



## Molecular modelling and bioinformatics group

Our long term objective is to decipher the behaviour of living organisms by means of theoretical models, whose roots are anchored in the basic principles of physics and chemistry. For this purpose, we work with a range of methodologies, from the mining of biological databases to classical dynamics and quantum chemistry calculations. The use of such diverse approaches allows us to explore problems as diverse as drug design and genome analysis. Special emphasis is given to connecting basic interactions with the global properties of biological systems. During this period, our work has focused on three major areas: i) the study of small model systems; ii) the analysis of stressed or unusual nucleic acids; and iii) the dynamics of proteins.

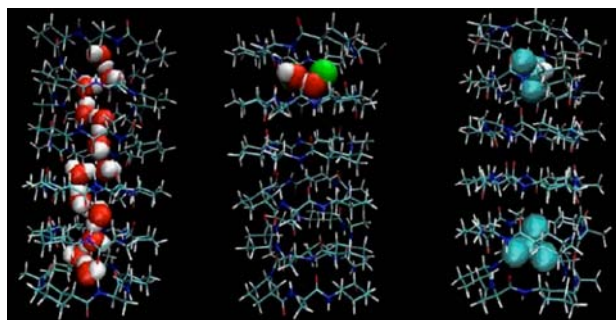


Figure 1. Detail of solvent migration in a  $\alpha,\gamma$ -peptide nanotube.

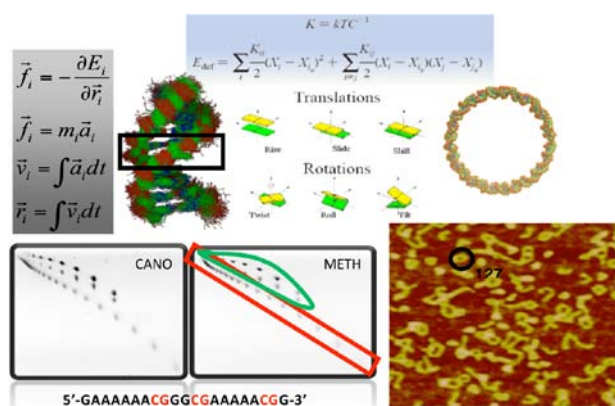


Figure 2. Derivation of mesoscopic descriptors of DNA flexibility and validation by cyclization experiments at EBL.

### Small model systems

The study of simple systems can provide clues to enhance our understanding of the behaviour of much more complex biological molecules. In this area of work our effort has been traditionally focused on the development of methods for the treatment of solvation and the analysis of supramolecular systems of biological impact. During 2009, we have advanced in the refinement of our MST method for the representation of solvent (Klamtz *et al*, 2009), which in recent blind tests (Soteras *et al*, 2009) was found to be superior to most self-consistent reaction field approaches, providing not only good estimates of solvation free energies, but also enthalpies (Bidon-Chanal *et al*, 2009). Applying this and other methods, we have studied several supramolecular systems of biological/technological impact. Special mention is given to our work on  $\alpha,\gamma$ -peptide nanotubes (Garcia-Fandiño *et al*, 2009; Figure 1), in which we characterise the structure, dynamics and transport properties of these molecules, which have been proposed as potential carriers of small molecules across membranes.

### Analysis of stressed or unusual nucleic acids

Major breakthroughs in the field of nucleic acid simulations emerged from work performed by our group in 2007 related to the development of the PARMBSCO force-field. This formalism has become the default force-field for nucleic acid simulations and was selected by the Ascona B-DNA consortium to simulate the sequence-dependent geometrical and dynamic properties of DNA (Lavery *et al*, 2009; reviewed in Laughton & Orozco, 2009). Using this force-field, we have advanced considerably in the characterisation of the elastic properties of DNA and their dependence on environmental conditions, external stress or chemical modifications. In summary, during 2009 we have analysed the impact of ionic atmosphere on DNA (Noy *et al*, 2009), the structure of unusual loop-structures in DNA quadruplexes (Fadrna *et al*, 2009), the elastic properties of all DNA tetraplexes in physiological conditions (Lavery *et al*, 2009), and the changes induced by epigenetic modifications on DNA (Perez *et al*, in preparation). Our models are now being validated through



cyclation experiments performed at the Experimental Bioinformatics Laboratory (EBL). Following this line of work, we are in the process of completing the development of a new nucleosome predictor, which is currently being validated by next-generation sequencing and tiling arrays.

