Molecular modelling and bioinformatics group

Our long-term objective is to explain 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 general aim, we work with several methodologies, from the mining of biological databases to classical dynamics and quantum chemistry calculations. The use of such diverse techniques allows us to explore a wide range of problems, from drug design to genome analysis. Special emphasis is placed on connecting basic interactions with the global properties of biological systems. In general terms, our work centres on the following 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.

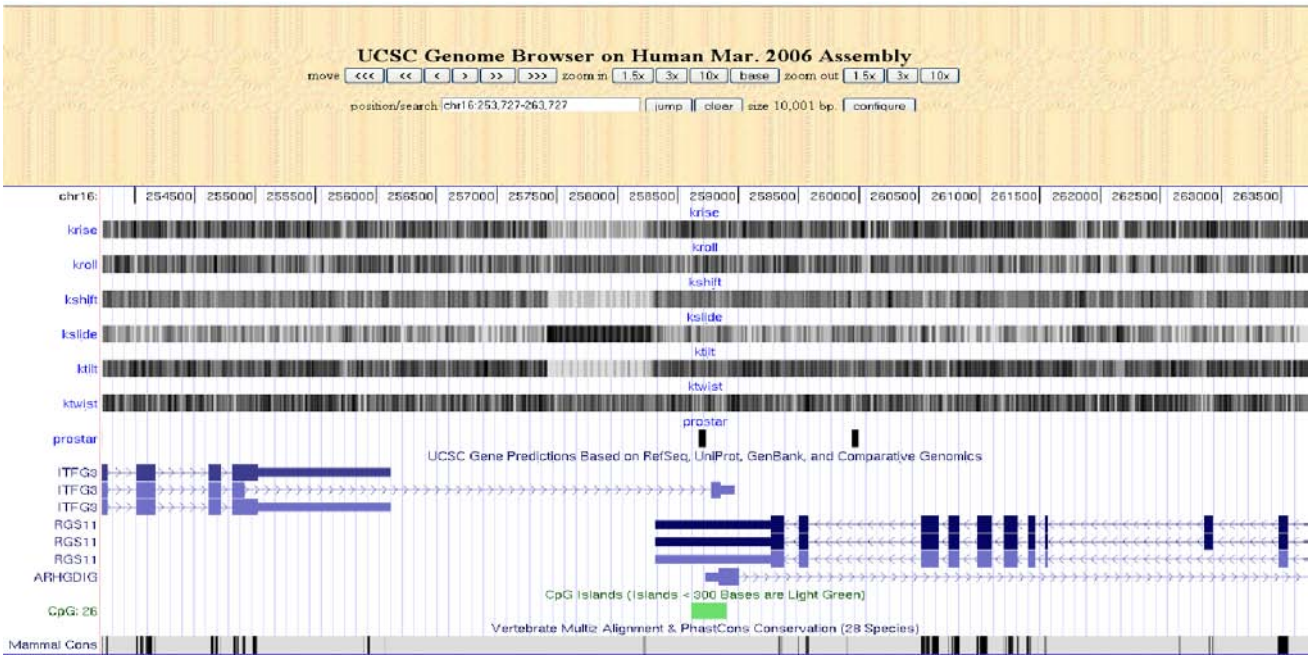

Figure 1. *Example of a classical 2D output of DNAAlive. Top figures correspond to physical properties derived from quasi-harmonic mesoscopic simulations of DNA (parameters obtained from MD simulations). Figures at the bottom correspond to biological annotations on the same genome fragment. Marked in red is a region of unusual physical properties, which is later found to correspond to a transcription start site.*

