



# Predicting Nucleosome Positioning in Genomes: Physical and Bioinformatic Approaches

Anita Scipioni, Pasquale de Santis

## ► To cite this version:

Anita Scipioni, Pasquale de Santis. Predicting Nucleosome Positioning in Genomes: Physical and Bioinformatic Approaches. Biophysical Chemistry, 2011, 155 (2-3), pp.53. <10.1016/j.bpc.2011.03.006>. <hal-00743843>

**HAL Id: hal-00743843**

**<https://hal.science/hal-00743843v1>**

Submitted on 21 Oct 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Accepted Manuscript

Predicting Nucleosome Positioning in Genomes: Physical and Bioinformatic Approaches

Anita Scipioni, Pasquale De Santis

PII: S0301-4622(11)00063-9  
DOI: doi: [10.1016/j.bpc.2011.03.006](https://doi.org/10.1016/j.bpc.2011.03.006)  
Reference: BIOCHE 5490

To appear in: *Biophysical Chemistry*

Received date: 22 February 2011  
Revised date: 16 March 2011  
Accepted date: 16 March 2011



Please cite this article as: Anita Scipioni, Pasquale De Santis, Predicting Nucleosome Positioning in Genomes: Physical and Bioinformatic Approaches, *Biophysical Chemistry* (2011), doi: [10.1016/j.bpc.2011.03.006](https://doi.org/10.1016/j.bpc.2011.03.006)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## Predicting Nucleosome Positioning in Genomes: Physical and Bioinformatic Approaches.

Anita Scipioni, Pasquale De Santis\*

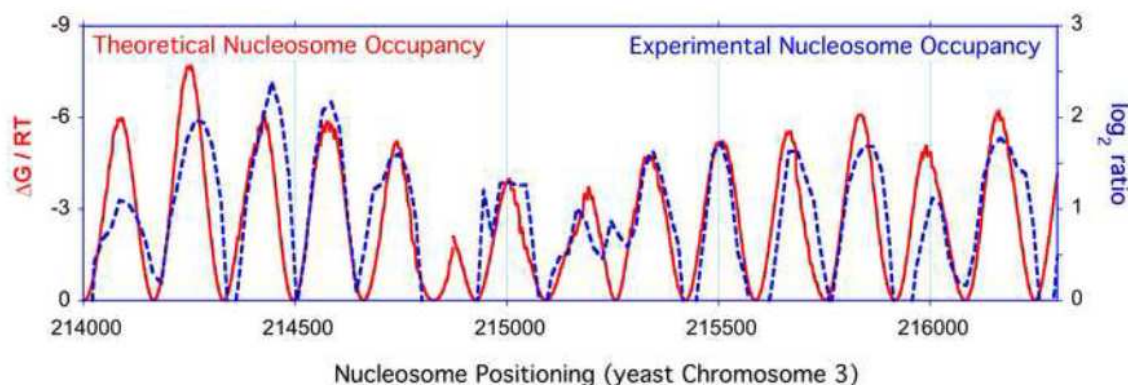
*Dipartimento di Chimica, Università di Roma La Sapienza, P.le A. Moro, 5 I-00185, Roma, Italy*

\* Corresponding author: Dipartimento di Chimica, Università di Roma La Sapienza, P.le A. Moro, 5 I-00185, Roma, Italy. Phone: +39.064453827, Fax: +39.064453827  
*E-mail address:* pasquale.desantis@uniroma1.it (P. De Santis)

## Abstract

In eukaryotic genomes, nucleosomes are responsible for packaging DNA and controlling gene expression. For this reason, an increasing interest is arising on computational methods capable of predicting the nucleosome positioning along genomes. In this review we describe and compare bioinformatic and physical approaches adopted to predict nucleosome occupancy along genomes. Computational analyses attempt at decoding the experimental nucleosome maps of genomes in terms of certain dinucleotide step periodicity observed along DNA. Such investigations show that highly significant information about the occurrence of a nucleosome along DNA is intrinsic in certain features of the sequence suggesting that DNA of eukaryotic genomes encodes nucleosome organization. Besides the bioinformatic approaches, physical models were proposed based on the sequence dependent conformational features of the DNA chain, which govern the free energy needed to transform recurrent DNA tracts along the genome into the nucleosomal shape.

## Graphical abstract



## Highlights

- > Theoretical models for predicting nucleosome positioning along genomes are reviewed.
- > Bioinformatics and physical approaches are compared and discussed.
- > Mechanical properties of DNA rule thermodynamic stability of nucleosomes.
- > Nucleosomes are regularly spaced along genomes as required for chromatin fibers.

## Keywords

Thermodynamic stability of nucleosomes; nucleosome code; nucleosome distribution along genomes; bioinformatic approaches for the prediction of nucleosome occupancy; physical models predicting nucleosome occupancy.

## 1. Introduction

Nucleosome positioning plays an essential role in cellular processes by regulating the accessibility of DNA to proteins and by providing the array of structural units for the DNA compaction in chromatin.

Whilst the problem of nucleosome positioning has been investigated by more than two decades, only recently genome wide experimental collections *in vitro* and *in vivo* of the nucleosome maps excited rather heated discussions [1-3] about the predictability of nucleosome positioning in terms of the sequence. In particular, some authors proposed the existence of a sequence code [4-15].

In fact, the pioneering paper by Satchwell *et al.* [16] introduced the idea of a sort of nucleosome code in that the chicken erythrocyte nucleosomal sequences resulted characterized by a statistical fluctuation of AA/TT dinucleotide step in phase and GC out of phase with the B-DNA period. A successive paper by Drew and Travers [17] demonstrated that the AA/TT small groove faced in toward the histone core and GC outward as occurred in the circularization of the same DNA tract. These findings and the notion of the occurrence in phase of AA/TT dinucleotide step in curved DNAs were the basis of a theoretical model we proposed for predicting nucleosome positioning along DNA tracts based on the correlation between the calculated intrinsic curvature map of a given sequence and that averaged on the 177 chicken erythrocyte nucleosomal sequences [18]. The comparison with the experimental nucleosome positioning of a 1000 bp tract of *Sv40* DNA, the only experimentally known at that time, was rather satisfactory. This agreement suggested that the intrinsic curvature is a determinant of nucleosome positioning as well as the recurrence of AA/TT dinucleotides in phase with the DNA helical periodicity given the similarity of the characteristic maps associated to the chicken erythrocyte nucleosomal DNAs. A few years later, we proposed the first physical model for predicting the nucleosome positioning along a DNA sequence based on the calculation of the elastic energy needed to distort each recurrent 147 bp tract from its intrinsic curvature into the nucleosomal shape [19]. The trend of the energy function was very similar to that previously obtained using the correlation function of curvature map. The agreement between the minima of the energy function and the experimental nucleosome positions of *Sv40* DNA suggested a physical explanation for the intrinsic sequence-dependence of histone-DNA association.

As a matter of fact, genome-wide analyses reveal that nucleosomes exhibit intrinsic DNA sequence preferences and a common nucleosomal pattern in that nucleosomes are often absent at enhancer, promoter, and terminator regions [20]. Furthermore, comparison between nucleosomes assembled *in vitro* on yeast genomic DNA, and on *Escherichia coli* DNA under the same conditions, strongly indicates that eukaryotic cells are evolutionarily selected for sequences that favor nucleosome formation: in fact, the ratio of nucleosome occupancy on yeast reach a value of 9 to 1 with respect to the prokaryotic *E. coli* [21].

Adopting bioinformatic approaches, Trifonov and colleagues [22-26] and later Segal *et al.* [4], Field *et al.* [13], Kaplan *et al.* [14], and Tillo *et al.* [15] attempted to decode the experimental nucleosome maps of genomes in terms of certain dinucleotide step periodicity observed along DNA. Such investigations showed that highly significant information about the occurrence of a nucleosome along DNA is intrinsic in certain features of the sequence suggesting that DNA sequence of eukaryotic genomes encodes nucleosome organization.

On principle, the differential binding of the histones to DNA is governed by the pertinent free energy of the association process. In the absence of competing selective binding of components different from the histones, as it occurs in nucleosome reconstitution *in vitro*, the differential free energy function along the sequence, depends mainly on the intrinsic affinity of the histones to the different DNA tracts and eventually on the interactions between the nucleosomes in chromatin fibers. Furthermore, the change of topological state of the DNA chain consequent to the nucleosome formation should be considered as an additional factor as well as the competitive stabilization of G-quadruplex DNA structures in G-rich genome regions, as recently proposed [27].

However, *in vitro* data could not be representative of direct histone-DNA binding *in vivo* since other factors may drive the reposition of nucleosomes [28-30]. In addition, models based on sequences dependent nucleosome positioning require a reference sequence that could change among functional categories of DNA and organisms. In fact, the competition with specific DNA-binding activator proteins could dramatically affect the nucleosome occupancy *in vivo*.

