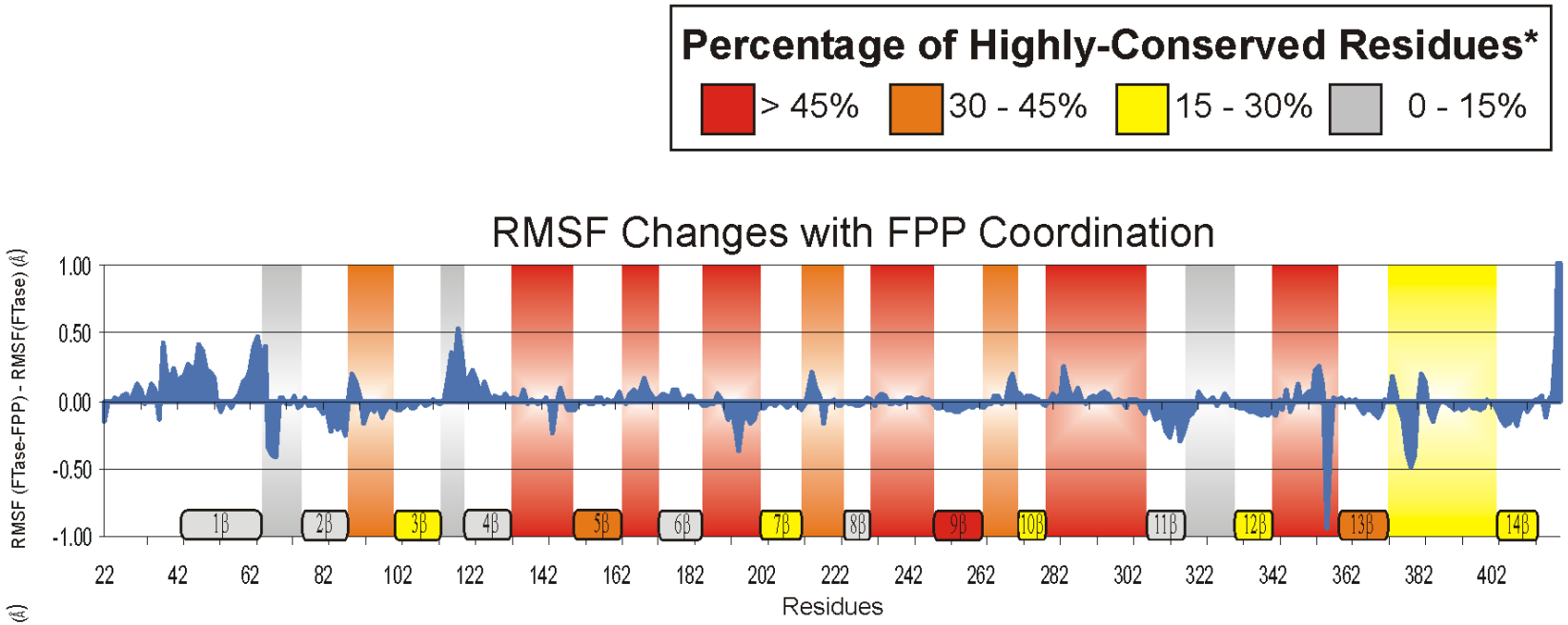
The Root Mean Square Fluctuation (RMSF)

Flexibility difference upon substrate binding? What changes in flexibility does ligand binding induce?

