# Modeling macro-biomolecules

Valentina Tozzini

v.tozzini@sns.it

Collaborators
P. Amat
S. Bonella
A. Di Fenza
G.N. Embrapa
J.A. McCammon
R. Nifosì
W. Rocchia
F. Trovato
J. Trylska
K. Voltz

*Biological matter is characterized by an overwhelming complexity: In order to "live" an organism must undergo a variety of events that occur on different scales of length and time. These range from a few Å, the size of the active site of proteins, where the ultra-fast triggering steps of the biochemical reactions take place, up to the level of the cells and organs, where the macroscopic physiological effects are detectable by the naked eye, and involve the nano- and micro- scale as intermediates. This hierarchical organization is responsible for the complexity, because each single process inherently involves a cascade of events occurring on different scales.*
*In the period 2005-2006 we developed and tested modeling methodologies tailored to address the nano-micro scale typical of the macromolecular aggregates, namely the Coarse Grained models. In the subsequent period (2007-2009), we extended those models and applied them to specific problems (fundamental steps of HIV replication and in general of DNA replication, and the Fluorescent Proteins structural and optical properties). In addition, we started a process of integration of the CG models within a more general multi-scale approach, which, we believe, is essential for a coherent description and a deep understanding of any biological processes.*

The HIV protease (HIV pr) is a virus-specific enzyme whose action is essential for the virus maturation: it cleaves the Gag viral poly-protein in functional parts that will be structural and functional elements of the new viral particles. In 2005-2006 we developed a CG model for HIV pr based on the one-interacting-center-per-amino-acid description (one-bead (OB) model[1]). The model was shown to accurately describe the protein dynamics on time scales large enough to address the slow (μsec-msec) process of the active site opening[2], that is the first step of the action mechanism of this enzyme. In 2007-2008, we applied the model to the study of the whole cleavage process: the model was able to clarify some details of the binding with a model substrate (a part of the Gag poly-peptide) and of the product release after cleavage[3] (Fig. 1 (a)). A similar simulation study with the whole Gag poly-protein is currently in progress. These studies are relevant to indicate novel therapeutic strategies exploiting the HIV pr inhibition. Additionally we studied the interactions of HIV pr with currently used inhibitor drugs[4] (Fig. 1 (b)). In this case a multi-scale approach involving also all-atom simulations was adopted in order to address the more chemically complex interactions occurring with non-peptide ligands.

The same methodology was also applied to the study of HIV-Integrase – an enzyme responsible for the viral DNA integration into the cell nucleus. Since the acetylation of some of its specific Lysine residues is involved in the process, using molecular recognition and docking methodologies, we built models of Integrase complexed with the two acetyl transferases (HATs) p300 and GCN5 (Fig 2). We showed that the obtained models are compatible with experimental observations and possess stereochemical quality. On the basis of the models we hypothesize that the residue which is more likely to be acetylated by both HATs is Lys 273. Finally, we applied an Essential Dynamics analysis to both models which outlined two critical regions of Integrase which can provide possible allosteric sites on the enzyme for the inhibition of the viral genome integration in the host cell [5].

The one-bead model was then extended to nucleic acids, for which a one-bead-per-nucleotide model was developed. In particular our DNA model is able to accurately predict the dynamics of supercoiling and local denaturation of torsionally stressed DNA nanorings

