# Molecular modelling and bioinformatics unit

**Principal Investigator**
Modesto Orozco (UB)

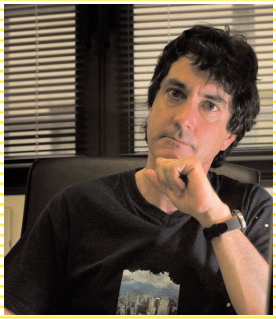
**Research Associates**
Xavier de la Cruz (ICREA)
Agusti Emperador
Juan Fernández
Josep Lluis Gelpí

**Postdoctoral Fellows**
Oliver Carrillo
Carles Ferrer
Rebeca García
Manuel Rueda

**PhD Students**
Damian Blanco
Solène Grosdidier
Adam Hospital
Sergio Lois
Iván Marchán
Tim Meyer
Agnés Noy
Alberto Pérez
David Piedra
Josep Ramón Goñi
Albert Solernou
David Talavera

**Research Assistants**
Iván Párraga
Margarita Pedro
Iván Sánchez

**Visitor**
Jiali Gao (USA)


Modesto Orozco

Our long term objective is to understand the behaviour of living organisms by means of theoretical models, the roots of which are anchored in the basic principles of physics and chemistry. With this aim, we work with a number of methodologies, from mining of biological databases to classical dynamics and quantum chemistry calculations. This wide range of methodologies has allowed us to explore topics as diverse as enzyme reactivity and genome analysis.

**Small model systems**
Our group has a long trajectory in the study of small model systems of biological importance (nucleobase complexes, drugs, isolated complexes of amino acids, stacked or hydrogen-bonded complexes). The analysis of these simple systems can contribute to our understanding of the behaviour of more complex biological molecules. Almost a decade ago, we realised that these studies were simple in the gas phase but very difficult in aqueous solution, thereby hampering the real application of the information obtained to the biological scenario. This consideration led us to the development of methods to describe solvent, some of which are regarded as the "state of the art" in the field (for a review see Orozco and Luque, 2000).

During 2006 we have continued to develop and improve methods to study condensed phases and to study the application of these and other methods to describe model systems of biological interest. Thus, in collaboration with Tomasi's group, we have improved continuum methodology by introducing dispersion contributions in an accurate way (Curutchet *et al*, 2006), a methodology successfully applied to complex chemical processes (Soteras *et al*, 2006; Blas *et al*, 2006). We have also made progress in transferring the methodology developed towards the macromolecular scenario (nucleic acids: Muñoz-Muriedas *et al*, 2006 or proteins: Talavera *et al*, 2006). Regarding application, we should highlight the first complete *ab initio* atlas of nucleobases-nucleobase stacking (Sponer *et al*, 2006) done by our group in collaboration with the team led by Hobza. We have also worked intensively on the analysis of expanded oligonucleotides, non-natural derivatives of DNA bases that can be incorporated in DNA, thereby changing its properties. We have explored the behaviour of these molecules and the differences they are expected to introduce in the structure/reactivity of DNA (Huertas *et al*, 2006a, 2006b; Fuentes *et al*, 2006). These model calculations have made it possible to conduct current work on the simulation of large fibers of expanded nucleobases (Huertas *et al*, unpublished).

***Structural studies on proteins***
Our group has extensive experience in the study of individual proteins of biological importance. Our studies are typically in collaboration with experimental groups. For example, during 2005 and 2006, we collaborated with Prof. Estrin's group in the analysis of mechanisms of ligand diffusion and reaction of *Mycobacterium tuberculosis* truncated-hemoglobin-N (Crespo *et al*, 2005; Bidon-Chanal *et al*, 2006). This enzyme is responsible for NO detoxification, and its overexpression leads to the resistance of *Mycobacterium tuberculosis* to macrophage action. In particular, during 2006 we characterised the main diffusion pathways for NO and $O_2$ in the enzyme, thereby locating the bottlenecks and control points of the process. An additional project, in collaboration with a pharmaceutical company, has led to the defi-