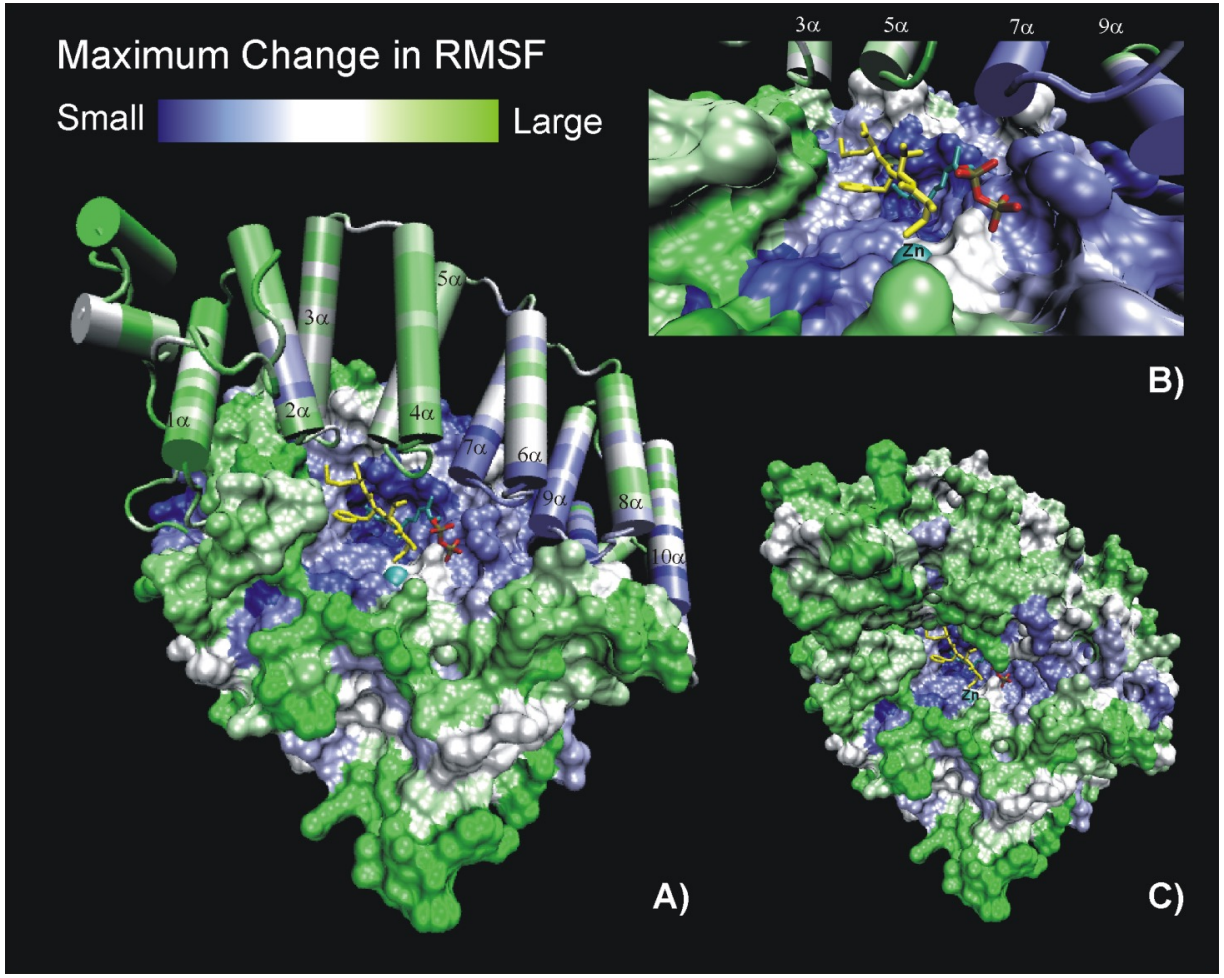


The Root Mean Square Fluctuation (RMSF)

