## 1. Summary

In the framework of the recently created Spanish CECAM (http://www.cecam.org) node, called ZCAM and located in Zaragoza, the workshop took place on September 2-4, 2010. About 30 national, European and international experts met to discuss about the theoretical and computational issues related to the **imposition of constraints** in molecular modeling.

The original aims of the workshop were, on the one hand, to achieve a general overview of this important problem with implications in many distinct fields related to Computational Chemistry and Physics, as well as Engineering, placing special emphasis in the open questions and in the possible future lines of research that are likely to bring significant advances in the field. The participation, in a unique forum, of researchers with different backgrounds, ranging from theoretical physicists to engineers expert in robot kinematics, facilitated a fruitful interdisciplinary discussion and an understanding of the problem at different levels.

Additionally, it was also an important objective of the workshop, and of others that will also take place in 2010 and 2011, to present the Spanish CECAM node to the Spanish scientific community and inform of its future activities.

All the objectives were met and the meeting was a success both from the point of view of the organizers and the attendants.

In the web http://neptuno.unizar.es/events/constraints2010/ more details about the workshop can be found, including most of the presentations.

## 2. Description of the scientific content of and discussion at the event

The workshop structure was based in key oral expositions followed by short discussion sessions longer than normal questions turns. Also, at the end of each day, round tables were held in which the most interesting points that appeared were tackled in more depth.

The content of the talks follow:

### *The canonical equilibrium of constrained molecular models*
**Pablo Echenique, Claudio Cavasotto, Pablo García-Risueño and José Luis Alonso**

With the aim of establishing a framework for some of the discussions that shall follow, we first introduce a classification for the constraints typically used in molecular simulations. Next, we discuss the loss of accuracy associated to the different type of constraints as evaluated by measuring the influence of the correcting terms that appear in the equilibrium probability density functions. We do this in small molecules and in peptide systems, using classical force fields.

### *Flexible constraints*
**Berk Hess**

In classical molecular simulations chemical bonds and bond angles have been modeled either as rigid constraints, or as nearly harmonic oscillators. However, neither model is a good description of a chemical bond, which is a quantum oscillator that in most cases occupies the ground state only. A so called flexible constraint is a more faithful representation. A flexible constraint constrains the effective force, potential plus rotational, to zero. This talk with explain the formalism behind and algorithms for flexible constraints. Furthermore, the use of flexible constraints will be illustrated through simulations of polarizable water in different phases.

### *Constrained macromolecular chains at thermal equilibrium: A quantum-mechanical approach*
**Ramón F. Álvarez-Estrada and Gabriel F. Calvo**

We review an approach yielding models for three-dimensional macromolecular chains, at thermal equilibrium, based on Quantum Mechanics. We start with Hamiltonians for N($\geq$ 1) non-relativistic quantum particles, interacting among themselves through potentials which include rapidly-varying vibrational ones. By using the Peierls inequality, for large vibrational frequencies, one gets effective quantum Hamiltonians for different kinds of constrained macromolecules: unhindered, hindered, open or closed, in which not only bond lengths are constrained but also angles can be constrained. The approach is related to the Born-Oppenheimer approximation. Those effective quantum Hamiltonians account for the slow motions of the remaining unconstrained (angular) variables. At about room temperature, we carry out a classical approximation of the quantum partition functions leading to classical ones, thereby justifying consistently previous purely classical approaches. As applications of the above program, both the consistency with the standard Gaussian model and approximate classical models for DNA are outlined. Based upon both the above general analysis and the well-known model of Peyrard and Bishop for the thermal denaturation of DNA, a model for a three-dimensional classical closed double-stranded macromolecular chain is treated succinctly.

### *Constraints: A geometric approach*
**José F. Cariñena**

The concept of constraint is reviewed from a geometric point of view. Both holonomic and nonholonomic constraints are considered and Dirac's theory of constraints for singular systems is also analysed from this geometric perspective. The use of an extended singular Lagrangian including some Lagrange multipliers is also pointed out.