Besides the bioinformatic approaches, theoretical models were proposed for the prediction of the nucleosome positioning along DNA sequences. These are mainly based on the sequence-dependent conformational features of the DNA chain, which regulate the energy required to transform recurrent 147 nt DNA tracts along the genome into the nucleosomal structure. Nonspecific histone-DNA interactions are generally modeled with a first-order elastic potential that penalizes the deviations of the intrinsic structure of a DNA tract from

the superhelix obtained by fitting the nucleosome X-ray crystal structures [31]. These would represent the main differential contribution to the average interactions of the histone octamer with the sugar-phosphate chain.

In fact, experimental results achieved in different laboratories showed that the thermodynamic binding constant of the histone octamer changes with the sequence of DNA. Furthermore, free energy data of competitive nucleosome reconstitution experiments obtained with a large pool of different DNA tracts strongly suggested a significant relation with the DNA intrinsic curvature [32-49].

The curvature function is generally obtained by integrating along the DNA the conformational deviations of the dinucleotide steps from the canonical B-DNA structure. Such deviations were evaluated theoretically on the basis of conformational energy calculations [50] as well as empirically from the average deviations observed in X-ray crystal structures of a significant number of double helical oligonucleotide tracts. The first moment of the empirical distributions of dinucleotide geometries of a set of double helical oligonucleotides [51] as well as those of protein-DNA X-ray crystal structures [31,52] represents an evaluation of the deviations of dinucleotide steps from the canonical B-DNA structure. The central second moment is adopted to evaluate the relative flexibilities.

However, it should be noted that the oligonucleotide crystal structures are generally characterized by the presence of AT rich central regions whilst the GC rich sequences are segregated at the ends and then more affected by crystal packing effects. Therefore, DNA-protein complexes are at present adopted to evaluate the deviations of each dinucleotide step and the relative flexibility since considered slightly influenced by the crystal packing. However, also in this case the conformation of the dinucleotide steps could be sensitive to the specific interactions with the different proteins that cannot be cancelled by averaging over the pool of the nucleoprotein complexes. On the contrary, the interactions between the nucleosomal DNA tract and histones more extended than those with the regulative proteins and the relatively low variance of the affinity with different nucleosomal sequences, suggest that the main interactions are non-specific and involve the phosphodiester chain prevalently.

As matter of fact, the two sets of dinucleotide parameters obtained by surveying the x-ray crystal structures of double-helix oligonucleotides [51] and DNA-protein complexes [31,52] appear weakly correlated (the correlation factor is 0.64). Furthermore, Morozov *et al.* [31] provided roll angles, the major determinant of the DNA curvature, different from those proposed by Olson *et al.* [52] in spite of the identical method used to obtain them.

However, the poor agreement between the two sets of data could be due to the more extensive database adopted by Morozov *et al.* [31].

It is significant that the normalized roll angles proposed by Morozov *et al.* [31] are very similar to those achieved by minimizing the conformational energy of the different dinucleotide steps we proposed about 25 years ago [50] and more recently confirmed adopting similar calculations by Packer *et al.* [53]. Fig. 1 reports the comparison among the roll angles of the dinucleotide steps by Morozov *et al.* [31], those proposed by Packer *et al.* [53] and ours [50]. The roll angles are normalized with respect to their own average value in order that a DNA tract with random sequence appears straight. Interestingly, the correlation factor between the roll parameters by Morozov *et al.* [31] and our theoretical roll values [50] is 0.97, significantly higher than the correlation factor with the roll proposed by Olson *et al.* [52] ( $R=0.68$ , data not reported in the figure).

Most theoretical models assume that the histone-DNA potential is a quadratic function of the deviations of the dinucleotide parameters from those of the ideal superhelix whose pitch and radius are inferred from the nucleosome crystal structures [54]. On the contrary, in the approach we firstly proposed [18], the elastic energy is calculated as the minimum energy needed to distort the intrinsic curvature of the naked DNA in the nucleosomal shape. In such a way, the specificity of the sequence is implicitly taken into account even with those particular sequences such as oligoA/T or G-rich tracts that are considered to inhibit the nucleosome formation.

This paper critically analyzes the different approaches for the prediction of nucleosome positioning and their validity by comparing the theoretical results with the experimental nucleosome occupancy along genomes. The comparison between the bioinformatic and physical models is also reported and critically discussed.

## **2. Genome-wide nucleosome occupancy and nucleosome code.**

The isolation of nucleosomal DNAs is generally obtained from the purified chromatin stabilized with formaldehyde and digested with micrococcal nuclease, which cleaves the DNA linkers, and successively treated with proteinase to remove proteins as well as to reverse the formaldehyde cross-links.

To measure nucleosome occupancy on a genomic scale, Yuan *et al.* [20] developed a DNA microarray method to identify nucleosomal and linker DNA sequences on the basis of



sensitivity of linker DNA to micrococcal nuclease. Nucleosomal DNA was isolated, labeled with Cy3 fluorescent dye (green), and mixed with Cy5-labeled total genomic DNA (red). This mixture was hybridized to microarrays printed with overlapping 50-mer oligonucleotide probes tiled every 20 base pairs across chromosomal regions of interest.

Sequence specificity of micrococcal nuclease and sequencing bias are then removed by normalizing the microarray data with the hybridization data of the pertinent naked DNA genome tracts. A profile of the logarithm of the ratio between the nucleosomal and genomic DNA hybridization data represents the nucleosome density diagram along the genome. Such a procedure provides a map of nucleosome occupancy *in vivo*. Reconstituted chromatin from purified genomic DNA fragments and histones, are used to produce a nucleosome occupancy profile along the sequence *in vitro*.

As an alternative experimental method, Valouev *et al.* [55] used the ultra-high-throughput Applied Biosystems SOLiD (Sequencing by Oligonucleotide Ligation and Detection) sequencing technology to characterize nucleosome positions in a mixed-tissue population of *Caenorhabditis. elegans* cells. As an experimental control and reference data set, the SOLiD technology was used to sequence *C. elegans* genomic DNA digested with micrococcal nuclease. These analyses provide nucleosome maps at extremely high density and resolution and a global view of the chromatin architecture of a multicellular animal.

Nucleosome occupancy reflects the average histone levels on a given region of DNA within a cell population; however, it does not address where an individual nucleosome is positioned with respect to a certain DNA sequence. Nucleosome positioning generally refers to the rotational and translational relationships between the histone octamer and the DNA chain. Rotational positioning defines the orientation of the DNA helix on the histone surface and is driven by the minimization of the DNA bending energy, reflecting also possible preferences for dinucleotides that face inwards or outwards with respect to the histone core. The translational position of a nucleosome refers to the specific 147 nt sequence covered by the histone octamer, and is generally assigned to the midpoint of this sequence in the nucleosome positioning diagram. Differently positioned but rotationally phased nucleosomes within a given genomic region will contribute to nucleosome density. Therefore, positioning of nucleosomes along DNA and nucleosome occupancy are distinct, although related issues.

The profile of nucleosome occupancy constitutes the basic information to achieve positioning of nucleosomes along the genome adopting the Hidden Markov Model (HMM) procedure [20]. This is a powerful method for assigning probabilities to hidden

nucleosomal states once the nucleosomal length, variable linker lengths as well as the density of nucleosomes are introduced. Thus, the resulting nucleosomal map along the genome is conditioned by these constraints and the range of probes.

*In vivo* nucleosome maps reflect the combined action of multiple factors, including the action of chromatin remodelers, competition with site-specific DNA-binding proteins, and the DNA sequence preferences of the nucleosomes themselves.

As a matter of fact, the statistical analysis of the nucleosome distributions along genomes in terms of the recurrences of the different dinucleotide steps along the nucleosomal DNAs results in the statistical preference of AA, TT, and AT dinucleotide small groove to face in towards the histone core and GC outward. Such periodicities and phase relationships were early derived from a collection of 177 natural nucleosomes from chicken erythrocytes [16]. Later, Lowary and Widom [39] carried out a *SELEX* experiment starting with a large pool of chemically synthetic random DNA molecules to identify those with the highest affinity for histone octamer. A set of highest-affinity molecules were selected, cloned, and sequenced, and their free energies in nucleosome reconstitution measured. Non-random occurrences of  $n$ -mers where  $n = 2$  to 10 arose in different analyses carried out in the selected sequences. In particular, a statistical recurrence of TA in phase with the B DNA periodicity was also revealed. These results supported the hypothesis of sequence rules, which govern nucleosome affinity and positioning. However, since it was not known whether these sequence preferences have a significant influence on nucleosome position *in vivo*, Segal *et al.* (2006) [4] isolated nucleosome-bound sequences at high resolution from yeast and used these sequences in a new computational approach to construct and validate experimentally a nucleosome-DNA interaction model, and to predict the genome-wide organization of nucleosomes. Sequence-based model for nucleosome positioning uses the *in vitro* map to represent the sequence preferences of nucleosomes by devising a probabilistic model that assigns a score to every 147-bp sequence. This model is based on both the position-dependent and global preferences of sequences of length 5 base-pair that the authors characterized previously. Then, the scores were used to compute the genome-wide distribution over nucleosome positions, taking into account steric hindrance constraints between neighboring nucleosomes [4].