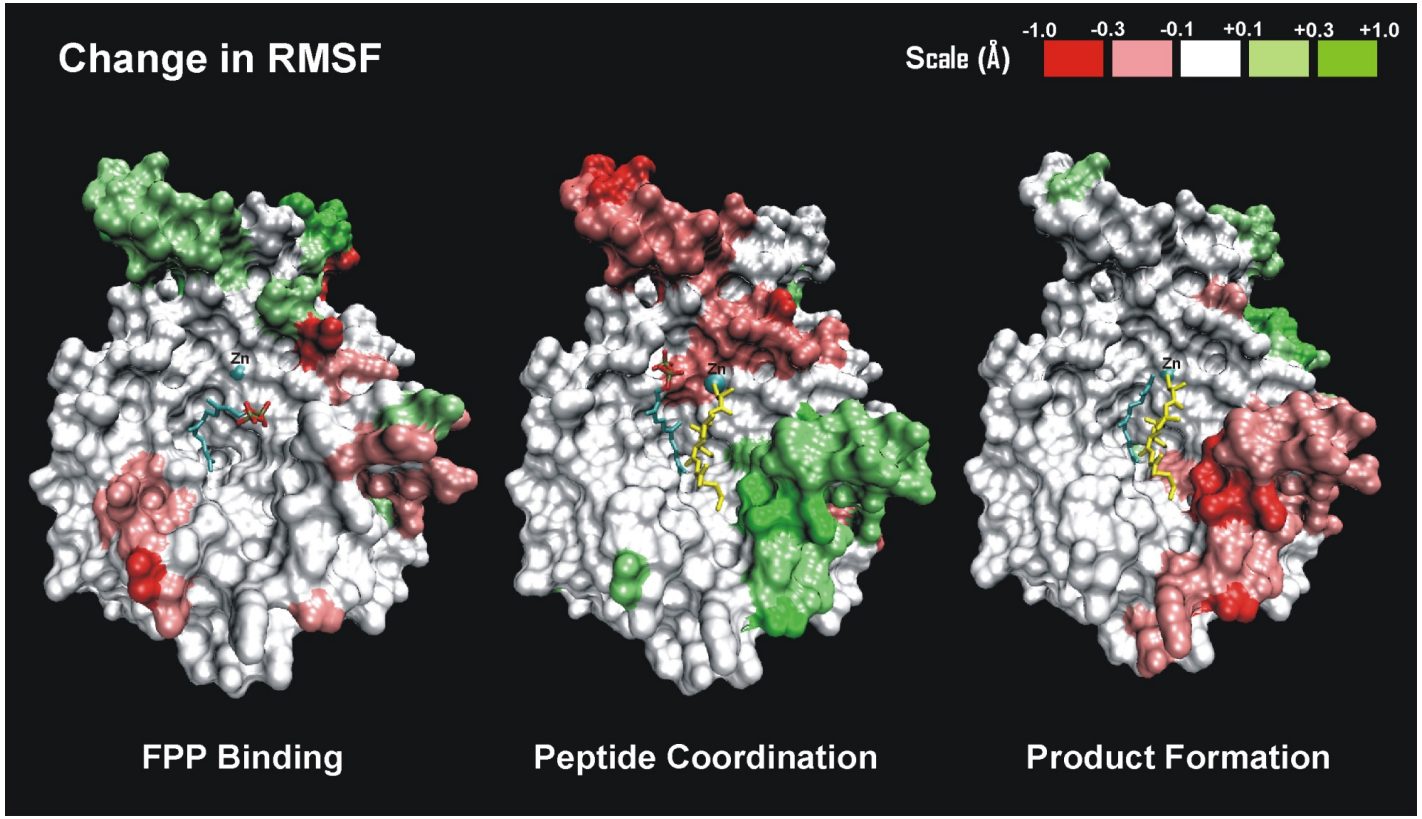
Flexibility difference upon substrate binding? What changes in flexibility does ligand binding induce?