### *The impetus-striction formulation of constrained Lagrangian dynamics*
**John Maddocks**

We are used to thinking that there is one essentially unique and most `natural' way to transform second-order Lagrangian dynamics into a first-order dynamical system, namely by introducing the appropriate momentum variable and obtaining the Hamiltonian form. I will explain why this uniqueness is lost in the presence of holonomic constraints. A different conjugate variable, the impetus, can be introduced which differs from the standard momenta by the striction, which is the Lagrange multiplier associated with the time differentiated constraint. The resulting Hamiltonian dynamics is unconstrained with the imposed constraints arising, by construction, as invariant sets of the dynamics. Moreover these invariant sets can be made to be attracting.

## *Holonomic constraints in statistical mechanics*

**Jürgen Schlitter**

Holonomic constraints are a most general means for facilitating the theoretical treatment of any system. They are used whenever variables are assumed to be constant and are to be regarded only as parameters. The method is somewhat artificial in nature, but gives three advantages: first, it simplifies the treatment by discarding undesired degrees of freedom, second, it feeds in external information, and third yields results for precisely defined situations. Constraints on bond lengths in molecular dynamics simulations are the most wide-spread example of the first category. For the second category, the maximum estimate of entropy from numerical fluctuations is a good example. The third category comprises many applications. The computation of equilibrium distributions from the maximum entropy principle is a most prominent case. In the theory of simulations, constraint forces were studied for their use at the calculation of free energy. While it was clear for a long time that free energy along a process coordinate should be the potential of mean force, the computation of that force remained difficult or even impossible. Introducing a constraint, however, may interfere with the dynamics. It was believed, but only lately proven that the mean force is the (negative) mean constraint force with a simple correction which often vanishes. Finally, equilibria can be defined by constraints, and their release allows to consider the transition to relaxation without exchange of energy. Recently we have shown that second-law conform behaviour of relaxation can be derived using this approach.

## *Strong confinement limit and two distinct notions of free energy*

**Carsten Hartmann**

Blue moon sampling is a widely-used technique for the sampling of rare events, e.g., for the calculation of free energy profiles. The idea behind it is to constrain a system to those regions in state space that would be rarely visited otherwise (i.e., once in a blue moon). The probability distribution of the rare events is eventually reconstructed by an appropriate reweighting of the constrained trajectories. Yet there has been (and still is) some confusion about how the weight is affected by the presence of momenta or velocities in the system. I will argue that reweighting is also an issue for first-order systems (e.g., Brownian motion) as can be explained by the often ignored difference between constrained systems and systems that are confined by infinitely stiff restraints. I will analyse confinement limits of first- and second-order systems and discuss its implications for the calculation of free energies.

## *Revisiting and parallelizing SHAKE*

**Yael Weinbach and Ron Elber**

An algorithm is presented for running SHAKE in parallel. SHAKE is a widely used approach to compute molecular dynamics trajectories with constraints. An essential step in SHAKE is the solution of a sparse linear problem of the type Ax = b, where x is a vector of unknowns. Conjugate gradient minimization (that can be done in parallel) replaces the widely used iteration process that is inherently serial. Numerical examples present good load balancing and are limited only by communication time. Finally a new approach is suggested (which has a fixed constraint matrix) for improved parallelization.

## *On the tensor character of the Brownian propagator*

**Wim Briels**

I will start by writing down the Lagrange equations of motion in generalized coordinates and introduce a bit of differential geometry. The resulting equations of motion are (by definition) manifestly covariant. The same holds for the Langevin propagator and the corresponding Fokker-Planck equation. After eliminating the velocities one obtains the Smoluchowski equation, which again is manifestly covariant. The corresponding Brownian propagator seems to lack this property. I will show that this inconsistency is only apparent.

## *Control of measure using degenerate diffusions*

**Benedict Leimkuhler**

The simulation of molecular and other systems requires formulations and numerical methods which constrain the invariant measure on phase space. Techniques such as Langevin dynamics can provide such a control, but this method may severely perturb dynamical autocorrelation functions, and it is delicate to implement a rigorous stochastic method in the presence of holonomic constraints and solution invariants. Instead, we argue here for the introduction of a highly degenerate diffusion in which stochastic perturbations are shielded from the physical variables by a dynamical coupling. These methods provide rapid convergence to equilibrium and in doing so introduce a much milder perturbation of autocorrelation functions: i.e. they are more "efficient." They are also easily implemented for constrained dynamics.