Principal Investigator Modesto Orozco **Research Associates** Xavier de la Cruz, Agustí Emperador, Josep Lluís Gelpí **Postdoctoral Fellows** Montse Barbany, Oliver Carrillo, Chiara Catellazzi, Marco d'Abramo, Rebeca Garcia, Agnes Noy, Alberto Pérez, Guillem Portella, Nadine Utz **PhD Students** Annalisa Arcella, Ozgen Deniz, Ignacio Faustino, Carles Fenollosa, Josep Ramon Goñi, Adam Hospital, Sergio Lois, Tim Meyer, Jordi Morata, Laura Orellana, Antonella Paladino, David Piedra, David Talavera **Research Assistants** José Antonio Alcántara, Jordi Camps, Chiara Castellazi, Damjan Cicin **Lab Technician** Mª Luz Ruiz **Administrative Assistants** Margarita Pedro, Cristina Villanueva **Visiting Scientists** Andrea Amadei (Italy), Sarah Harris (United Kingdom) **Visiting Students** Ariadna Giró (Spain), Venu Kurella (India), Eva Novoa (Spain), Francesco Raimondi (Italy)

Small model systems

Our group has a long trajectory in the study of small model systems of biological relevance (nucleobase complexes, drugs, isolated complexes of amino acids, stacked or hydrogen bonded complexes, etc.). The study of these simple systems can shed light on the behaviour of much more complex biological molecules. Almost a decade ago, we realised that such studies were simple in the gas phase but very difficult in aqueous solution, thereby hampering the true application of the information obtained to the biological scenario. This led us to develop methods to describe solvent systems, some of which are considered the current state of the art in the field, and to develop approaches for accurate representation of molecular interactions.

As planned, during 2008 we have advanced in the development of methods for the efficient representation of polarisation effects (Soteras *et al*, 2008a; Soteras *et al*, 2008b). We have also considerably improved our MST continuum model, which has now been adapted to provide complete thermodynamics of solvation with high accuracy (Bidon-Chanal *et al*, 2008; Klamtz *et al*, 2008).

In collaboration with colleagues at Minnesota (Xie *et al*, 2008), we have developed a novel method to perform full quantum mechanical calculations in proteins. The method is based on the partition of the entire system in inner (treated as QM clusters) and external (treated classically) fragments. We are now parallelising the code to allow extension of these quantum mechanical/molecular dynamics (QM-MD) calculations to the nanosecond time scale. The potential use of all these new methodologies in the context of drug development and nucleobase design have been explored in detail (Cozzini *et al*, 2008; Vázquez-Mayogoita, 2008) during 2008.