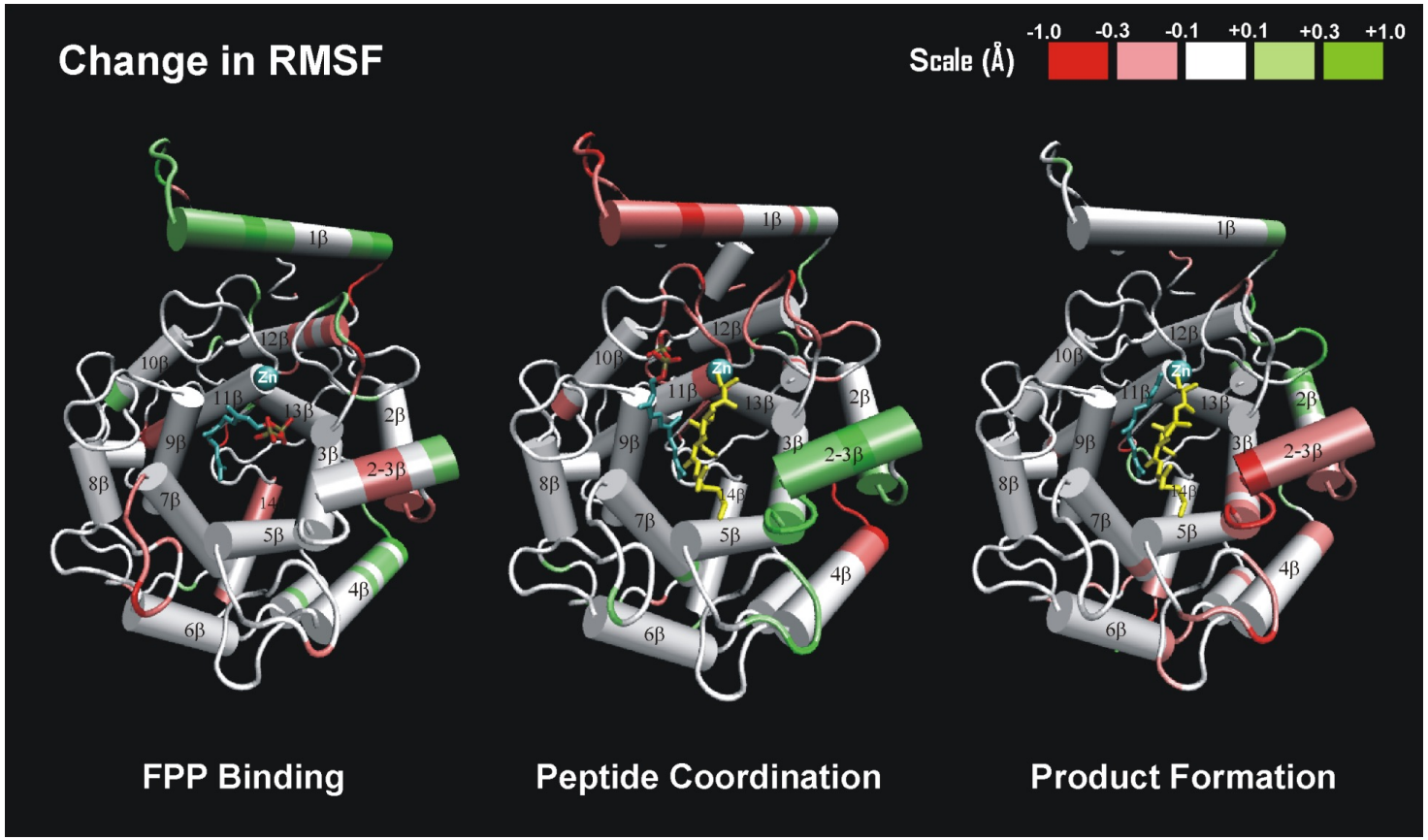
The Root Mean Square Fluctuation (RMSF)



The Root Mean Square Fluctuation (RMSF)