## *Isothermal-isobaric molecular dynamics with constraints: A simplified version of the ROLL algorithm and extension to the prediction of crystalline polymorphs*

**Mark E. Tuckerman**

While molecular dynamics calculations are often performed in the canonical or NVT ensemble, it is sometimes necessary to perform calculations under conditions of constant pressure. The Rigorous formulation of an isothermal-isobaric or NPT molecular dynamics algorithm for systems subject to holonomic constraints with a full atomic virial is a relatively recent development in the molecular dynamics field. A somewhat greater challenge is the generation of time-reversible integration algorithms for the equations of motion that rigorously preserve phase space volume. In 1996, we introduced the so-called ROLL algorithm that showed how to adapt the usual SHAKE and RATTLE algorithms to incorporate constraints under isothermal-isobaric conditions, however, the integration algorithm failed to preserve the phase space volume exactly. More recently, we introduced a simplified integrator for the NPT equations of motion also preserves the phase space volume, and we have now extended that approach to treat systems subject to holonomic constraints. The resulting algorithm achieves this while simultaneously simplifying the ROLL procedure. In addition, we have combined this new approach with an adiabatic enhanced free-energy sampling scheme for the prediction of crystalline polymorphs. This talk will focus on these recent developments and demonstrate their performance on a number of illustrative examples.

## *Corrected potentials for constrained dynamics*

**Robert Skeel**

Atomic oscillations present in classical molecular dynamics restrict the step size that can be used. Multiple time stepping schemes offer only modest improvements, and implicit integrators are costly and inaccurate. The best approach may be to actually remove the highest frequency oscillations by constraining bond lengths and bond angles, thus permitting perhaps a 4-fold increase in the step size. However, omitting degrees of freedom produces errors in statistical averages, and rigid angles do not bend for strong excluded volume forces. Presented here is an enhanced treatment for constrained dynamics that addresses these difficulties using ideas in papers of Fixman (1974) and Reich (1995, 1999). In particular, the goal is to simplify and clarify this prior work.

## *Extending dynamic calculations to the biological scale*

**Modesto Orozco**

Simulation of the dynamical behavior of proteins advances by many combined efforts: i) improvement of force-fields, ii) development of new computational platforms, iii) improvement of molecular dynamics algorithms, iv) generation of automatization tools for analysis and simulation set-up and v) creation of simplified methods able to reach biological time and size scales with a reduced computational cost. I will review here approaches to increase the impact of molecular simulation in biology, making special emphasis in new methods to approach simulations to the biological scale.

## *Constraints in protein-ligand and protein-protein docking*

**Juan Fernández-Recio**

The structural prediction of protein complexes is particularly important in the field of Structural Biology. Proteins interact with a variety of biomolecules to perform their functions, and the formation of complexes is a key point to understand the most important biological processes at molecular level. From a computational point of view, docking is the only technique that can provide ab initio structural information on the orientation modes between two interacting biomolecules. Unrestricted molecular dynamics are too costly to have predictive capabilities on the formation of complexes at a reasonable time, so several big constraints have to be put in place. The first constraint is the rigid-body assumption: the majority of docking methods constraint the internal flexibility of the interacting molecules, especially if they are proteins. Other methods like HADDOCK allow flexibility during the search, but in order to avoid excessive computational times, they need to impose some NMR-style restraints derived from previous knowledge about the system. The use of such restraints a posteriori, i.e. to filter previously generated docking orientations, gives similar practical results in faster times, like in our pyDockRST protocol and can be extended for multi-domain interactions. Further constraints can be imposed from low-resolution experimental techniques like SAXS. We will discuss here the advantages and limitations of these types of restraints in docking, and the application to real-scenario cases. The bottom-line is that the use of constraints in protein docking is not only a technical approach useful to overcome the high conformational search cost, but it is often an unavoidable option derived from the noise in the energy description and the incomplete representation of a system involved in multi-protein interactions.