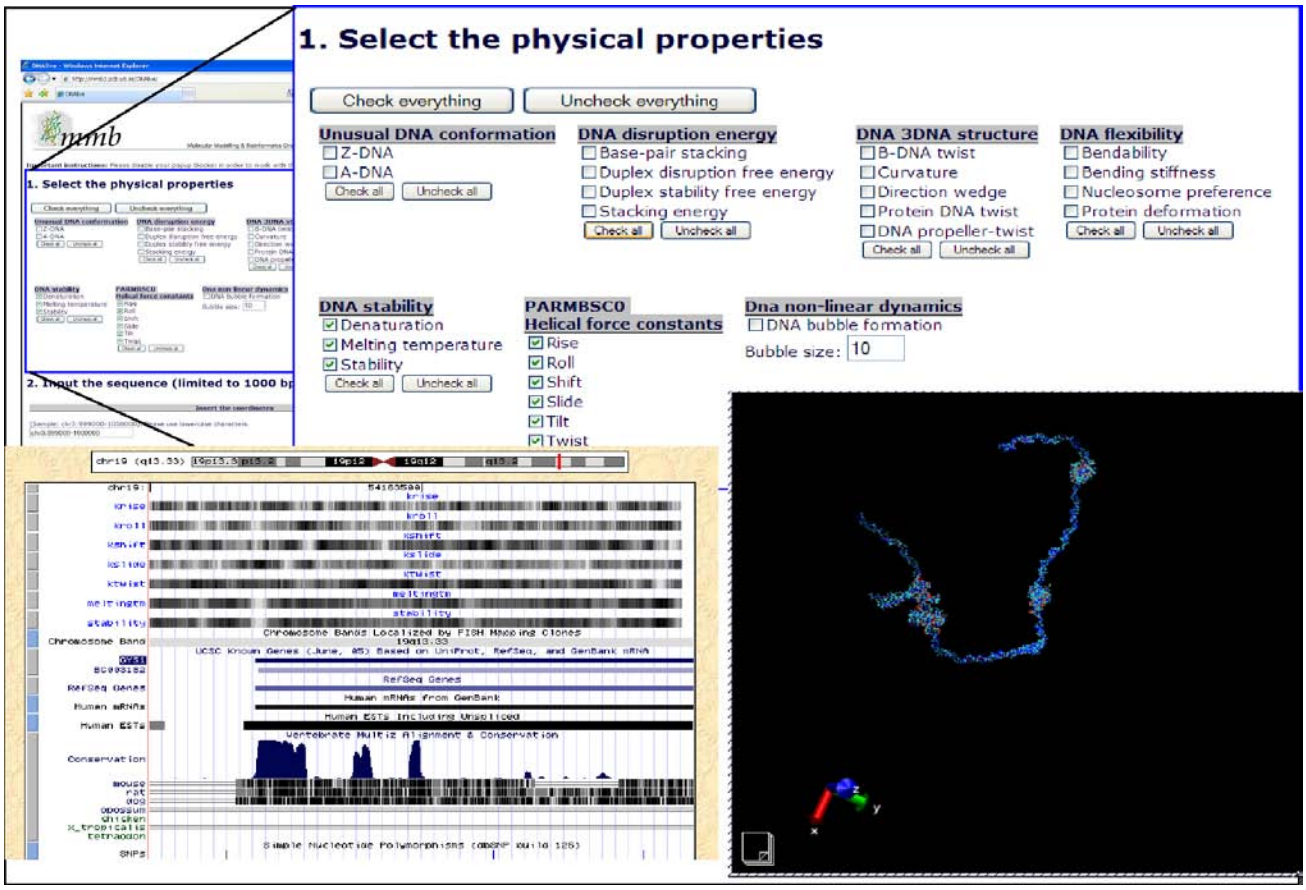

Figure 2. *Example of the possibilities of DNA in its physical description. Bottom left panel corresponds to mesoscopic simulations of the structure and dynamics of the chromatin fiber, including proteins.*

Analysis of stressed or unusual nucleic acids

Major breakthroughs in the field of nucleic acid simulations emerged from the work performed by our group in 2007 related to the development of PARMBSC0 force-field, which, after additional tests (Svozil *et al*, 2008), has become the default force-field for nucleic acid simulations (Orozco *et al*, 2008). Thanks to this refined force-field and to the massive computer power of MareNostrum, we have deciphered the mechanism by which RNAase H differentiates between hybrids to be degraded and homopolymers that must be resistant to the enzyme (Noy *et al*, 2008). Using the same force-field, we have successfully designed modified nucleobases that display a unique effect on G-DNAs, stabilising or destabilising them depending on the position where they are inserted (Gross *et al*, 2008).

Considerable effort has been focused on the characterisation of the equilibrium sequence-dependent stiffness of duplex DNA through PARMBSC0-powered molecular dynamics simulations (Pérez *et al*, 2008). These studies have allowed us to derive a full mesoscopic model of the flexibility of DNA, which can be used to study chromatin deformability (Perez *et al*, 2008, Goñi *et al*, 2008). In this context, the development of DNAlive (Goñi *et al*, 2008) has been a major landmark since it has allowed us for the first time to obtain an integrated view of chromatin by including physical information and biological annotations in a single tool. DNAlive is the perfect link between DNA simulation and genomics (Figures 1-2). We are currently using DNAlive intensively to study several aspects of chromatin structure and dynamics and during 2009 we plan to extend it to account for epigenetic changes.

Protein dynamics

The creation of the MODEL (Molecular Dynamics Extended Library) database has been the focus of a massive amount of work by the group. During 2008 we have finished the first draft of the database and started to work on the derivation of complementary databases, such as that of proteins of pharmacological interest, kinases and protein complexes. We have also almost completed the analysis of the trajectories to decipher the patterns of protein flexibility and their association with solvent structure (Figure 3).