Following a well-established research line in the laboratory, we have explored the structure of nucleic acids containing modified nucleotides. In particular, alone or in collaboration with other research groups, we have analysed the impact of introducing restricted nucleotides in the G-DNA structure of the thrombin aptamer (Saneyoshi *et al*, 2009), the impact of thio-thymines (Faustino *et al*, 2009), seleno derivatives of thymine (Vázquez-Mayagoitia *et al*, 2009) and fluoro-arabino compounds (Watts *et al*, 2009) in duplexes, and the 8-amino purine derivatives in hairpins and triplexes (Aviño *et al*, 2009). Finally, of note are our recent efforts in the characterisation of nucleic acid folding and unfolding pathways (Pérez *et al*, in press; Portella *et al*, in preparation), studies that will crystallise during 2010.

### **Dynamics of proteins**

The creation and data-mining of the MODEL (Molecular Dynamics Extended Library) database has involved a huge amount of work. We have not only completed the database, but used it to characterise the connection between evolutionary and physical deformation patterns in proteins (Velázquez-Muriel *et al*, 2009). This general analysis has been centred on the RAS family (Raimondi *et al*, 2009), where flexibility was found to be crucial to explain the relation between flexibility and function. We have also used the information in MODEL to parametrise a wide variety of coarse-grained methods, which have greatly facilitated the study of deformability in systems containing thousands of proteins (Emperador *et al*, 2009). Recently these different approaches are being implemented in automatic tools for the description of protein flexibility, such as FlexServ (<http://mmb.pcb.ub.es/FlexServ>; Figure 4).

The latest technological breakthrough in the group in this field is related to the development of applications for automatic trajectory generation (MdWeb), an application for which a beta-version is already available (Figure 5), and the COCO utility for combining NMR and physical description of flexibility (Laughton *et al*, 2009).

### **Research Group Members**

#### **Principal Investigator:**

Modesto Orozco

#### **Research Associates:**

Agustí Emperador, Josep Lluís Gelpí, Manuel Rueda

#### **Postdoctoral Fellows:**

Neva Besker, Oliver Carrillo, Kathryn Collinet, Santiago Esteban, Athi Narayanan, Guillem Portella, Nadine Simone

#### **PhD Students:**

Annalisa Arcella, Özgen Deniz, Ignacio Faustino, Óscar Flores, Adam Hospital, Laura Orellana

#### **Research Assistants:**

José Alcántara, Carlos Fenollosa, Chiara Lara, Margarita Pedro





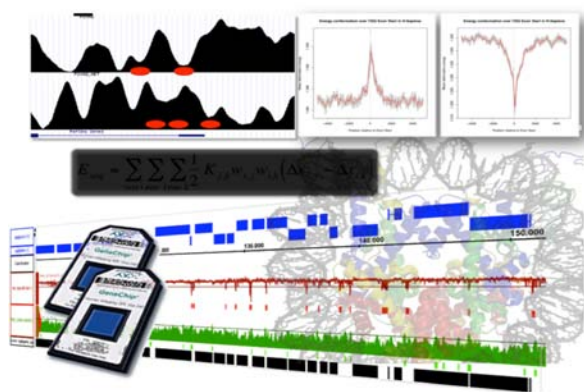


Figure 3. Prediction and experimental validation by tiling array of nucleosome positioning in *S. cerevisiae* genome.

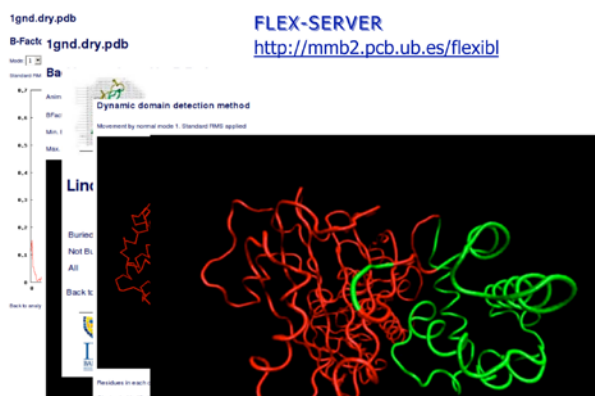


Figure 4. Examples of output of our FlexServ application.