## *To constrain or not to constrain*

**Sebastian Reich**

Conservative mechanical systems with highly oscillatory degrees of freedom are difficult to treat numerically. To simplify computations, one often makes a scale separation argument and replaces the oscillatory degrees of freedom by holonomic constraints. While computationally being efficient, there are two major conceptional difficulties associated with such an approach:

(i) scale separation rarely occurs in practice, and
(ii) if under scale separation fast and slow degrees can interact in a nontrivial manner through resonances.

Hence it seem preferable to develop true multi-scale integration methods that both resolve the fast oscillations while also being computational efficient. If the slow interactions form the most computationally demanding contributions, then mollified multiple-time-stepping methods (MOLLY) are a good option. MOLLY has been proposed by Skeel and co-workers. Using constraints to define the mollification yields a particularly robust approach (the method EQUILIBRIUM). If (ii) from above is not a major concern, then alternative constraint formulations can be used to address (i). These are essentially given by higher order correction terms to the constraining manifold and have led in joined work with Bernie Brooks to the concept of soft constraints.

## *Efficient formulations for the dynamics of multibody systems in mechanical engineering*

**Javier García de Jalón**

This paper reviews some ways to improve the efficiency of multibody dynamic formulations that have been published recently. First, two global formulations that speed up the dynamic simulations using dependent Cartesian coordinates (i. e., that maintain the descriptor form) are considered; one of them enforces the constraint equations by Lagrange multipliers and the other one by the penalty method. These new formulations are faster than the conventional formulations that use the descriptor form, but they can never attain the efficiency of the formulations that use relative or joint coordinates, which can easily take advantage of the system topology. The ideas behind these improvements of global formulations can be used to improve the topological formulations when they are applied to closed-loop multibody systems.

With relative coordinates there are no constraint equations for open chain systems, but systems with closed loops shall include the closure of the loop constraint equations. These constraint equations are more complicated to take into account with fully recursive O(n) formulations, which are the fastest for long, open-chain systems. For this reason, several semi-recursive formulations have been developed in the last few years. They are expected to be less efficient than the fully recursive ones, but they are simpler and easier to implement. In the second part of this paper, some semi-recursive formulations are reviewed and a new variant, that is simpler and more general, is described in detail. A simple way to introduce the topology of the spanning tree is explained. Special attention is paid to closed-loop multibody systems with rods, and the benefits of opening the loops by removing these rods −while keeping its inertia forces exactly− are explained. Some examples and numerical results illustrate the aforesaid theoretical developments.

## *A biokinematic approach for the computational simulation of protein motion*

**Mikel Díez**

Proteins play an essential role in biological processes that occur in human cells. For this reason it is important to analyse their movement to understand the molecular mechanisms in which they are involved. Recently a new approach based on the parallelism between proteins and spatial mechanisms have opened new ways to calculate their atomic trajectories using concepts of kinematics of mechanisms. In this communication is presented a methodology for simulating the movement of a protein that is based on the evaluation of its potential energy along its motion, eliminating the need for a minimization process. This procedure presents a good compromise between computational effort, and accuracy obtained. It also implements an algorithm to normalize the protein structure focused on achieving efficient simulation of its movement. Finally we present the results of implementing the procedures submitted to the inorganic pyrophosphatase (famila II) from Streptococcus gordonii.

## 3. Assessment of the results and impact of the event on the future direction of the field

As a result of the discussions in the meeting, many new collaborations between the attendants have emerged, which, in the view of the organizers, is one of the most relevant signs of success for a scientific event.

Another achievement has been to gather together in the same place most of the experts that are nowadays interested in the issues related to the imposition of constraints in molecular simulations. This has allowed to establish a very precise overview of the whole problem, avoiding the risk of leaving important points out.

The more relevant problems that have been identified and extensively discussed in the workshop are:

1. The existence of many different interpretations of the free energy of a constrained system, including the correcting terms that appear in the equilibrium distribution. A number of collaborations have emerged in the meeting to tidy and rationalize this issue.
2. The question of whether or not constraints can be imposed beyond bond lengths (e.g., in bond angles) without altering too much the physical model. If this could be done, the time step used in simulations could be doubled, together with the efficiency of the calculation. This question, which is very much related to the possibility of implementing flexible constraints, has also implications in the convergence and parallelizability of existing algorithms, and a number of attendants have agreed in exploring it further.
3. The possibility of justifying constrained models (from the dynamical point of view, which is more difficult than in equilibrium) from the more fundamental unconstrained scheme, and through techniques like averaging or coarse-graining. A number of attendants are engaged in the investigation of this point.
4. The need to use quantum mechanical arguments to properly describe the hardest constrained degrees of freedom in typical molecular systems. Two groups that have attended the workshop are exploring this line of research.