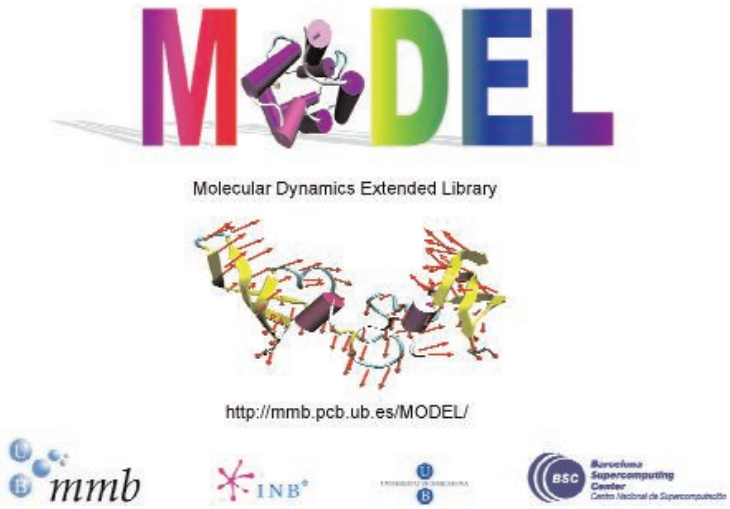



***Figure 1**. Main page of the MODEL webserver and database.The analysis of this massive database of trajectories gives a complete consensus picture of protein dynamics. The proteins are found to be melted solids (the core being solid-like and the surface being similar to a dense liquid). The essential dynamics of the proteins is clearly printed in their 3D structure, independently of the force-field and exact simulations conditions. Hydrogen bond interactions are the stiffer ones inside the proteins, while quite surprisingly, the saline bridge appears to be highly labile (Rueda* et al*, 2006). Overall, our study tested the capacity of force-field simulations to describe protein dynamics in physiological conditions.*

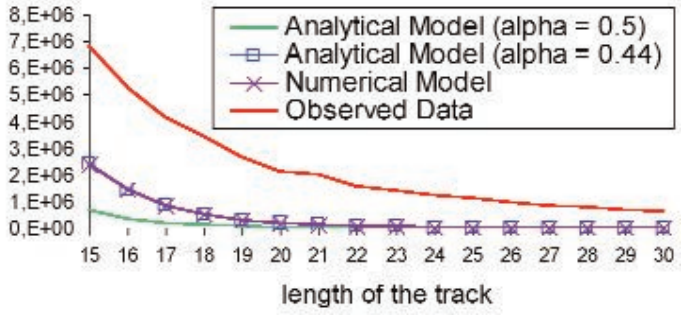



***Figure 2**. Representation of the number of nucleotides in triplex-forming oligonucleotides of different lengths compared with two background models.*

nition of the 3D recognition properties of a new family of $P_{38}\alpha$ MAP Kinase inhibitors which are now being tested as powerful anti-inflammatory compounds in phase II clinical trials (Soliva *et al*, 2006).

Since 2004 we have expanded our interest in proteins to the study of more general aspects of their structure and flexibility. The installation of the *MareNostrum* computer at the Barcelona Supercomputer Center (BSC-CNS) has opened up unexpected possibilities to perform proteome-scale studies. This is the origin of MODEL (Molecular Dynamics Extended Library; see Figure 1), the largest database of molecular dynamics simulations, currently containing information on the dynamics of around 1300 proteins (see http://mmb.pcb.ub.es/MODEL). The project has produced the first results, published during 2006 (Meyer *et al*, 2006 and Rueda *et al*, 2006a). The most remarkable of these studies is the definition of γMODEL, a restricted set of around 40 highly representative proteins which was simulated for 10 (some cases 100) ns using the four most commonly used force-fields (Rueda *et al*, 2006a).

**Genome mining studies**

Like many bioinformatics groups, our interest lies in obtaining information on key biological processes by means of the analysis of available biological databases. Our efforts have focussed on two directions: i) analysis of genomes to determine a pattern of DNAs with unusual structures and ii) development of predictive tools to determine the pathological character of single nucleotide polymorphisms.

Within the first project, during 2006, we have completed the mapping of TTS (triplex targeted sequences) in the human genome (Goñi *et al*, 2006). We have found that these triplexes are many times more common in the human genome than in background models (see Figure 2), and that they are overconcentrated in promoter regions, especially of those genes related to the control of cellular functions. These results open up exciting possibilities for triplex strategies, such as anti-gene therapies, since triplex formation at promoters are known to knock-out or knock-down genes.