On the basis of these experimental results, Segal *et al.* [4], Field *et al.* [13], Kaplan *et al.* [14] devise a computational model in which nucleosome occupancy is governed only by the intrinsic sequence preferences of nucleosomes. Moreover, since *in vitro*, nucleosome depletion is evident at many transcription factor binding sites and around gene start and end

sites, they suggest that also nucleosome depletion at these sites *in vivo* is partly encoded in the genome.

Different experimental findings were showed by Caserta *et al.* [56] who derived from *in vivo* positions, accurately mapped by partial micrococcal nuclease digestion, a translational positioning signal that identifies the approximate midpoint of DNA bound by a histone octamer. These authors conclude that the midpoint is, on average, highly A/T rich and, in particular, the dinucleotide TpA occurs preferentially at this and other outward facing minor grooves and suggested that the enrichment of AT-containing dinucleotides at the centre is required for local untwisting.

More recently, the analysis of the *C. elegans* genome allowed the Trifonov group to identify a consensus sequence that occurs along the nucleosomal sequences [25,26]. Based on a strong ~10 base periodicity of AA and TT dinucleotides associated with the nucleosomes evidenced in the genome of *C. elegans* [57], Salih *et al.* [25] and Gadbank *et al.* [26] measured prevailing distances between all dinucleotide steps in the database of the nucleosome fragments of *C. elegans* and identified the consensus sequence (CGGAAATTTCCG) of the nucleosome DNA repeat of this genome.

Accordingly, if the existence of a nucleosome code is assumed, the pattern of nucleosome positioning is primarily determined by genomic DNA sequence and hence can be predicted.

### **3. Experimental evaluation of differential nucleosome stability by competitive nucleosome reconstitution.**

Competitive nucleosome reconstitution method provides quantitative estimates of the sequence-dependent differential thermodynamic nucleosome stability, and strongly indicates that nucleosome positioning along a DNA sequence occurs with different affinity [32,33].

This experimental approach measures the ability of a labeled DNA fragment to compete with bulk DNA for a limited number of histone octamers during the nucleosome reconstitution procedure.

A DNA competitor labeled with  $^{32}\text{P}$  is added to a solution of standard nucleosomes (e.g. erythrocyte nucleosomes obtained by micrococcal nuclease digestion of the chromatin) and of the corresponding naked nucleosomal DNAs. At high ionic strength the nucleosomes dissociate; the successive stepwise dilution [32] or dialysis [39] procedure provide the

competitive nucleosome reconstitution with the labeled DNA. The percentage of a DNA fragment that had been incorporated into nucleosomes was assayed by separating free DNA from complexed DNA in polyacrylamide gels. The concentration excess of the standard nucleosomes permits the evaluation of the differential association constant and the corresponding differential free energy by measuring the ratio of radioactivity of the reconstituted nucleosome and the corresponding free DNA.

It should be noted that the length of the DNA tracts investigated generally exceeds the nucleosomal length and therefore the free energy contains an entropy contribution in account of the virtual multiplicity of the energy minima distributed along the sequence with a period of about 10 nt.

Differential free energy data were collected by different authors for synthetic and natural DNA tracts with different length and intrinsic curvatures [32-38,40-49] as well as sequences with virtually highest affinity as obtained by *SELEX* procedures [39]. Despite the DNA tracts investigated are characterized by a variety of sequences, lengths and curvatures, the free energy range is limited to a few kilocalories per nucleosome. It is interesting to note that the free energies involved in nucleosome formation on the telomeres are the highest among the synthetic and biological sequences investigated so far [38,40,41,45-47]. The low affinity of telomeric sequences for the histone octamer is consistent with 6-8-bp repeated lengths, therefore out of phase with the B-DNA period.

#### 4. DNA architecture and common features of the X-ray nucleosome structures

The first model of nucleosome was proposed by Kornberg in 1977 [58] and validated with the neutron scattering studies performed on dilute solutions [59]. The structure was later obtained at low resolution by electron microscopy [60]. The nucleosome particle is characterized by a solenoid-like structure where a 146 bp DNA tract is wrapped around the protein core of the histone octamer. The crystal structure of the nucleosome core particle, solved to 7 Å resolution by Richmond *et al.* (1984) [61], firstly evidenced that B-DNA superhelix on the outside contains several sharp bends. High-resolution electron density map later obtained by Lüger *et al.* (1997) [62] confirms the previous low-resolution structure and shows atomic details of both DNA and histone proteins revealing a pseudo-dyad symmetry. This nucleosome structure was obtained on crystals of nucleosome reconstituted *in vitro* with a 146 bp palindromic DNA sequence, which was expected to

adopt the dyad symmetry of the protein core. Actually, the structure is characterized by a pseudo-dyad axis that relates the four pairs of histones and the two phosphodiester chains but lies on the average plane of a central base pair. This feature was found in all the crystal structures of nucleosomes reconstituted with different DNA sequences. This finding suggests that the nucleosome is mainly stabilized by the interactions of histone octamer with the phosphodiester chains while the interactions with the bases should play a minor role.

In fact, the detailed analysis of the 24 X-ray structures of nucleosomes at relatively high-resolution, shows that the DNA superhelical shape appears to be highly conserved and have common features practically independent of the sequence. A picture emerges where the superhelix is characterized by the similar pattern of roll angles and slide displacements of dinucleotide steps along the nucleosomal DNA while the base-pair twisting remains close, on average, to that in solution [54].

In recent years, many authors proposed a structural approach as the basis of computational methods for predicting the nucleosome positioning. Tolstorukov *et al.* [7] emphasized the structural role of the lateral displacements of adjacent base pairs and evaluated the energy coupling between roll and slide deformations suggesting that the lower cost of deforming DNA on the nucleosome occurs at sites of large positive slide and negative roll, where the DNA bends into the minor groove. With a similar method, Wang *et al.* [63] showed that kink and slide deformations are likely to be stabilized by the arginine residues of histones interacting with the minor groove of DNA. In fact, the arginines are positioned asymmetrically in the minor groove, being closer to one strand. These asymmetric interactions should facilitate lateral displacement of base pairs across the DNA grooves. These findings allowed the authors to propose a simple stereochemical model based on the evaluation of the sequence-dependent kink and slide deformation energy.

Xu and Olson [64] focused on DNA conformational signals found in the growing library of known high-resolution nucleosome structures and the ways in which these features may contribute to the positioning of nucleosomes on specific DNA sequences.

## **5. Prediction of the sequence-dependent nucleosome differential stability based on energy calculations.**

Among the numerous theoretical methods proposed for predicting nucleosome positioning

in terms of the sequence, only a few of these are based on the nucleosome differential thermodynamic stability. In fact, experimental investigations established that the affinity of histone octamer to DNA is sequence-dependent providing also the values of the binding free energy difference with respect to a chosen standard nucleosome, as discussed in the paragraph 3.

The physical basis of nucleosome stability lies in the sequence-dependent propensity of DNA to adopt the tightly bent configuration imposed by the binding of the histone proteins. The first attempt to predict the differential stability of nucleosomes in terms of DNA sequence was published about two decades ago. The method was based on the evaluation of the elastic energy to transform the DNA tract with its intrinsic curvature into the nucleosomal shape [18,19]. These calculations were carried out adopting the set of intrinsic dinucleotide internal parameters early obtained by conformational energy calculations [50]. These parameters were successful in predicting the gel electrophoretic mobility of a thousand DNA tracts with different sequences and length [65] as well as the circularization thermodynamic constants of a pool of DNA tracts [66]. More recently, curvature and flexibility profiles of long DNA tracts, directly derived from AFM images, were theoretically reproduced on the basis of these dinucleotide parameters [67,68].

The differential thermodynamic stability of the association between the histone octamer and different DNA tracts is governed by the relative differential free energy function. The evaluation of the free energy difference would require the calculation of a huge number of local interactions involving the histone amino acid residues and the different nucleotides, water molecules and counterions. Furthermore, the inter-nucleosomal interactions should be considered. Theoretical calculations, which account for all kinds of interactions, are not computationally feasible. However, the complexity of the system, allows a statistical thermodynamic approach for evaluating the differential free energy of nucleosome formation based on the collective features of different DNA regions virtually involved in the association with the histone core.

Experimental findings show that nucleosomal DNAs have a similar superhelical shape of the double helix around the histone core largely independent of sequence [54]. Therefore, we could consider the nucleosome differential stability mainly dependent on the conformational states of the naked DNA sequences due to the practical invariance of the histone octamer and the nucleosomal shape.