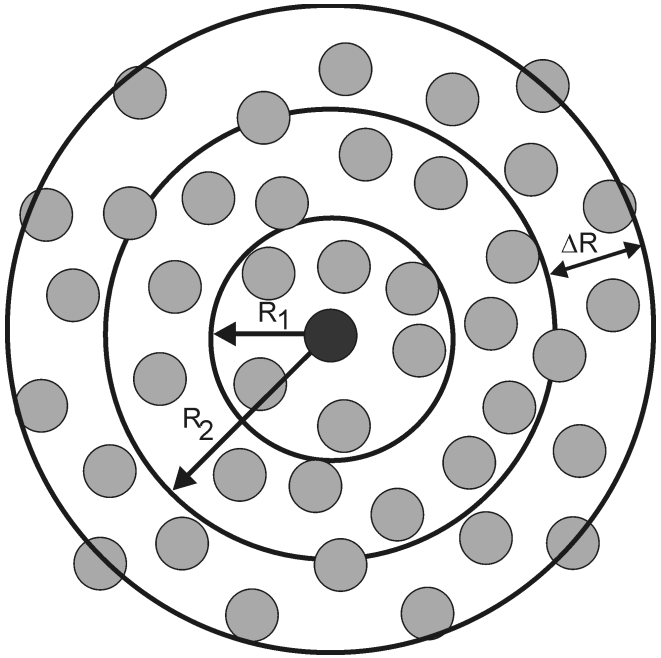


The Root Mean Square Fluctuation (RMSF)



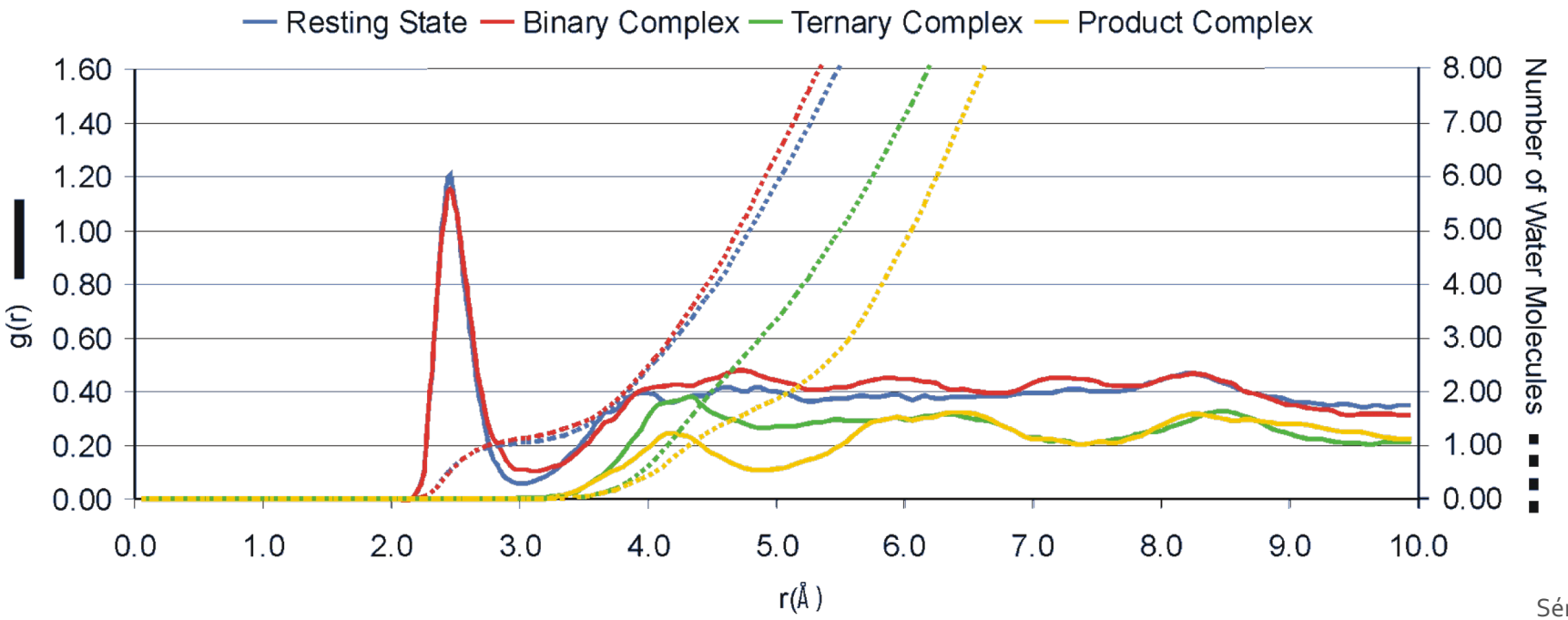
Radial Distribution Functions (RDFs)

Very useful in biomolecular simulations to evaluate the organization of solvent molecules (water) around biologically relevant groups on a protein for example.