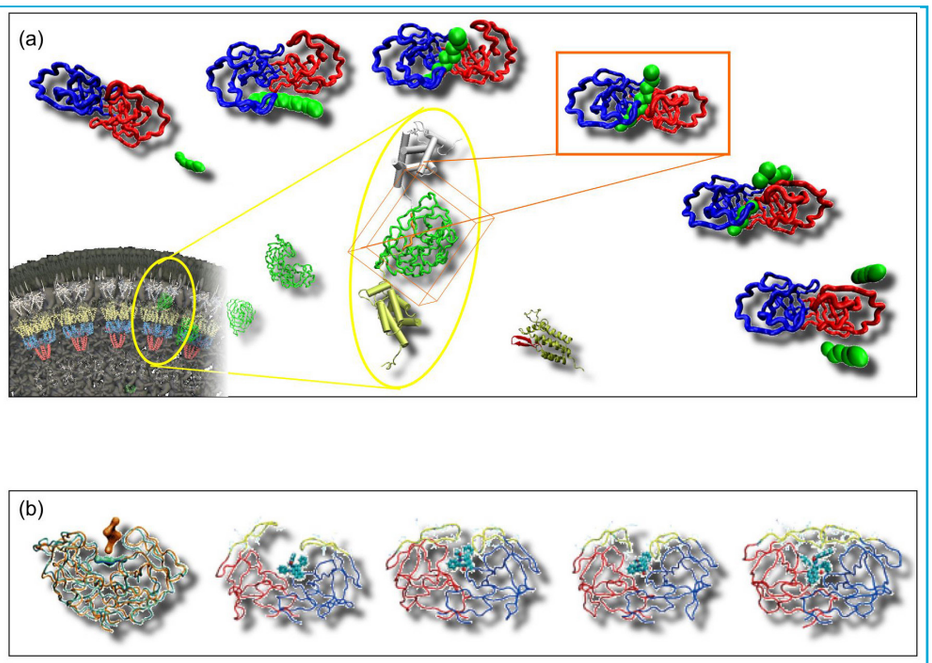


Fig. 1
Modeling the HIV pr action mechanism. (a) In red-blue-green: subsequent snapshots taken from the simulation of the model-substrate (in green) capture, cleavage and product release by the HIV pr (red-blue). The orange rectangle highlights the snapshot where the substrate HIV-1pr complex is formed. The same portion of the HIV-1pr (green) complexed with the Gag (white-yellow) is enclosed in the 3D orange parallelepiped. The cleaved yellow part undergoes a structural transformation before becoming a structural part of the capsid. The complex within the envelope of the immature virion is also pictorically represented in the left lower corner. (b) The multiscale simulation of the binding of an antiviral drug to HIV pr. In the first snapshot two steps of the approaching phase of the drug to the active site from a CG simulation are shown. The following four snapshots are taken from the all-atom MD simulation of the capture phase.

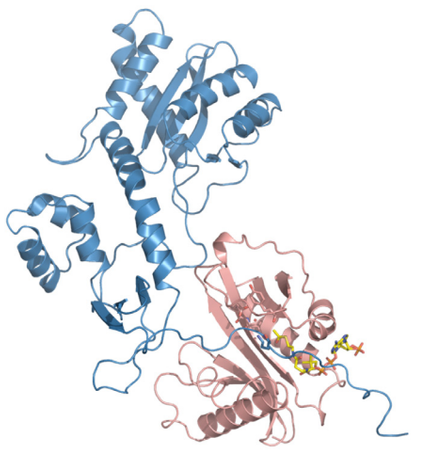

Fig 2
Cartoon view of the complex between HIV-1 Integrase (blue) and the acetylTransferase GCN5 (pink). The residue which undergoes acetylation (Lys 273), the co-factor molecule (AcetylCoA, yellow) are showed as sticks together with the residues of the catalytic site.

and of entire small plasmids [6] (Fig 3). This study indicates that the torsional stress physiologically present in DNA has an important role in the regulation of the interplay between compaction-decompaction and duplication of DNA, that requires the strand separation. A simplified version of the model was integrated with the one-bead models for proteins[7] and used within a multi-scale approach [8] to study the wrapping-unwrapping of DNA around the core proteins in the nucleosomes [9] (Fig 4).

The surface representation of bio-molecules, consisting in defining a bio-molecule by means of its solvent accessible surface, can be placed at the "very coarse" extreme in the multi-scale approaches. Surface representations are widely used in the docking and molecular recognition methodologies. Having in mind the idea of building an integrated framework for multi-scale approaches, the accurate and efficient representation of the electrostatics has an important role. In order to make some steps forward in this field, the