Very recently, Morozov *et al.* [31] developed a biophysical model for the sequence dependence of DNA bending energies, and validated it against a collection of *in vitro* free

energies of nucleosome formation and a set of in vitro nucleosome positions mapped at high resolution. The authors also made an *ab initio* prediction of nucleosomal DNA geometries, and checked its accuracy against the nucleosome crystal structure reproducing with good accuracy the roll and tilt dinucleotide angles, whereas the other parameters showed discrepancies.

In this model the total energy  $E$  of a nucleosomal DNA is given by a weighted sum of two quadratic potentials:

$$E = E_{el} + wE_{sh} \quad (1)$$

where  $E_{el}$  is the sequence-specific DNA elastic energy and  $E_{sh}$  represents non-specific histone-DNA interaction energy. The DNA elastic energy is formulated as:

$$E_{el} = 1/2 \sum_{s=1}^N \left[ \mathbf{R}_s (\mathbf{a}^s - \langle \mathbf{a}^{n(s)} \rangle) \right]^T \mathbf{F}^{n(s)} \left[ \mathbf{R}_s (\mathbf{a}^s - \langle \mathbf{a}^{n(s)} \rangle) \right] \quad (2)$$

where  $\mathbf{a}_s$  is the six component vector of angles (twist, roll and tilt) and displacements (rise, shift and slide) for each dinucleotide step and  $\langle \mathbf{a}^{n(s)} \rangle$  are the average values of the local degrees of freedom computed for all basesteps ( $n=AA, AC, AG, \dots, TT$ ) using a collection of oligonucleotides extracted from a set of 101 non-homologous protein-DNA structures. The matrix of force constants  $\mathbf{F}^{n(s)}$  is evaluated by inverting the covariance matrix of deviations of local geometric parameters from their average as observed in the set of crystal structures. The matrix  $\mathbf{R}_s$  transforms the local parameters into the global frame in which the nonspecific histone-DNA interactions are given. These empirical parameters for each of the six local degrees of freedom of the ten independent dinucleotide steps are reported in the Supplementary Tables 1-3 of the reference 31.

Nonspecific histone-DNA interactions are modeled with a quadratic potential that penalizes deviations of nucleosomal DNA from the ideal superhelical shape:

$$E_{sh} = \sum_{s=1}^N (\mathbf{r}_s - \mathbf{r}_s^0)^2 \quad (3)$$

where  $\mathbf{r}_s$  and  $\mathbf{r}_s^0$  are the nucleosomal DNA and the ideal superhelix radius vectors in the nucleosome particle, respectively

The energy was then minimized with respect to all the six geometrical parameters, angles

and displacements, which determine the local structure of the dinucleotide steps.

The final conformation of a given nucleosomal DNA is the one corresponding to the minimum of the total energy  $E$  of a DNA sequence given in arbitrary units. The empirical parameter  $w$  in Eq. 1 was obtained by optimizing the average correlation between the distribution of geometric parameters observed in a high-resolution nucleosome crystal structure and the corresponding predictions.

The statistical sum over all possible configurations of non-overlapping nucleosomes in a DNA sequence is given by:

$$Z = \sum_{conf} \exp(-E(conf)) \quad (4)$$

where the configurational energy  $E(conf)$  is in arbitrary dimensionless units. It was obtained by recursively computing of partial statistical sums, after adding each nucleosome to the configuration. This procedure permits the calculation of the partial free energy and the probability of finding a nucleosome at a sequence position, consequently.

In alternative, the approach we previously proposed evaluates the minimum free energy to transform a free DNA tract into the nucleosomal shape preserving the maximum of intrinsic features of the sequence [69,70]. A mathematical formulation based on the Parseval equality, which allows the integration of quadratic differences of two functions in terms of the sum of the differences between the related Fourier transform amplitudes, was adopted [71].

In this model the elastic energy of a nucleosomal DNA is expressed as a sum of bending and twisting contributions:

$$\Delta E_{el}^o(k) = \frac{b(k)}{2L} |\mathbf{A}_n(\mu) - \mathbf{A}_f^o(\mu)|^2 + \frac{t(k)}{2} (2\pi \Delta T_w(k))^2 \quad (5)$$

$$\text{with } \mathbf{A}_f^o(\mu) = \sum_{k-L/2}^{k+L/2} \mathbf{C}_f^o(s) \exp\left(-\frac{2\pi i \mu s}{L}\right) \quad \text{and} \quad \mathbf{A}_n(\mu) = \sum_{k-L/2}^{k+L/2} \mathbf{C}_n(s) \exp\left(-\frac{2\pi i \mu s}{L}\right)$$

$b(k)$  and  $t(k)$  are the apparent isotropic bending and twisting force constants, and  $\Delta T_w(k)$  represents the change of twisting number of the  $k$ th free DNA tract.  $\mathbf{C}_n(s)$  and  $\mathbf{C}_f^o(s)$  are the curvature vectors, represented in the complex plane, of the nucleosomal DNA and that pertinent to the free DNA relative to the  $s$ th basestep in the  $k$ th tract;  $L=146$  is the length of



nucleosomal DNA in bp.

On the basis of the Parseval equality, the bending energy is conveniently approximated to a single term with frequency  $\mu = -0.18$ .  $A_n(\mu)$  represents the curvature function in the Fourier space and its value is  $10.9 \text{ rad}$ . This approximation was successfully adopted in our previous papers to predict the propensity to circularization of a pool of DNA fragments [66] and the writhe transformations of circular DNAs [72]. Assuming that all the amplitude differences vanish, except that characterized by the periodicity  $\mu$ , the maximum part of the intrinsic curvature features is conserved and the elastic energy is minimized. This is compatible with the hypothesis that the deviations from the ideal uniform superhelix observed in the x-ray structure could be due in part to the intrinsic curvature of the DNA sequence.

This formulation of the elastic energy permits the calculation of the canonical partition function, and in turn the evaluation of the elastic free-energy difference pertinent to the  $k$ th nucleosome. The thermodynamic affinity,  $\beta\Delta G_{el}$  ( $\beta = 1/RT$ ), of the DNA tract relative to the standard straight DNA with a random sequence:

$$\beta\Delta G_{el}(k) = \beta\Delta E_{el}^o(k) + Z(k) - Z(k)\cos\phi - \frac{3}{2}L\ln\left(\frac{b(k)}{b^*}\right) \quad (6)$$

where  $Z(k) = -\beta b(k)A_n(\mu)A_f^o(\mu)/L$ .  $A_n(\mu)A_f^o(\mu)$  represents the modulus of the correlation between the superstructure of the nucleosomal DNA and that of the free form, according to the convolution theorem [71].  $\Delta E^o(k)$  contains the twisting and ground state bending energy contributions when the phase angle of  $A_n(\mu)$  and that pertinent to the  $k$ th DNA tract,  $A_f(\mu)$ , are equal;  $\phi$  is the phase angle between the nucleosome dyad axis and that of the intrinsic curvature of the free DNA tract. The ratio of the bending force constants of the  $k$ th DNA tract and the standard DNA,  $b(k)/b^*$ , is represented by the ratio of the dinucleotide melting temperatures [73] (in thermodynamic scale) averaged over the tract considered. As discussed in our previous papers [69,70], the normalized melting temperatures were adopted as parameters to estimate the rigidity of the dinucleotide steps. This assumption was found consistent with the statistical mechanics analysis of AFM images of a number of DNA tracts [67,68]. A similar evaluation of DNA flexibility was later proposed by Travers and Thompson [74].

The free energy difference is characterized by periodical fluctuations with the period of the

B-DNA, which determine the phasing of nucleosomes along the sequence. The statistical mechanics averaging of the different phasing of nucleosome over a turn of DNA leads to:

$$\beta\Delta G_{el}(k)=\beta\Delta E_{el}^o(k)+Z(k)-\ln(J_0(iZ(k)))-\frac{3}{2}L\ln\left(\frac{b(k)}{b^*}\right) \quad (7)$$

where  $J_0(iZ(k))$  is the zero-order Bessel function of the imaginary argument  $Z(k)$ . The enveloping curve of the profiles of the free energy minima of Eq. (6) practically coincides with those obtained with Eq. (7).

The last three terms in Eqs. (6) and (7) represent the entropy contribution to the thermodynamic stability of a  $k$ th nucleosome along the DNA sequence. These three terms vanish in the case of straight DNA tracts with random composition, where  $A_f = 0$  and  $b(k) = b^*$ .