Using our μMODEL reduced set, we have examined the impact of environmental changes on protein structure. Multi-microsecond-long simulations on urea have been performed during 2008 (they will finish in 2009) to illustrate the atomistic mechanism of chemo-thermal unfolding. This year we have completed the mapping of the gas phase proteome (Meyer *et al*, 2008), again using the reduced μMODEL database. This study has provided surprising evidence of the stability of proteins under electrospray-like conditions, thereby opening up the possibility to use gas phase structural information derived from x-free electron laser microscopy to obtain the protein structure in solution (Figure 4).

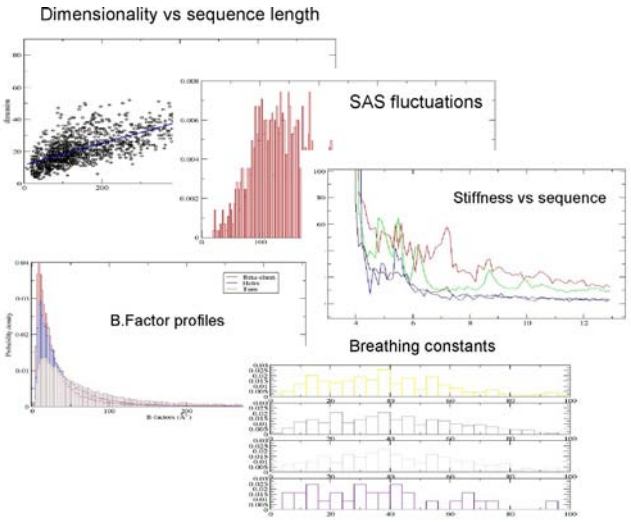

Figure 3. *Examples of dynamic analysis of protein flexibility derived from our MODEL database.*

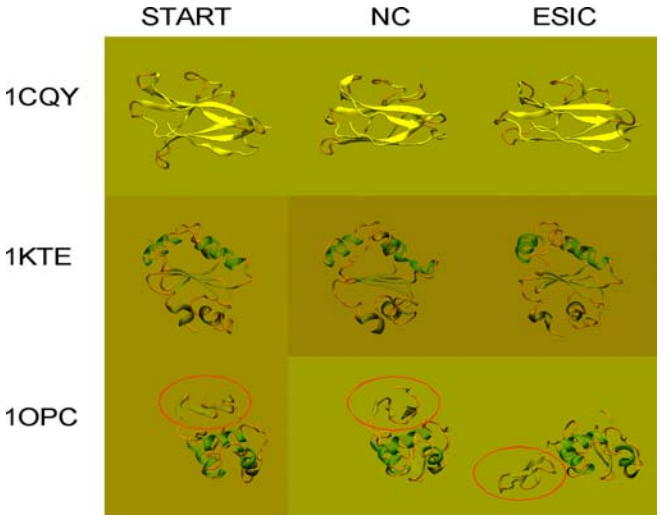

Figure 4. *Structures of 3 ultra-representative proteins in solution (start) and upon ideal (NC) and normal (ESIC) electrospray vaporisation.*

The availability of several terabytes of trajectories in μMODEL has opened up many possibilities for proteome-scale analysis, including systematic exploration of essential dynamic movements in proteins and their connection with function and evolution (Velázquez-Muriel *et al*, 2008). It has also provided the opportunity to study protein channels following the MDGRID methodology developed by our group (Carrillo *et al*, 2008) and to perform genome-scale ensemble docking experiments to identify new drugs.

MD trajectories deposited in μMODEL have helped us to derive new coarse-grained models designed to obtain a fast first-order approximation to protein flexibility using very simple potential functions. During 2008 we have developed and optimised pseudo-harmonic methods based on normal model analysis and Brownian MD (Emperador, 2008a). We have explored the use of ballistic equations for recovering dynamics trajectories using either residue or atom-resolution representations of proteins (Emperador, 2008b). Given their speed, these methods are expected to allow a full-proteome representation of protein flexibility, even in protein mixtures. We are currently finishing the development of an entire platform (named FlexServ) that will automatise these types of studies.