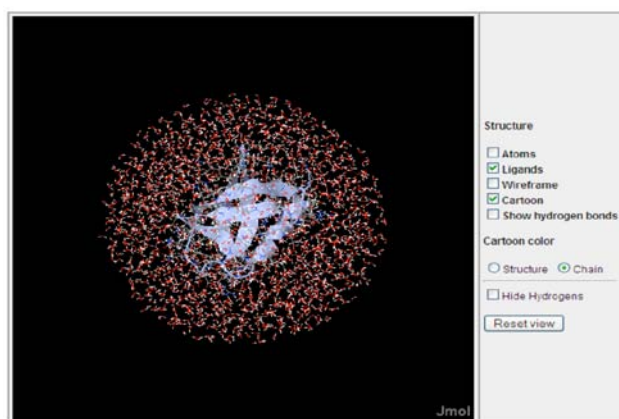


Figure 5. Graphical interface in the MDWeb application.

Significant work has been done during 2009 on the description of protein structure and flexibility under extreme conditions, particularly on gas phase (Meyer *et al*, 2009). The MD simulations demonstrated that the gas-phase ensemble of conformations is not far from the solution conformation, and, in fact, we demonstrated that collision cross-sections determined for gas-phase ensembles contribute to refining structural models derived from threading, *ab initio* or homology modelling (D'Abramo *et al*, 2009). The study opens up the possibility to use gas-phase structural information derived from X-free electron laser microscopy to obtain the protein structure in solution.

## Scientific output

### Publications

Bidon-Chanal A, Huertas O, Orozco M and Luque FJ. Solvation enthalpies of neutral solutes in water and octanol. *Theor Chem Acc*, **123**, 11-20 (2009)

Camps J, Carrillo O, Emperador A, Orellana L, Hospital A, Rueda M, Cicin-Sain D, D'Abramo M, Gelpí JL and Orozco M. FlexServ: an integrated tool for the analysis of protein flexibility. *Bioinformatics*, **25**(13), 1709-10 (2009)

D'Abramo M, Meyer T, Bernadó P, Pons C, Fernández-Recio J and Orozco M. On the use of low-resolution data to improve structure prediction of proteins and protein complexes. *J Chem Theory Comput*, **5**, 3127-37 (2009)

Emperador A, Meyer T and Orozco M. Protein flexibility from discrete molecular dynamics simulations using quasi-physical potentials. *Proteins*, Epub Aug 5 (2009)

Fadrná E, Spacková N, Sarzyńska J, Kocá J, Orozco M, Cheatham III T, Kulinski T and Sponer J. Single stranded loops of quadruplex DNA as key benchmark for testing nucleic acids force fields. *J Chem Theor Comput*, **5**, 2514-30 (2009)

Faustino I, Aviño A, Marchán I, Luque FJ, Eritja R and Orozco M. Unique tautomeric and recognition properties of thioketothymines? *J Am Chem Soc*, **131**(35), 12845-53 (2009)

García-Fandiño R, Granja JR, D'Abramo M and Orozco M. Theoretical characterization of the dynamical behavior and transport properties of alpha,gamma-peptide nanotubes in solution. *J Am Chem Soc*, **131**(43), 15678-86 (2009)

Klamt A, Mennucci B, Tomasi J, Barone V, Curutchet C, Orozco M and Luque FJ. On the performance of continuum solvation methods. A comment on 'Universal approaches to solvation modeling'. *Acc Chem Res*, **42**(4), 489-92 (2009)

Laughton CA and Orozco M. Nucleic acid simulations themed issue. *Phys Chem Chem Phys*, **11**(45), 10541-42 (2009)

Lavery R, Zakrzewska K, Beveridge D, Bishop TC, Case DA, Cheatham T 3rd, Dixit S, Jayaram B, Lankas F, Laughton C, Maddocks JH, Michon A, Osman R, Orozco M, Perez A, Singh T, Spackova N and Sponer J. A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Res*, Epub Oct 22 (2009)

Meyer T, de la Cruz X and Orozco M. An atomistic view to the gas phase proteome. *Structure*, **17**(1), 88-95 (2009)

Noy A, Soteras I, Luque FJ and Orozco M. The impact of monovalent ion force field model in nucleic acids simulations. *Phys Chem Chem Phys*, **11**(45), 10596-607 (2009)