As a consequence, the curvature function of a nucleosomal DNA can be considered sequence-dependent in that the intrinsic features of the DNA tract modulate the ideal superhelix shape and implicitly define which bases face toward or outward the histone core. The comparison between the experimental values of the free energy difference of a large pool of DNA tracts, different for sequence and length, with the theoretical elastic free energy, gives a satisfactory agreement for slightly curved DNA but increasing deviations for rather curved DNA. However, the analysis of these deviations showed a strict correlation with the  $\langle A_f^o \rangle$  ( $R = .98$ ) namely, with the average integral curvature of the free DNA that appears to destabilize the nucleosome [48]. This result supports the hypothesis of an additional energy contribution to the free energy difference, dependent on the free DNA curvature. This contribution was interpreted as due to the differential interactions of the water and counterions with the reduced small groove in curved DNA: e.g. the spine of water as found in x-ray crystal structures [75] and in solution [76] in AA·TT rich tracts of double-helical oligonucleotides. Incidentally, the periodical repetitions of such dinucleotide steps also represent the sequence feature that produces the intrinsic curvatures of DNA tracts. Therefore, the intrinsic curvature of the free DNA plays two opposite roles in the nucleosome formation. One stabilizes the nucleosome by reducing the energy required transforming the DNA tract into the nucleosome shape; the second, related to the differential interactions of curved DNA with water solution, destabilizes the nucleosome.

This energy contribution, introduced as an empirical term ( $4.5 \cdot \langle A_f^o \rangle^{1.5}$  in  $RT$  units) in the Eqs. (6) and (7), led to a satisfactory agreement ( $R=0.92$ ) between experimental and predicted free energy values of about hundred DNAs investigated in different laboratories [32-49]. The comparison between the experimental and theoretical nucleosome reconstitution free energies is reported in Fig. 2.

It is interesting to note that the last term in the Eqs. (6) and (7) decreases the free energy if the DNA tract is relatively rigid. Thus, the flexibility of the DNA tract also plays a dual role; the bending energy decreases due to the lower bending constant while the entropy change increases from a flexible state of the free form to the essentially rigid nucleosomal state.

The good agreement between experimental and theoretical free energy differences settles the base for a physical model capable of predicting the most stable positioning of nucleosomes along genomic DNA [77-78].

A direct test of how accurately the two models proposed by Morozov *et al.* [31] and Anselmi *et al.* [69] predict nucleosome position on DNA can be provided by a comparison of energy and free energy (in  $RT$  units) trends for six sequences where *in vitro* nucleosome positions are known with 1-2 bp accuracy, as illustrated in Fig. 3 (rearranged from Figure 5 of the reference 31). Unlike longer genomic sequences, only one nucleosome can form on the shorter sequences considered here and its position is typically determined by the global energy or free energy minimum, except for two sequences with two experimentally mapped alternative positions. The profiles of energy calculated according to Morozov *et al.* [31] (red lines) and free energy calculated according to our model (blue lines) are overlapped and reported along the sequence. Nucleosome positioning is given with respect to the dyad axis. The vertical lines and arrows (green) indicate the experimentally known nucleosome positions. As illustrated in Fig. 3, the two models produce comparable results in good agreement with the experimentally mapped nucleosomes.

It should be noted that the theoretical free energy values as well as the experimental nucleosome occupancy data are related to a large number of copies of the DNA tract. However, the nucleosome positioning resulting from the localization of the minima of the free energy function of the virtual mononucleosome does not account for the contributions arising from the multiplicity of the free energy map around a minimum as well as the interaction effects with other nucleosomes in compacted chromatin fibers. To regard these effects, the mononucleosomal free energy can be averaged over an interval compatible with

the experimental average nucleosome density represented by the nucleosome repeat length (*NRL*).

$$\beta\Delta G(k) = -\ln \left[ \sum_{j=k-n}^{k+n} \exp(-\beta\Delta G(j)) \right] \quad n = (NRL - L)/2 \quad (8)$$

Furthermore, to account for the effects of nucleosome packing into chromatin fibers the free energy profile can be modulated with a simple periodical function with a periodicity equal to *NRL*:

$$\frac{1}{2} \left[ 1 + \cos \left( \frac{2\pi(k - k_o)}{NRL} \right) \right] \quad (9)$$

where  $k_o$  represents the dyad position of strongly positioned nucleosomes, which could represent nucleation sites of the nucleosome compaction in chromatin fibers.

In fact, Valouev *et al.* [55] recently pointed out that relative positioning of nucleosomes on *C. elegans* genome appears to be a significant property of chromatin structure.

A physical approach, similar to that we had early proposed [69,70], was suggested by Miele *et al.* [79]. The authors predict the nucleosome occupancy along the genomes of *Saccharomyces cerevisiae* and *Drosophila melanogaster* using only sequence-dependent DNA flexibility and intrinsic curvature to evaluate the free energy required for the nucleosome formation transforming the intrinsic curvature of a recurrent DNA tract into the nucleosomal superhelix. For the wrapping length, Miele *et al.* [79] chose  $L=73$  bp, which provided the best correlation with the experimental data. Furthermore, these authors adopt our set of roll, tilt and twist angles associated to the different dinucleotide steps and the normalized melting temperatures to evaluate the entropy contribution according to the last term of Eqs. (6) and (7) [69]. As a consequence, they obtained results comparable with ours as regards the prediction of nucleosome occupancy along the yeast genome. In the fly, Miele *et al.* predict promoter strength as encoded in distinct chromatin architectures characteristic of strongly and weakly expressed genes. In addition, they showed that the other dinucleotide parameters (rise, shift and slide) as well as the fine features, which modulate the superhelical shape in the nucleosome crystal structures, do not add significant contributions to the free energy [79]. This paper confirms our first hypothesis that only

intrinsic mechanical properties of the free DNA are necessary to describe a substantial part of chromatin structural complexity, thus providing the physical basis for methods capable of predicting nucleosome occupancy sounder than those based on recurrent DNA sequence motifs.

## 6. Comparative results and discussion.

The relations between the bioinformatic and biophysical approaches are illustrated in Fig. 3 where the rotational distribution of AA/TT/TA/-GC dinucleotide steps per nucleosome along a tract of yeast genome (210000-228000 bp of chromosome III) is compared with the minimum distortion energy needed to transform the intrinsic curvature into the nucleosomal shape calculated according to Eq. (6) adopting Anselmi *et al.* [69] (Fig. 3A) and Morozov *et al.* [31] (Fig. 3B) internal dinucleotide parameters (roll, tilt, twist and elastic constants). The profiles are quite similar indicating that the periodicity of these dinucleotide steps represents an important contribution to the nucleosome conformational energy. Interestingly, the energy profiles are characterized by the periodical fluctuations coherent with the B-DNA structure and represent the effect of nucleosome phasing (see the insert in Fig. 3A). Furthermore, the similarity of the energy profiles indicates the self-consistency of the two sets of dinucleotide parameters obtained theoretically by conformational energy calculations [50] and empirically by analysis of X-ray crystal structures of DNA-protein association complexes [31].

However, the comparison of the energy profile with that of the experimental nucleosome occupancy, represented by the  $\log_2$  ratio of the hybridization values of nucleosomal and genomic DNA carried out by Yuan *et al.* [20], does not show a satisfactory agreement suggesting that the role of entropy contributions to the free energy is important.

In fact, the profiles of the experimental nucleosome occupancy [20] and the entropy contribution calculated according to the last term of Eqs. 6 and 7 for a tract of the yeast genome (210000-228000 bp of chromosome III) (Fig. 4) have similar trend showing that an important part of the experimental trend is ruled by the entropy contribution. Therefore, the stability of nucleosomes appears favored in DNA tracts characterized by higher stiffness since the entropy term is a monotonous function of the normalized force constants. As early discussed in our papers [69,70], the DNA stiffness plays a dual role in the nucleosome stability: it increases the energy required for reaching the nucleosomal shape but decreases

the loss of entropy in transforming the free DNA in solution in a more rigid nucleosomal structure.

As a matter of fact, the free energy profile and that obtained from the Boltzmann averaging according to Eq. (8) are compared with that of the corresponding nucleosome occupancy. Some features of the chromosomal map are shown evidencing the genes along the sequence. The experimental and theoretical profiles show a good correlation although single deviations are present (Fig. 5). Moreover, the free energy profile shows maxima in those regions where *in vivo* map [20] identifies nucleosome-depletion around transcription start and stop sites. It should be noted that the weighted free energy contains the entropy terms due to the multiplicity of nucleosome occupancy on the near free energy minima, is a continuous function. The trend is well reproduced and also a significant part of the periodical fluctuations of the nucleosome occupancy due to the packing of the nucleosomes in the chromatin fibers.

In fact, compaction of nucleosomes in chromatin requires regularity in the nucleosome array along the DNA chain. It is plausible that the most stable nucleosome positions nucleate the nucleosome occupancy along the chain with multiple steps of DNA turns.