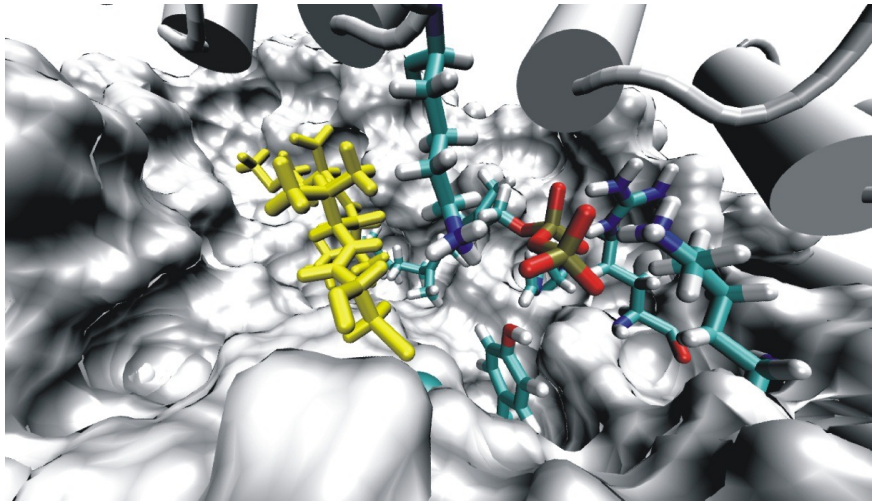
Radial Distribution Functions (RDFs)

Very useful in biomolecular simulations to evaluate the organization of solvent molecules (water) around biologically relevant groups on a protein for example.



Hydrogen Bonding Analysis

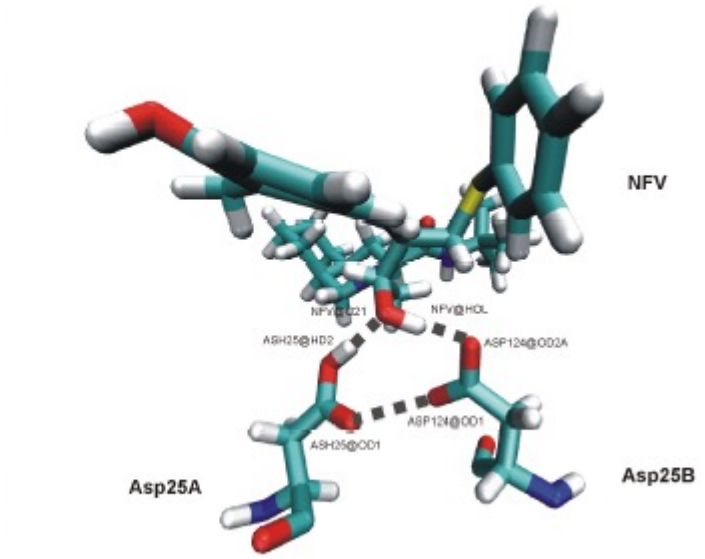
Statistical view of the hydrogen bonds formed and broken as a function of time