Saneyoshi H, Mazzini S, Aviñó A, Portella G, González C, Orozco M, Marquez VE and Eritja R. Conformationally rigid nucleoside probes help understand the role of sugar pucker and nucleobase orientation in the thrombin-binding aptamer. *Nucleic Acids Res*, **37**(17), 5589-601 (2009)

Soteras I, Forti F, Orozco M and Luque FJ. Performance of the IEF-MST solvation continuum model in a blind test prediction of hydration free energies. *J Phys Chem B*, **113**(27), 9330-34 (2009)

Vázquez-Mayagoitia A, Huertas O, Brancolini G, Migliore A, Sumpter BG, Orozco M, Luque FJ, Di Felice R and Fuentes-Cabrera M. Ab initio study of the structural, tautomeric, pairing, and electronic properties of seleno-derivatives of thymine. *J Phys Chem B*, **113**(43), 14465-72 (2009)

Velázquez-Muriel JA, Rueda M, Cuesta I, Pascual-Montano A, Orozco M and Carazo JM. Comparison of molecular dynamics and superfamily spaces of protein domain deformation. *BMC Struct Biol*, **9**, 6 (2009)

Xie W, Orozco M, Truhlar D and Gao J. X-pol potential: An electronic structure-based force field for molecular dynamics simulation of a solvated protein in water. *J Chem Theor Comput*, **5**(3), 459-67 (2009)

## Research networks and grants

### *Acción complementaria*

Spanish Ministry of Science and Innovation, ExpandingBio (2009)  
Principal investigator: Modesto Orozco

### *Estudio de formas inusuales o tensionadas del DNA. Implicaciones biotecnológicas y biomédicas*

Spanish Ministry of Science and Innovation, BIO2006-01602 (2006-2009)  
Principal investigator: Modesto Orozco

### *IFLPD*

European Commission, Marie Curie-FP7-PEOPLE-2007-4-1-IOF (2009-2013)  
Principal investigator: Manuel Rueda

### *Molecular recognition*

'Marcelino Botin' Foundation, IO FMBotin-M Orozco (2007-2010)  
Principal investigator: Modesto Orozco

### *Red temática de investigación cooperativa en biomedicina computacional (COMBIOMED)*

Carlos III Health Institute (ISCIII), RD07/0067/0009 (2008-2012)  
Principal investigator: Modesto Orozco

### *Simulaciones de formas inusuales o tensionadas de los ácidos nucleicos de potencial interés biotecnológico o biomédico*

Spanish Ministry of Science and Innovation, BIO2009-10964 (2009-2012)  
Principal investigator: Modesto Orozco

### *Supercomputación y eCiencia*

Spanish Ministry of Science and Innovation, CSD2007-00050 (2007-2012)  
Principal investigator: Modesto Orozco

## Collaborations

### *Development of new tools for computer assisted drug design*

Francisco Javier Luque, University of Barcelona (Barcelona, Spain)

### *Drug design*

Lluís Ribas de Pouplana, IRB Barcelona (Barcelona, Spain)

### *Dynamics of proteins*

Francesca Fanelli, University of Modena (Modena, Italy); José María Carazo, Centro Nacional de Biotecnología-CSIC (Madrid, Spain)

### *Mixed QM-MM methods for protein simulations*

Donald Truhlar, University of Minnesota (Minnesota, USA); Jiali Gao, University of Minnesota (Minnesota, USA)

### *SCRF solvation methods*

Andrea Klamtz, Cosmologic Inc (Leverkusen, Germany); Francisco Javier Luque, University of Barcelona (Barcelona, Spain); Jacopo Tomasi, University of Pisa (Pisa, Italy)

### *Study of modified nucleobases*

Ramon Eritja, IRB Barcelona (Barcelona, Spain); Miguel Fuentes-Cabrera, Oak Ridge National Laboratory (Oak Ridge, USA); Carlos González, Rocasolano Institute-CSIC (Madrid, Spain); Francisco Javier Luque, University of Barcelona (Barcelona, Spain)

### *Study of physical properties of DNA*

ABC Consortium; Charlie Laughton, Centre for Biomolecular Sciences, Nottingham University (Nottingham, UK); Richard Lavery, Institut de Biologie et Chimie des Protéines, University of Lyon (Lyon, France)

## Awards and honours

Distinguished fellow, 'Marcelino Botin' Foundation  
Awardee: Modesto Orozco