We have continued to make progress in the analysis of the trends that determine when a mutation is neutral or pathological. Improvements in our PMUT server (http://mmb.pcb.ub.es/PMUT) have been made (see Figure 3). This server is now one of the most used for prediction of the pathological nature of SNPs. Furthermore, a comparative analysis on neutral, pathological and correlated mutations (Ferrer-Costa *et al*, 2006) has shown that each individual mutation of the correlated pair is characterised by a low pathogenicity index. This observation indicates that individual mutations can persist for long periods, without having a large impact on the species, before being stabilised by a complementary mutation (Ferrer-Costa *et al*, 2006).

**Structural studies on unusual nucleic acids**

This field has traditionally been one of great activity in the group (for a review see Orozco *et al*, 2003). During 2006 we have focussed our efforts on both the methodological aspects of analysis of the molecular dynamics trajectories of nucleic acids (NA) (Noy *et*

*al*, 2006) and on the detailed study of several unusual conformations. In particular, following previous studies (Cubero *et al*, 2003), we have explored the possibility of short fragments of Hoogsteen anti-parallel DNA embedded into long B-DNA duplexes. Extended molecular dynamics simulations demonstrate that, quite surprisingly, not only is the parallel Hoogsteen a stable structure, but the H-B junction does not introduce dramatic distortions into the overall structure of the helix. We also focus our attention on quadruplex DNA (G-DNA), a minor conformation of DNA that contributes to telomere stabilisation and is a major target for telomerase inhibition. We have reported for the first time that, in the presence of monovalent ions, G-DNA is so stable that it can survive to full vaporisation (Rueda *et al*, 2006b). If fully confirmed experimentally (available experimental information supports our hypothesis), G-DNA will be the first biopolymer to show stability in gas phase conditions, which would open up unexpected biotechnological applications for DNA quadruplexes, especially in the field of conductor design.

Another long-standing focus of our research efforts is anti-gene therapy. Here we seek to target a duplex DNA by means of a triplex-forming oligonucleotide (TFO), which binds to the major groove of the duplex generating a triplex. Once the triplex is formed the gene is knocked out, especially when triplex formation occurs in the promoter region. Considerable effort by several groups in our lab focuses on designing TFOs that lead to more stable triplexes. However, very often the formation of the triplex is hampered by that of an alternative tetraplex (see Figure 4). Part of our work during 2006 has addressed the design of nucleobase derivatives with the capacity to prevent tetraplex formation at the TFO. Our simulations indicate that excellent results can be obtained with 8-aminoguanine (Cubero *et al*, 2006). This result has been confirmed in agreement with experimental data.


***Figure 3***. *PMUT(2006) Web server (http://mmb.pcb.ub.es/PMUT).*

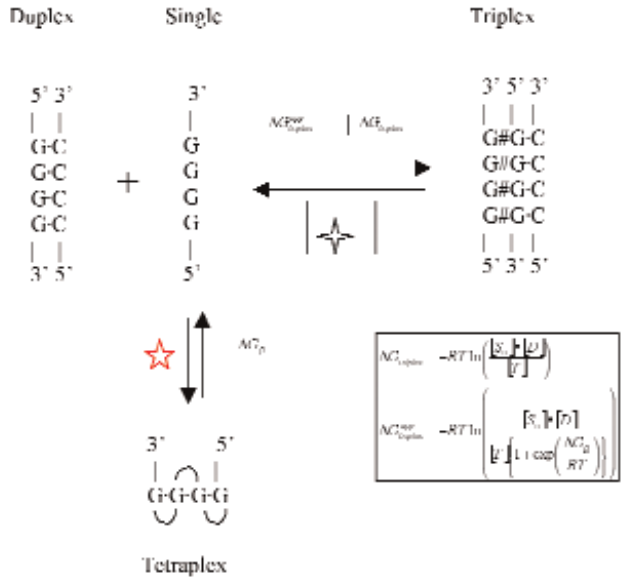



***Figure 4***. *Scheme of the inhibition of triplex formation by tetraplexes.*

**PUBLICATIONS**

Alonso D, Dorronsoro I, Rubio L, Muñoz P, García-Palomero E, del Monte M, Bidon-Chanal A, Orozco M, Luque FJ, Castro A, Medina M and Martínez A (2005) Donepezil-tacrine hybrid related derivatives as new dual binding site inhibitors of AchE. Bioorg Med Chem, 13:6588-6597