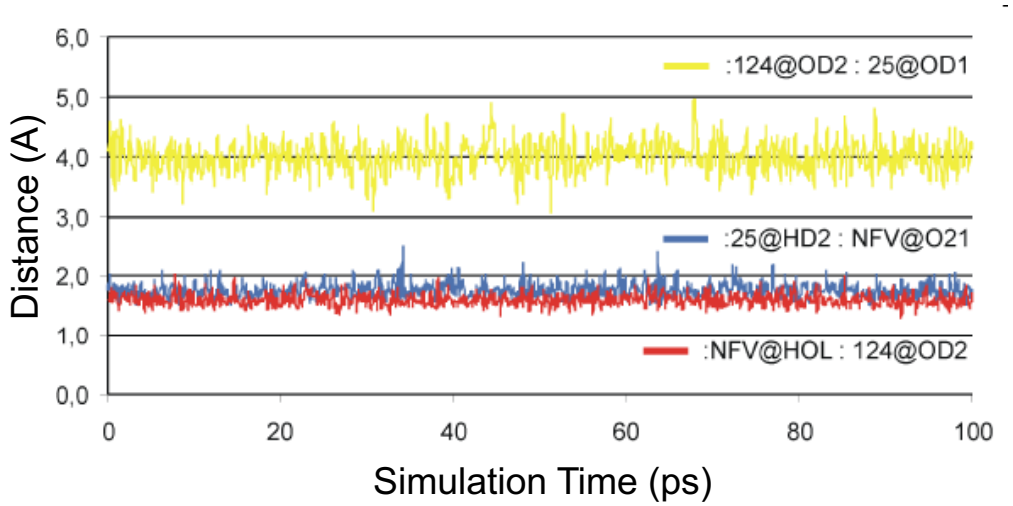


				%Occupied	Distance		Angle		Lifetime		MaxOcc
:732@O2A	Lys164A	:110@HZ3	:110@NZ	37.69	**2.795**	*0.09*	**26.98**	*13.74*	**5.4**	*6.4*	56
:732@O2A	Lys164A	:110@HZ2	:110@NZ	16.69	**2.793**	*0.09*	**28.17**	*14.12*	**4.2**	*6.6*	57
:732@O2B	HIS248B	:555@HE2	:555@NE2	6.22	**2.897**	*0.08*	**29.69**	*11.43*	**1.2**	*0.6*	6
:732@O2B	ARG291B	:598@HE	:598@NE	66.73	**2.811**	*0.09*	**24.15**	*9.67*	**11.9**	*14.7*	127
:732@O1B	LYS294B	:601@HZ3	:601@NZ	31.93	**2.766**	*0.10*	**35.26**	*13.91*	**4.0**	*5.4*	66
:732@O1B	LYS294B	:601@HZ1	:601@NZ	17.13	**2.762**	*0.09*	**35.64**	*14.22*	**3.8**	*6.8*	110
:732@O2B	TYR300B	:607@HH	:607@OH	13.73	**2.775**	*0.12*	**18.58**	*9.86*	**4.0**	*6.3*	76

Critical Distances – A Statistical view

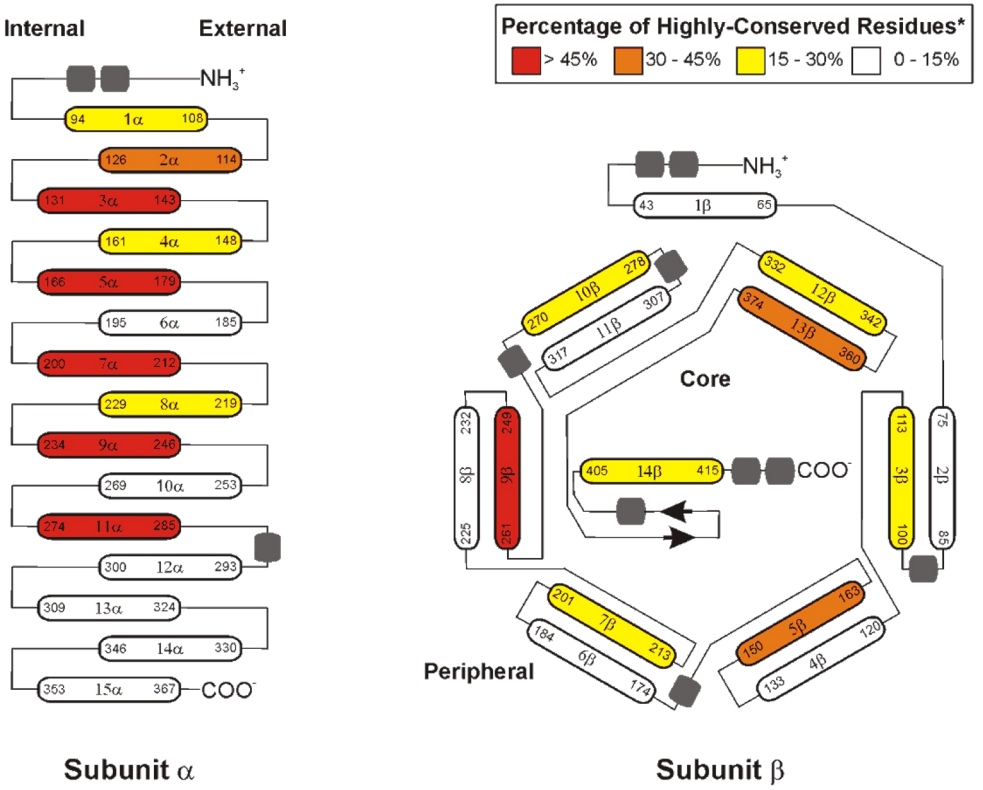
Dynamic View taking into consideration time evolution