Finally, we should remark that dynamic information, obtained from either MD, coarse-grained methods or even evolutionary methods (Velázquez-Muriel *et al*, 2008) can help to improve the structural representation of proteins obtained for ensemble techniques such as NMR, as demonstrated in the COCO web server developed by our group, in collaboration with colleagues in Nottingham and at the European Bioinformatics Institute (Laughton *et al*, 2008).

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**(1-2), 11-20 (2008)

BioMoby Consortium, Wilkinson MD, Senger M, Kawas E, Bruskiewich R, Gouzy J, Noirot C, Bardou P, Ng A, Haase D, Saiz Ede A, Wang D, Gibbons F, Gordon PM, Sensen CW, Carrasco JM, Fernández JM, Shen L, Links M, Ng M, Opushneva N, Neerincx PB, Leunissen JA, Ernst R, Twigger S, Usadel B, Good B, Wong Y, Stein L, Crosby W, Karlsson J, Royo R, Párraga I, Ramírez S, Gelpi JL, Trelles O, Pisano DG, Jimenez N, Kerhornou A, Rosset R, Zamacola L, Tarraga J, Huerta-Cepas J, Carazo JM, Dopazo J, Guigo R, Navarro A, Orozco M, Valencia A, Claros MG, Pérez AJ, Aldana J, Rojano MM, Fernandez-Santa Cruz R, Navas I, Schiltz G, Farmer A, Gessler D, Schoof H and Groscurth A. Interoperability with Moby 1.0--it's better than sharing your toothbrush! *Brief Bioinform*, **9**(3), 220-31 (2008)

Carrillo O and Orozco M. GRID-MD-A tool for massive simulation of protein channels. *Proteins*, **70**(3), 892-99 (2008)

Cozzini P, Kellogg GE, Spyrakis F, Abraham DJ, Costantino G, Emerson A, Fanelli F, Gohlke H, Kuhn LA, Morris GM, Orozco M, Pertinhez TA, Rizzi M and Sotriffer CA. Target flexibility: an emerging consideration in drug discovery and design. *J Med Chem*, **51**(20), 6237-55 (2008)

Emperador A, Carrillo O, Rueda M and Orozco M. Exploring the suitability of coarse-grained techniques for the representation of protein dynamics. *Biophys J*, **95**(5), 2127-38 (2008a)

Emperador A, Meyer T and Orozco M. United-atom discrete molecular dynamics of proteins using physics-based potentials. *J Chem Theor Comput*, **4**(12), 2001-10 (2008b)

Goñi JR, Fenollosa C, Pérez A, Torrents D and Orozco M. DNAlive: A tool for the physical analysis of DNA at the genomic scale. *Bioinformatics*, **24**(15), 1731-32 (2008)

Gros J, Aviñó A, Lopez de la Osa J, González C, Lacroix L, Pérez A, Orozco M, Eritja R and Mergny JL. 8-Amino guanine accelerates tetramolecular G-quadruplex formation. *Chem Commun (Camb)*, **25**, 2926-28 (2008)

Laughton CA, Orozco M and Vranken W. COCO: a simple tool to enrich the representation of conformational variability in NMR structures. *Proteins*, Epub ahead of print (2008)

Noy A, Luque FJ and Orozco M. Theoretical analysis of antisense duplexes: determinants of the RNase H susceptibility. *J Am Chem Soc*, **130**(11), 3486-96 (2008)

Orozco M, Noy A and Pérez A. Recent advances in the study of nucleic acid flexibility by molecular dynamics. *Curr Op Struct Biol*, **18**(2), 185-93 (2008). Front cover paper

Pérez A, Lankas F, Luque FJ and Orozco M. Towards a molecular dynamics consensus view of B-DNA flexibility. *Nucleic Acids Res*, **36**(7), 2379-94 (2008)