Compaction of nucleosomes in chromatin is directed by their positioning along DNA and therefore by the length of the DNA linkers that bridge and mutually orient the adjacent nucleosomes in the space. Their average length varies among species, cell types within the species as well as among nucleosomes within the same cell type. A statistical analysis of their length distribution in chromatin reveals that nucleosome repeat length differs in multiples of about 10 bp, close to the DNA helical repeat [80]. Such a finding suggests the existence of orientation constraints between the nucleosomes and between the nucleosomes and the fiber axis. Such features are plausibly dictated by the optimization of the mutual orientation of the nucleosomal units as a necessary condition for the best packing in the chromatin fibers [81]. In fact, it is interesting that DNA curvature as well as the rotational distribution of dinucleotide AA/TT/TA/-GC, which represents the nucleosome code [4-13-15], along tracts of yeast genome are characterized by a significant coherence of phase. As matter of fact, the analysis of the experimental nucleosome occupancy along the genome reveals a quasi periodical modulation as a result of the periodicity in the nucleosome positioning required for an effective nucleosome packing in chromatin [20,82].

This collective assembly of nucleosomes in chromatin fibers could compete with the intrinsic mononucleosome affinity pattern giving rise to a possible reposition of

nucleosomes which would occur in the genome regions with relatively low nucleosome affinity.

To account for the effects of nucleosome packing into chromatin fiber, the free energy profile was modulated with a simple periodical function as reported in Eq. (9). This approach represents a rough but operative modeling to interpret nucleosome positioning imposing the need of the intrinsic mononucleosome stability and the necessary conditions for the packing into the chromatin fiber. As matter of fact, the comparison between theoretical and experimental nucleosome occupancy profiles along a tract of yeast genome *in vivo* (214000-225000 bp of chromosome III) appears rather satisfactory as illustrated in Fig. 6. The different colors indicate opposite phases. The profile is characterized by successive blocks of nucleosomes with opposite phases in connections to inversions of free DNA curvature and plausibly in relation with the superstructure of the chromatin fiber.

As a general conclusion, the genome-wide analysis argues that the sequence preference for nucleosome formation is consistent with the idea that DNA intrinsic curvature and flexibility drive the rotational and translational positioning *in vitro* as well as *in vivo*.

This result seems not consistent with the conclusion reported by Zhang *et al.* [21] that intrinsic histone-DNA features are not the major determinant of nucleosome positioning *in vivo* and with the suggestion that the apparent similarities of histone densities between *in vitro* and *in vivo* samples are likely to be inflated by the DNA sequence specificity of micrococcal nuclease that is applied to the analysis of the mononucleosomal samples. It is however supported by the issue that translational positioning *in vivo* is influenced by relatively constant spacing between nucleosomes, which is presumably favored by the action of nucleosome-remodeling complexes nearby the minima of the free energy profile. Furthermore, it agrees with the findings that the *in vitro* map, in which nucleosome occupancy is governed only by the intrinsic sequence preferences of nucleosomes, is found similar to *in vivo* nucleosome maps generated in different growth conditions [6].

Finally, the rather high similarity between the rotational distribution map of the dinucleotide steps AA/TT/TA in phase and GC out of phase with the B-DNA periodicity, which characterizes the nucleosomal sequences, and the map of the elastic deformation energy required to transform the intrinsic DNA structure into the nucleosomal shape, provides a physical basis for the hypothesis of a nucleosome code (see Fig. 3). In fact, these dinucleotide steps are the major determinant of the DNA curvature as illustrated in Fig. 7, where the comparison between the intrinsic curvature of recurrent 146-bp sequences along

a yeast genome tract ( $|C_f|$ ) and the rotational distribution of these dinucleotide steps are reported in modulus (A) and phase (B). Therefore, the fact that histones prefer these sequences in the nucleosome formation is conceptually different from the issue that these sequences interact preferentially with the histones. Actually, the periodical recurrence of these dinucleotide steps determines the DNA curvature; then, it is such a mechanical property that plays an important role in the nucleosome stability.

The DNA sequences of yeast genome were downloaded from *Saccharomyces* Genome Database (<http://www.yeastgenome.org>).



**Acknowledgements**

This research was supported by the the Progetto Ateneo 60% 2006-2007 of the Università la Sapienza and the Istituto Pasteur Fondazione Cenci-Bolognetti.

## References

1. N. Kaplan, I. Moore, Y. Fondufe-Mittendorf, A.J. Gossett, D. Tillo, Y. Field, T.R. Hughes, J.D. Lieb, J. Widom, E. Segal, Nucleosome sequence preferences influence in vivo nucleosome organization, *Nature* 17 (2010) 918-920.
2. Y. Zhang, Z. Moqtaderi, B.P. Rattner, G. Euskirchen, M. Snyder, J.T. Kadonaga, X.S. Liu, K. Struhl, Evidence against a genomic code for nucleosome positioning, *Nature* 17 (2010) 920-923.
3. B.F. Pugh, A preoccupied position on nucleosomes, *Nature* 17 (2010) 923.
4. E. Segal, Y. Fondufe-Mittendorf, L. Chen, A. Thåström, Y. Field, I.K. Moore, J.Z. Wang, J. Widom, A genomic code for nucleosome positioning, *Nature* 442 (2006) 772-778.
5. P. Ioshikhes, I. Albert, S.J. Zanton, B.F. Pugh, Nucleosome positions predicted through comparative genomics, *Nat. Genet.* 38 (2006) 1210-1205.
6. H.E. Peckham, R.E. Thurman, Y. Fu, J.A. Stamatoyannopoulos, W.S. Noble, K. Struhl, Z.P. Weng, Nucleosome positioning signals in genomic DNA, *Genome Res.* 17 (2007) 1170-1177.
7. M.Y. Tolstorukov, A.V. Colasanti, D.M. McCandlish, W.K. Olson, V.B. Zhurkin, A novel roll-and-slide mechanism of DNA folding in chromatin: implications for nucleosome positioning, *J. Mol. Biol.* 371 (2007) 725-738.
8. W. Lee, D. Tillo, N. Bray, R.H. Morse, R.W. Davis, T.R. Hughes, C. Nislow, A high-resolution atlas of nucleosome occupancy in yeast. *Nat. Genet.* 39 (2007) 1235-1244.
9. S. Gupta, S. Dennis, R.E. Thurman, Y. Fu, J.A. Stamatoyannopoulos, W.S. Noble, Predicting human nucleosome occupancy from primary sequence, *PLoS Comput. Biol.* 4 (2008) e1000134.

10. T.N. Mavrich, I.P. Ioshikhes, B.J. Venters, C. Jiang, L.P. Tomsho, J. Qi, S.C. Schuster, I. Albert, B.F. Pugh, A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome, *Genome Res.* 18 (2008) 1073-1083.
11. P.V. Kharchenko, C.J. Woo, M.Y. Tolstorukov, R.E. Kingston, P.J. Park, Nucleosome positioning in human HOX gene clusters, *Genome Res.* 18 (2008) 1554-1561.
12. G.C. Yuan, J.S. Liu, Genomic Sequence Is Highly Predictive of Local Nucleosome Depletion, *PLoS Comput. Biol.* 4 (2008) e13.
13. Y. Field, N. Kaplan, Y. Fondufe-Mittendorf, I.K. Moore, E. Sharon, Y. Lubling, J. Widom, E. Segal, Distinct Modes of Regulation by Chromatin Encoded through Nucleosome Positioning Signals, *PloS Comput. Biol.* 4 (2008) e1000216.
14. N. Kaplan, I.K. Moore, Y. Fondufe-Mittendorf, A.J. Gossett, D. Tillo, Y. Field, E.M. Leproust, T.R. Hughes, J.D. Lieb, J. Widom, E. Segal, The DNA-encoded nucleosome organization of a eukaryotic genome, *Nature* 458 (2009) 362-366.
15. D. Tillo, N. Kaplan, I.K. Moore, Y. Fondufe-Mittendorf, A.J. Gossett, Y. Field, J.D. Lieb, J. Widom, E. Segal, T.R. Hughes, High Nucleosome Occupancy Is Encoded at Human Regulatory Sequences, *PLoS ONE* 5 (2010) e9129.
16. S. Satchwell, H.R. Drew, A.A. Travers, Sequence periodicities in chicken nucleosome core DNA, *J. Mol. Biol.* 191 (1986) 659-675.
17. H.R. Drew, A.A. Travers, DNA bending and its relation to nucleosome positioning, *J. Mol. Biol.* 186 (1985) 773-790.
18. D. Boffelli, P. De Santis, A. Palleschi, M. Savino, The curvature vector in nucleosomal DNAs and theoretical prediction of nucleosome positioning, *Biophys. Chem.* 39 (1991) 127-136.