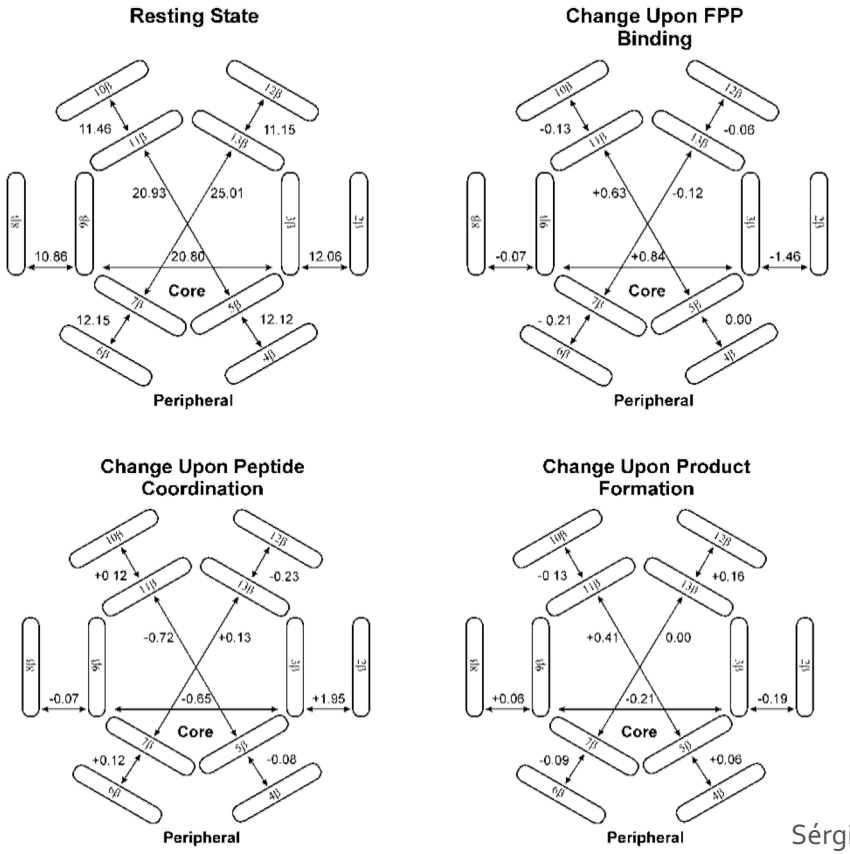
Critical Distances – A Statistical view

Dynamic View taking into consideration time evolution



Critical Distances – A Statistical view

Dynamic View taking into consideration time evolution



Critical Distances – A Statistical view