We believe that both the thorough overview of the field achieved in the workshop and the identification of the most important problems that remain unsolved will help to focus the efforts and to advance in the knowledge of constraints techniques, which are central in the search for computational efficiency in molecular simulations.

# 4. Final programme of the meeting

## Day 1
Thursday, September 2

*Theoretical and modeling aspects*

| | |
|---|---|
| 9:00 | Registration |
| 9:30 | *Welcome and introduction to the workshop*<br>**The organizers** |
| 9:50 | *Introductory talk to ZCAM*<br>**Michel Mareschal, ZCAM Director** |
| 10:10 | *The canonical equilibrium of constrained molecular models*<br>**Pablo Echenique** |
| 10:35 | Discussion |
| 10:50 | Coffee |
| 11:20 | *Flexible constraints*<br>**Berk Hess** |
| 11:45 | Discussion |
| 12:00 | *Constrained macromolecular chains at thermal equilibrium: A quantum-mechanical approach*<br>**Ramón F. Álvarez-Estrada** |
| 12:25 | Discussion |
| 12:40 | *Constraints: A geometric approach*<br>**José F. Cariñena** |
| 13:05 | Discussion |
| 13:20 | Lunch |
| 15:20 | *The impetus-striction formulation of constrained Lagrangian dynamics*<br>**John Maddocks** |
| 15:45 | Discussion |
| 16:00 | *Holonomic constraints in statistical mechanics*<br>**Jürgen Schlitter** |
| 16:25 | Discussion |
| 16:40 | *Strong confinement limit and two distinct notions of free energy*<br>**Carsten Hartmann** |
| 17:05 | Discussion |
| 17:20 | Coffee |
| 17:50 | General discussion of the day (round table) |
| 19:00 | End of day 1 |

## Day 2
Friday, September 3

*Numerical aspects and applications*

| | |
|---|---|
| 9:30 | *Revisiting and parallelizing SHAKE*<br>**Ron Elber** |
| 9:55 | Discussion |
| 10:10 | *On the tensor character of the Brownian propagator*<br>**Wim Briels** |
| 10:35 | Discussion |
| 10:50 | Coffee |
| 11:20 | *Control of measure using degenerate diffusions*<br>**Ben Leimkuhler** |
| 11:45 | Discussion |
| 12:00 | *Isothermal-isobaric molecular dynamics with constraints: A simplified version of the ROLL algorithm and extension to the prediction of crystalline polymorphs*<br>**Mark Tuckerman** |
| 12.25 | Discussion |
| 12:40 | *Corrected potentials for constrained dynamics*<br>**Robert Skeel** |
| 13.05 | Discussion |
| 13:20 | Lunch |
| 15:20 | *Extending dynamic calculations to the biological scale*<br>**Modesto Orozco** |
| 15:45 | Discussion |
| 16:00 | *Constraints in protein-ligand and protein-protein docking*<br>**Juan Fernández-Recio** |
| 16:25 | Discussion |
| 16:40 | *To constrain or not to constrain*<br>**Sebastian Reich** |
| 17:05 | Discussion |
| 17:20 | Coffee |
| 17:50 | General discussion of the day (round table) |
| 19:00 | End of day 2 |

## Day 3
Saturday, September 4

*Robot kinematics and conclusions*

| | |
|---|---|
| 9:30 | *Efficient formulations for the dynamics of multibody systems in mechanical engineering*<br>**Javier García de Jalón** |
| 9:55 | Discussion |
| 10:10 | *A biokinematic approach for the computational simulation of protein motion*<br>**Mikel Díez** |
| 10:35 | Discussion |
| 10:50 | Coffee |
| 11:20 | *Summary talk*<br>**Giovanni Ciccotti** |
| 12:00 | General discussion of the workshop (round table) |
| 13:00 | Closing of the workshop |