Bidon-Chanal A, Martí MA, Crepo A, Milani M, Orozco M, Bolognesi M, Luque FJ and Estrin DA (2006) Ligand-induced dynamical regulation of NO conversion in *Mycobacterium tuberculosis* truncated hemoglobin-N. Proteins, 64:457-465

Blas JR, López-Bes JM, Márquez M, Sessler JL, Luque FJ and Orozco M (2007) Exploring the dynamics of Calix[4]pyrrole: Effect of solvent and fluorine substitution. Chemistry, 13:1108-1116

Conde L, Vaquerizas JM, Ferrer-Costa C, de la Cruz X, Orozco M and Dopazo J (2005) PupasView: A visual tool for selecting suitable SNPs with putative pathological effects in genes for geotyping purporses. Nucleic Acids Res, 33:501-505

Crespo A, Martí MA, Kalko SG, Morreale A, Orozco M, Gelpí JL, Luque FJ and Estrín D (2005) Theoretical study of the truncated hemoglobin HbN: Exploring the molecular basis of the NO detoxification mechanism. J Am Chem Soc, 127:4433-4444

Cubero E, Luque FJ and Orozco M (2006) Theoretical study of the Hoogsteen<->Watson-Crick junctions in DNA. Biophysical J, 90:1000-1008

Curutchet C, Bidon-Chanal A, Soteras I, Orozco M and Luque FJ (2005) MST continuum study of the hydration free energies of monovalent ionic species. J Phys Chem B, 109:3565-3574

Curutchet C, Orozco M, Luque FJ, Mennucci B and Tomasi J (2006) Dispersion and repulsion contributions to the solvation free energy: Comparison of quantum mechanical and classical approaches in the polarisable continuum model. J Comp Chem, 27:1769-1780

Ferrer-Costa C, Gelpí JL, Zamacola L, Párraga I, de la Cruz X and Orozco M (2005) PMUT: A Web-based tool for the annotation of pathological mutations on proteins. Bioinformatics, 21:3176-3178

Ferrer-Costa C, Orozco M and de la Cruz X (2005) How can non-human proteins accommodate as wild-type human disease-associated residues. Proteins, 61:878-887

Ferrer-Costa C, Orozco M and de la Cruz X (2007) Characterisation of compensates mutations in terms of structural and physico-chemical properties. J Mol Biol, 365:249-256

Fuentes-Cabrera M, Lipkowski P, Huertas O, Sumpter BG, Orozco M, Luque FJ, Wells JC and Leszczynski J (2006) Aromaticity induced changes in the electronic properties of size-expanded DNA bases. The case of xC. Int J Quantum Chem, 106:2339-2346

Goñi JR, Vaquerizas JM, Dopazo J and Orozco M (2006) Exploring the reasons for the large density of triplex-forming oligonucleotide target sequences in the human genome. BMC Genomics, 7:63-72

Huertas O, Blas JR, Soteras I, Orozco M and Luque FJ (2006) Benzoderivatives of nucleic acid bases as modified DNA building blocks. J Phys Chem A, 110:510-518

Huertas O, Poater J, Fuentes-Cabrera M, Orozco M, Sola M and Luque FJ (2006) Local aromaticity in natural nucleobases and their size-expanded benzo-fused derivatives. J Phys Chem B, 110:12249-12258

López de la Osa J, González C, Gargallo R, Rueda M, Cubero E, Orozco M, Aviñó A and Eritja R (2006) Destabilisation of quadruplex DNA by 8-aminoguanine. ChemBioChem, 7:46-48

Meyer T, Ferrer-Costa C, Pérez A, Rueda M, Bidon-Chanal A, Luque FJ, Laughton CA and Orozco M (2006) Essential dynamics: A tool for efficient trajectory compression and management. J Chem Theory Comput, 2:251-258

Morreale A, de la Cruz X, Meyer T, Gelpí JL, Luque FJ and Orozco M (2005) Partition of protein solvation into group contributions from molecular dynamics simulations. Proteins, 58:101-119