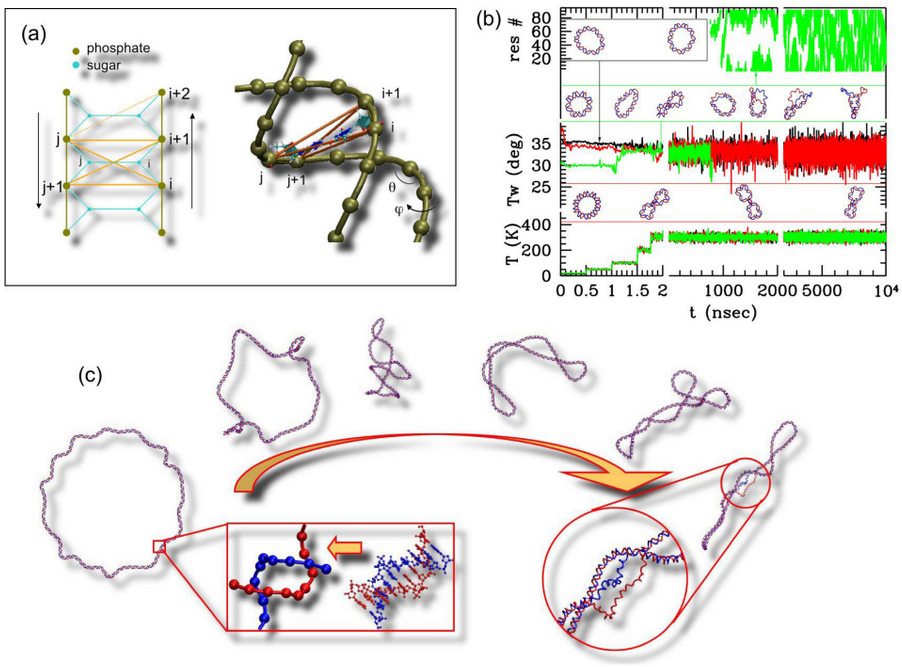


Fig. 3
(a) Illustration of the one-bead model for DNA. Topological and geometric representation. (b) Molecular dynamics simulations of DNA nanorings (90bp). Red, green and black lines (and rectangles) correspond to overtwisted, undertwisted and untwisted starting conformations. Temperature and average twist is shown during simulation. In the undertwisted case a denaturation bubble opens, which dynamically move around the circle. (c) A similar simulation for a plasmid (~900 bp) with the "physiological" value of torsional stress (weakly undertwisted). Starting from a nearly circular conformation the plasmid first super-coils, then a denaturation bubble opens.

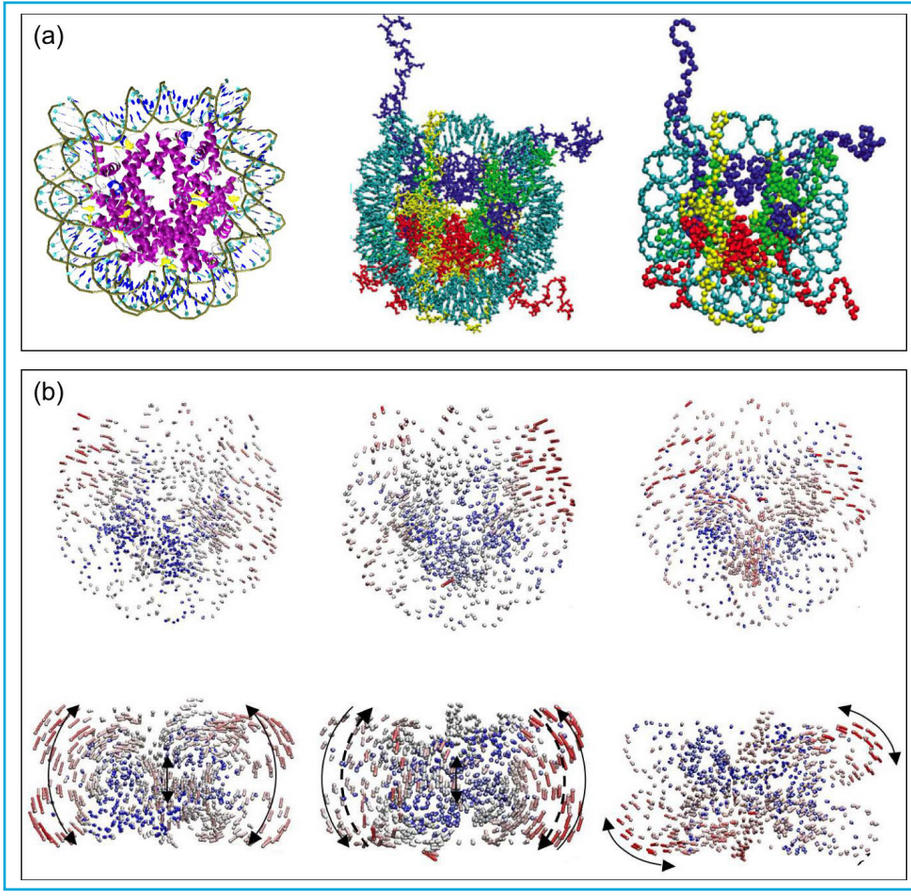


Fig 4
Multi-scale modeling of the nucleosome. (a) The nucleosome structure in cartoon, all-atom and CG representation. In the center and right structure the DNA is in cyan and the core proteins are in blue red and yellow. (b) The first three principal modes of the nucleosome (front and top views). The atom motion is represented by colored segments. The black arrows illustrate the motion of the structural units. These slow motions are related to the first steps of the DNA unwrapping.

electrostatic potential generated by all the existing PDB structures was calculated and stored in a relational DataBase in order to identify new patterns for predicting interesting structural properties related to molecular recognition[10].