19. P. De Santis, M. Fuà, A. Palleschi, M. Savino, Theoretical prediction of the gel electrophoretic retardation changes due to point mutations in a tract of SV40 DNA, *Biophys. Chem.* 46 (1993) 193-204.
20. G.C. Yuan, Y.J. Liu, M.F. Dion, M.D. Slack, L.F. Wu, S.J. Altschuler, O.J. Rando, Genome-Scale Identification of Nucleosome Positions in *S. Cerevisiae*, *Science* 309, (2005) 626-630.
21. Y. Zhang, Z. Moqtaderi, B.P. Rattner, G. Euskirchen, M. Snyder, J.T. Kadonaga, X.S. Liu, K. Struhl, Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo, *Nat. Struct. Mol. Biol.* 16 (2009) 847-853.
22. E.N. Trifonov, J.L. Sussman, The pitch of chromatin DNA is reflected in its nucleotide sequence, *Proc. Nat. Acad. Sci. USA* 77 (1980) 3816-3820.
23. G. Mengeritsky, E.N. Trifonov, Nucleotide sequence-directed mapping of the nucleosomes, *Nucleic Acids Res.* 11 (1983) 3833-3851.
24. Ioshikhes, A. Bolshoy, K. Derenshteyn, M. Borodovsky, E.N. Trifonov, Nucleosome DNA sequence pattern revealed by multiple alignment of experimentally mapped sequences. *J. Mol. Biol.* 262 (1996) 129-139.
25. F. Salih, B. Salih, E.N. Trifonov, Sequence structure of hidden 10.4-base repeat in the nucleosomes of *C. elegans* *J. Biomol. Struct. Dyn.* 26 (2008) 273-282.
26. Gadbank, D. Barash, E.N. Trifonov, Nucleosome DNA bendability matrix (*C. elegans*), *J. Biomol. Struct. Dyn.* 26 (2009) 403-411.
27. J.A. Capra, K. Paeschke, M. Singh, V.A. Zakian, G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in *Saccharomyces cerevisiae*, *PLoS Comput. Biol.* 6 (2010) e1000861.

28. C. Stockdale, A. Flaus, H. Ferreira, T. Owen-Hughes, Analysis of nucleosome repositioning by yeast [SW] and Chd1 chromatin remodeling complexes, *J. Biol. Chem.* 281 (2006) 16279-16288.
29. S. Pisano, E. Marchione, A. Galati, S. Mechelli, M. Savino, S. Cacchione, Telomeric nucleosomes are intrinsically mobile, *J. Mol. Biol.* 369 (2007) 1153-1162.
30. Y. Chaban, C. Ezeokonkwo, W.H. Chung, F. Zhang, R.D. Kornberg, B. Maier-Davis, Y. Lorch, F.J. Asturias, Structure of a RSC-nucleosome complex and insights into chromatin remodeling, *Nature Struct. Mol. Biol.* 15 (2008) 1272-1277.
31. A.V. Morozov, K. Fortney, D.A. Gaykalova, V.M. Studitsky, J. Widom, E.D. Siggia, Using DNA mechanics to predict in vitro nucleosome positions and formation energies, *Nucleic Acids Res.* 37 (2009) 4707-4722.
32. T.E. Shrader, D.M. Crothers, Artificial nucleosome positioning sequences, *Proc. Natl. Acad. Sci. USA* 86 (1989) 7418-7422.
33. T.E. Shrader, D.M. Crothers, Effects of DNA sequence and histone-histone interactions on nucleosome placement, *J. Mol. Biol.* 216 (1990) 69-84.
34. S. Godde, A.P. Wolffe, Nucleosome assembly on CTG triplet repeats, *J. Biol. Chem.* 271 (1996) 15222-15229.
35. S. Godde, S.U. Kass, M.C. Hirst, A.P. Wolffe, Nucleosome assembly on methylated CGG triplet repeats in the fragile X mental retardation gene 1 promoter, *J. Biol. Chem.* 271 (1996) 24325-24328.
36. Y. Wang, R. Gellibolian, M. Shimizu, R.D. Wells, J. Griffith, Long CCG triplet repeat blocks exclude nucleosomes: a possible mechanism for the nature of fragile sites in chromosomes, *J. Mol. Biol.* 263 (1996) 511-516.

37. H.R. Widlund, H. Cao, S. Simonsson, E. Magnusson, T. Simonsson, P.E. Nielsen, J.D. Kahn, D.M. Crothers, M. Kubista, Identification and characterization of genomic nucleosome-positioning sequences, *J. Mol. Biol.* 267 (1997) 807-817.
38. S. Cacchione, A.M. Cerone, M. Savino, In vitro low propensity to form nucleosomes of four telomeric sequences, *FEBS Lett.* 400 (1997) 37-41.
39. P.T. Lowary, J. Widom, New DNA sequence rules for high affinity binding to histone octamer and sequence directed nucleosome positioning, *J. Mol. Biol.* 276 (1998) 19-42.
40. L. Rossetti, S. Cacchione, M. Fuà, M. Savino, Nucleosome assembly on telomeric sequences, *Biochemistry* 37 (1998) 6727-6737.
41. M. Del Cornò, P. De Santis, B. Sampaolese, M. Savino, DNA superstructural features and nucleosomal organization on the two centromeres of *Kluyveromyces lactis* chromosome 1 and *Saccharomyces cerevisiae* chromosome 6 *FEBS Lett.* 431 (1998) 66-70.
42. H. Cao, H.R. Widlund, T. Simonsson, M. Kubista, TGGA repeats impair nucleosome formation, *J. Mol. Biol.* 281 (1998) 253-260.
43. D.J. Fitzgerald, J.N. Anderson, Unique translational positioning of nucleosomes on synthetic DNAs, *Nucleic Acids Res.* 26 (1998) 2526-2535.
44. D.J. Fitzgerald, J.N. Anderson, DNA distortion as a factor in nucleosome positioning, *J. Mol. Biol.* 293 (1999) 477-491.
45. Filesi, S. Cacchione, P. De Santis, L. Rossetti, M. Savino, The main role of the sequence-dependent DNA elasticity in determining the free energy of nucleosome formation on telomeric DNAs, *Biophys. Chem.* 83 (1999) 223-237.
46. S. Mattei, B. Sampaolese, P. De Santis, M. Savino, Nucleosome organization on *Kluyveromyces lactis* centromeric DNAs, *Biophys. Chem.* 97 (2002) 173-187.

47. S. Cacchione, J.L. Rodriguez, R. Mechelli, L. Franco, M. Savino, Acetylated nucleosome assembly on telomeric DNAs, *Biophys. Chem.* 104 (2003) 381-392.
48. A. Scipioni, S. Pisano, C. Anselmi, M. Savino, P. De Santis, Dual role of sequence-dependent DNA curvature in nucleosome stability: the critical test of highly bent *Crithidia fasciculata* DNA tract, *Biophys. Chem.* 107 (2004) 7-17.
49. J. Virstedt, T. Berge, R.M. Henderson, M.J. Waring, A.A. Travers, The influence of DNA stiffness upon nucleosome formation, *J. Struct. Biol.* 148 (2004) 2408-2418.
50. P. De Santis, A. Palleschi, M. Savino, A. Scipioni, A theoretical model of DNA curvature, *Biophys. Chem.* 32 (1988) 305-317.
51. A.A. Gorin, V.B. Zhurkin, W.K. Olson, B-DNA twisting correlates with base-pair morphology, *J. Mol. Biol.* 247 (1995) 34-48.
52. W.K. Olson, A.A. Gorin, X.-J. Lu, L.M. Hock, V.B. Zhurkin, DNA sequence-dependent deformability deduced from protein-DNA crystal complexes, *Proc. Natl. Acad. Sci. USA.* 95 (1998) 11163-11168.
53. M.J. Packer, M.P. Dauncey, C.A. Hunter, Sequence-dependent DNA Structure: Tetranucleotide Conformational Maps, *J. Mol. Biol.* 295 (2000) 85-103.
54. T.J. Bishop, Geometry of the Nucleosomal DNA Superhelix, *Biophys J.* 95 (2008) 1007-1017.
55. A. Valouev, J. Ichikawa, T. Tonthat, J. Stuart, S. Ranade, H. Peckham, K. Zeng, J.A. Malek, G. Costa, K. McKernan, A. Sidow, A. Fire, S.M. Johnson, A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning, *Genome Res.* 18 (2008) 1051-1063.
56. M. Caserta, E. Agricola, M. Churcher, E. Hiriart, L. Verdone, E. Di Mauro, A. Travers, A translational signature for nucleosome positioning in vivo, *Nucleic Acids Res.* 37 (2009) 5309-5321.