Soteras I, Lozano O, Escolano C, Orozco M, Amat M, Bosch J and Luque FJ. Structure-directed reversion in the pi-facial stereoselective alkylation of chiral bicyclic lactams. *J Org Chem*, **73**(19), 7756-63 (2008)

Soteras I, Orozco M and Luque FJ. Induction effects in metal cation-benzene complexes. *Phys Chem Chem Phys*, **10**(19), 2616-24 (2008)

Svozil D, Sponer JE, Marchan I, Pérez A, Cheatham TE 3rd, Forti F, Luque FJ, Orozco M and Sponer J. Geometrical and electronic structure variability of the sugar-phosphate backbone in nucleic acids. *J Phys Chem B*, **112**(27), 8188-97 (2008)

Vazquez-Mayagoita A, Huertas O, Fuentes-Cabrera M, Sumpter BG, Orozco M and Luque FJ. *Ab initio* study of naphtho-homologated DNA bases. *J Phys Chem B*, **112**(7), 2179-86 (2008)

Research networks and grants

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

European Life Science Infrastructure for Biological Information (ELIXIR)
European Commission, 211601 (2007-2010)
Principal investigator: Modesto Orozco

Molecular recognition
Marcelino Botin Foundation
Principal investigator: Modesto Orozco

Proyecto CONSOLIDER Supercomputación y eCiencia
Spanish Ministry of Science and Innovation, CSD2007-00050 (2007-2012)
Principal investigator: Modesto Orozco

Reconeixement molecular
Generalitat de Catalunya, 2005-SGR0286 (2006-2009)
Principal investigator: Modesto Orozco

Red temática de investigación cooperativa en biomedicina computacional (COMBIOMED)
Instituto de Salud Carlos III, COMBIOMED RD07/0067/0009 (2008-2012)
Principal investigator: Modesto Orozco

Structural bioinformatics
Genoma España, GN4 (2007-2009)
Principal investigator: Modesto Orozco

Other funding sources

Collaboration contracts with Salvat Laboratories, Neuropharma and Palau Pharma-Grupo Uriach

Collaborations

Design of nucleobase derivatives for stabilization of anomalous DNAs
Ramon Eritja, IRB Barcelona-CSIC (Barcelona, Spain)

Development of new tools for computer assisted drug design
Francisco Javier Luque, Faculty of Pharmacy, University of Barcelona (Barcelona, Spain)

Development of strategies for improvement of NMR-samplings
Charles A Laughton, Nottingham University (Nottingham, UK), Williams Vranken, European Bioinformatics Institute, EBI-EMBL (Cambridge, UK)

Force-field refinement for nucleic acid simulations
Thomas Cheatham, University of Utah (Salt Lake City, USA), Jiri Sponer (Brno University (Brno, Czech Republic) and Filip Lankas, Laussanne University (Lausanne, Switzerland)

Introduction of polarisation effects in force-field calculations
Cristophe Chipot, Nancy University (Nancy, France) and Francisco Javier Luque, University of Barcelona (Barcelona, Spain)

Mixed QM-MM methods for protein simulations
Donald Truhlar and J Gao, University of Minnesota (Minnesota, USA)

New algorithms for drug design
Paolo Cozzini, Peruggia University (Peruggia, Italy) and Robert Soliva, PalauPharma-Uriach Group (Barcelona, Spain)

Physical properties of DNA
David Torrents, Barcelona Supercomputing Center (Barcelona, Spain)

SCRF solvation methods
Jacopo Tomasi, Pisa University (Pisa, Italy), Andreas Klamtz, Cosmologic Inc. (Leverkusen, Germany) and Francisco Javier Luque, University of Barcelona (Barcelona, Spain)

Study of DNA-metal complexes
Miguel Fuentes-Cabrera, Oak Ridge National Laboratory (Oak Ridge, USA) and Francisco Javier Luque, University of Barcelona (Barcelona, Spain)

Awards and honours

Distinguished fellow of the Marcelino Botin Foundation
Awardee: Modesto Orozco