An extensive description of the coarse grained and atomic based approaches for proteins and of their integration with atomic-based models in multi-scale approaches was proposed in a review paper[11], that also illustrates these methods on two very important example-systems, namely the green fluorescent proteins (GFP) and the proteins involved in the replication of HIV. The foremost relevance of these two biomolecular systems was confirmed by the assignment of two of the Nobel prizes in 2008: in chemistry for the discovery of GFP and in medicine for the discovery of HIV.

An extensive description of the coarse grained and atomic based approaches for proteins and of their integration with atomic-based models in multi-scale approaches was proposed in a review paper[11], which also illustrates these methods on two very important example-systems, namely the proteins involved in the replication of HIV and the green fluorescent proteins (GFP). The foremost relevance of these two biomolecular systems was confirmed by the assignment of two of the Nobel prizes in 2008: in medicine for the discovery of HIV and in chemistry for the discovery of GFP. Further molecular modeling activity on the latter system include a study of the oligomerization propensity of Fluorescent proteins using a novel statistical algorithm[12] and an investigation on photophysical properties for specific imaging applications reported in the section: "Fluorescent proteins for multiphoton and super-resolution imaging".

## References


[1] V. Tozzini and J.A. McCammon *One-Bead Coarse Grained Models for Proteins* in "Coarse-graining of condensed Phase and Biomolecular Systems", G.A. Voth Ed, Taylor & Francis/CRC press, Chapter 19 (2008)

[2] V. Tozzini, J. Trylska, C-E Chang, J.A. McCammon *Flap opening dynamics in HIV-1 protease explored with a coarse-grained model* J Stuct Biol, 157 606-615 (2007)

[3] J. Trylska, V. Tozzini, C.-E. Chang, J.A. McCammon *HIV-1 protease substrate binding and product release pathways explored with coarse-grained molecular dynamics* Biophys J, 92 4179-4187 (2007)

[4] C.-E. Chang, J. Trylska, V. Tozzini, J.A. McCammon *Binding pathways of ligands to HIV-1 Protease: Coarse-grained and atomistic simulations* Chem Biol Drug Des, 69 5-13 (2007)

[5] A. Di Fenza, W. Rocchia, V. Tozzini Complexes of HIV-1 integrase with HAT proteins: multi-scale models, dynamics and hypotheses on allosteric site of inhibition Proteins 76, 946-958 (2009)

[6] F. Trovato, V. Tozzini *Supercoiling and Local Denaturation of Plasmids with a Minimalist DNA Model* J Phys Chem B 112, 13197-13200 (2008)

[7] K. Voltz, T. Wocjan, K. Klenin, J. Trylska, V. Tozzini, V. Kurkal, J. Smith, J. Langowski *DNA dynamics on the nucleosome simulated by coarse-grained models* Biophys J, Suppl S 540A-540A (2007)

[8] K. Voltz, J. Trylska, V. Tozzini, V. Kurkal-Siebert, J. Langowski, J. Smith, *Coarse-grained force field for the nucleosome from self-consistent multiscaling* J Comput Chem 29, 1429-1439 (2008)

[9] K. Voltz, J. Trylska, V. Tozzini, V. Kurkal-Siebert, J. Smith, and J. Langowski *Global motions in the nucleosome explored using a coarse-grained model* NIC Series 40, 141-144 (2008)

[10] W. Rocchia, G. Neshich *Electrostatic potential calculation for biomolecules - creating a database of pre-calculated values reported on a per residue basis for all PDB protein structures* Genetics and Molecular Research 6, 923-936 (2007)

[11] V. Tozzini Multi-Scale Modeling of Proteins Acc Chem Res, DOI: 10.1021/ar9001476

[12] Bonella, W. Rocchia, P. Amat, R. Nifosí and V. Tozzini SPDhound, a mutual information based method to investigate specificity determining positions Algorithms 2, 764-789 (2009)