Muñoz-Muriedas J, Barril X, López JM, Orozco M and Luque FJ (2006) A hydrophobic similarity analysis of solvation effects on nucleic acids, J Mol Model, Sept 21 [Epub ahead of print]

Muñoz-Muriedas J, Perspicace S, Bech N, Guccione S, Orozco M and Luque FJ (2005) Molecular alignment based on hydrophobic similarity as a tool in drug design. J Comput Aided Mol Des, 19:401-419

Muñoz-Ruiz P, Rubio L, García-Palomero E, Dorronsoro I, Monte-Millán M, Valenzuela R, Usán P, Austria C, Bartolini M, Andrisano V, Bidon-Chanal A, Orozco M, Luque FJ, Medina M and Martínez A (2005) Design, synthesis and pharmacological evaluation of dual binding site acetylcholinesterase inhibitors: New disease modifying agents for Alzheimer's disease. J Med Chem, 48:7223-7233

Noy A, Meyer T, Rueda M, Ferrer C, Valencia C, Pérez A, De la Cruz X, López-Bes JM, Luque FJ and Orozco M (2006) Datamining of molecular dynamics trajectories of nucleic acids. J Biomol Struct Dynam, 23:447-455

Noy A, Pérez A, Márquez M, Luque FJ and Orozco M (2005) Structure, recognition properties and flexibility of the DNA.RNA hybrid. J Am Chem Soc, 127:4910-4920

Pérez A, Blas JR, Rueda M, López-Bes JM, de la Cruz X and Orozco M (2005) Exploring the essential dynamics of B-DNA. J Chem Theory Comput, 1:790-800

Pérez A, Sponer J, Jurecka P, Hobza P, Luque FJ and Orozco M (2005) Are the RNA(A.U) hydrogen bonds stronger than the DNA(A.T) ones? Chemistry, 11:5062-5066

Rueda M, Ferrer C, Meyer T, Pérez A, Camps J, Hospital A, Gelpí JL and Orozco M (2007) A consensus view of protein dynamics. Proc Natl Acad Sci USA, 104:796-801

Rueda M, Luque FJ and Orozco M (2005) The nature of minor-groove binders-DNA complexes in the gas phase. J Am Chem Soc, 127:11690-11698

Rueda M, Luque FJ and Orozco M (2006) G-DNA can maintain its structure in the gas phase. J Am Chem Soc, 128:3608-3619

Soliva R, Gelpi JL, Almansa C, Virgili M and Orozco M (2007) Dissection of the recognition properties of p38 MAP kinase. Determination of the binding mode of a new pyridinyl-heterocycle inhibitor family. J Med Chem, 50:283-293

Soteras I, Curutchet C, Bidon-Chanal A, Orozco M and Luque FJ (2005) Extension of the MST model to the IEF formalism: HF and B3LYP parametrisations. J Mol Struct Theochem, 727:29-40

Soteras I, Lozano O, Gómez-Esqué A, Escolano C, Orozco M, Amat M, Bosch J and Luque FJ (2006) On the origin of stereoselectivity in the alkylation of oxazolopiperidone enolates. J Am Chem Soc,128:6581-6588

Sponer J, Jurecka P, Marchan I, Luque FJ, Orozco M and Hobza P (2006) Nature of base stacking. Reference quantum chemical stacking energies in ten unique B-DNA base pairs steps. Chemistry, 12:2854-2865

Talavera D, Morreale A, Hospital A, Ferrer C, Gelpi JL, de la Cruz X, Meyer T, Soliva R, Luque FJ and Orozco M (2006) A fast method for the determination of fractional contributions to solvation in proteins. Protein Sci, 15:2525-2533

**OTHER REFERENCES**

Cubero E, Abrescia N, Subirana JA, Luque FJ and Orozco M (2003) Theoretical study of a new structure of DNA: The antiparallel Hoogsteen duplex. J Am Chem Soc, 125:14603-14612

Orozco M and Luque FJ (2000) Theoretical methods for the representation of solvent in biomolecular systems. Chem Rev, 100:4187-4225

Orozco M, Pérez A, Noy A and Luque FJ (2003) Theoretical methods for the simulation of nucleic acids. Chem Soc Rev, 32:350-364

*Modesto Orozco's group, March 2006.*