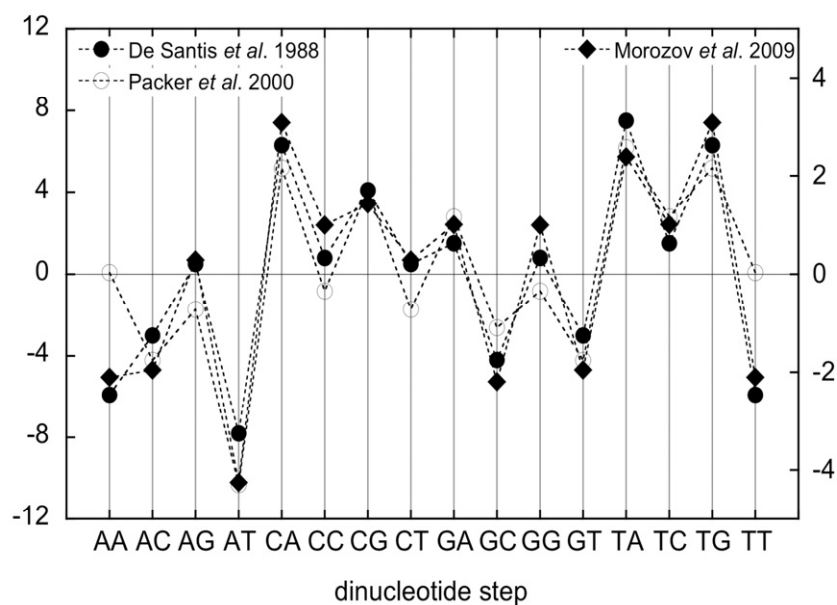
57. S.M. Johnson, F.J. Tan, H.L. McCullough, D.P. Riordan, A.Z. Fire, Flexibility and constraint in the nucleosome core landscape of *Caenorhabditis elegans* chromatin Genome Res. 16 (2006) 1505-1516.
58. R.D. Kornberg, Structure of chromatin. Annu. Rev. Biochem. 46 (1977) 931-954.
59. H.P. Hjelm, G.G. Kneale, P. Sabau, J.P. Baldwin, E.M. Bradbury, K. Ibel, Small angle neutron scattering studies of chromatin subunits in solution Cell 10 (1977) 139-151.
60. A. Klug, D. Rhodes, J. Smith, T.J. Finch, J.O. Thomas, A low resolution structure for the histone core of the nucleosome, Nature 287 (1980) 509 -516.
61. T.J. Richmond, J.T. Finch, B. Rushton, D. Rhodes, A. Klug, Structure of the nucleosome core particle at 7 Å resolution, Nature 311 (1984) 532-537.
62. K. Lüger, A.W. Mader, R.K. Richmond, D.F. Sargent, T.J. Richmond, Crystal structure of the nucleosome core particle at 2.8 Å resolution, Nature 389 (1997) 251-260.
63. D. Wang, N.B. Ulyanov, V.B. Zhurkin, Sequence-dependent Kink-and-Slide deformations of nucleosomal DNA facilitated by histone arginines bound in the minor groove, 27 (2010) 843-859.
64. F. Xu, W.K. Olson, DNA architecture, deformability, and nucleosome positioning, J. Biomol. Struct. Dyn. 27 (2010) 725-739.
65. C. Anselmi, P. De Santis, R. Paparcone, M. Savino, A. Scipioni, From the sequence to the superstructural properties of DNAs, Biophys. Chem. 95 (2002) 23-47.
66. P. De Santis, M. Fuà, M. Savino, C. Anselmi, G. Bocchinfuso, Sequence Dependent Circularization of DNAs: A Physical Model to Predict the DNA Sequence Dependent Propensity to Circularization and Its Changes in the Presence of Protein-Induced bending, J. Phys. Chem. 100 (1996) 9968-9976.



67. G. Zuccheri, A. Scipioni, V. Cavaliere, G. Gargiulo, M. Savino, P. De Santis, B. Samorì, Mapping the intrinsic curvature and the flexibility along the DNA chain, *Proc. Natl. Acad. Sci USA*, 98 (2001) 3074-3079.
68. A. Scipioni, C. Anselmi, G. Zuccheri, B. Samorì, P. De Santis, Sequence-Dependent DNA Curvature and Flexibility from Scanning Force Microscopy Images, *Biophys. J.* 83 (2002) 2408-2418.
69. C. Anselmi, G. Bocchinfuso, P. De Santis, M. Savino, A. Scipioni, Dual role of DNA intrinsic curvature and flexibility in determining nucleosome stability, *J. Mol. Biol.* 286 (1999) 1293-1301.
70. C. Anselmi, G. Bocchinfuso, P. De Santis, M. Savino, A. Scipioni, A Theoretical Model for the Prediction of Sequence-Dependent Nucleosome Thermodynamic Stability, *Biophys. J.* 79 (2000) 601-613.
71. M.R. Spiegel, *Fourier Analysis* (McGraw-Hill Book Company, New York, 1974).
72. C. Anselmi, G. Bocchinfuso, P. De Santis, M. Fuà, A. Scipioni, M. Savino, Statistical thermodynamic approach for evaluating the writhe transformations in circular DNAs, *J. Phys. Chem. B* 102 (1998) 5704-5714.
73. O. Gotoh, Y. Tagashira, Stabilities of nearest-neighbor doublets in double-helical DNA determined by fitting calculated melting profiles to observed profiles, *Biopolymers* 20 (1981) 1033-1042.
74. A.A. Travers, J.M. Thompson, An introduction to the mechanics of DNA, *Philos. Transact. A Math. Phys. Eng. Sci.* 362 (2004) 1265-1279.
75. H.R. Drew, R.E. Dickerson, Structure of a B-DNA dodecamer. III. Geometry of hydration, *J. Mol. Biol.* 151 (1981) 535-556.

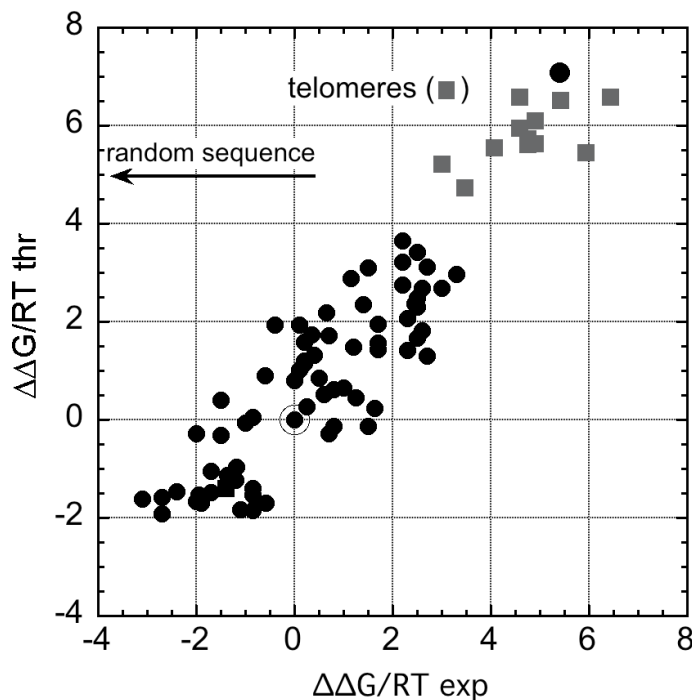
76. E. Liepinsh, G. Otting, K. Wuthrich, NMR observation of individual molecules of hydration water bound to DNA duplexes: direct evidence for a spine of hydration water present in aqueous solution, *Nucleic Acids Res.* 20 (1992) 6549-6553.
77. A. Scipioni, S. Morosetti, P. De Santis, A Statistical Thermodynamic Approach for Predicting the Sequence-Dependent Nucleosome Positioning Along Genomes, *Biopolymers* 91 (2009) 1143-1153.
78. A. Scipioni, S. Morosetti, P. De Santis, Prediction of Nucleosome Positioning in Genomes: Limits and Perspectives of Physical and Bioinformatic Approaches, *J. Biomol. Struct. Dyn.* 27 (2010) 747-764.
79. V. Miele, C. Vaillant, Y. d'Aubenton-Carafa, C. Thermes, T. Grange, DNA physical properties determine nucleosome occupancy from yeast to fly, *Nucleic Acid Res.* 36 (2008) 3746-3756.
80. J. Widom, A relationship between the helical twist of DNA and the ordered positioning of nucleosomes in all eukaryotic cells, *Proc. Natl Acad. Sci. USA* 89 (1992) 1095-1099.
81. A. Scipioni, G. Turchetti, S. Morosetti, P. De Santis, Geometrical, conformational and topological restraints in regular nucleosome compaction in chromatin, *Biophys. Chem.* 148 (2010) 56-47.
82. J.P. Wang, Y. Fondufe-Mittendorf, L. Xi, G.F. Tsai, E. Segal, J. Widom, Preferentially quantized linker DNA lengths in *Saccharomyces cerevisiae*, *PLoS Comput. Biol.* 4 (2008) e1000175.

**Fig. 1. Comparison among dinucleotide roll angles proposed by different authors.**



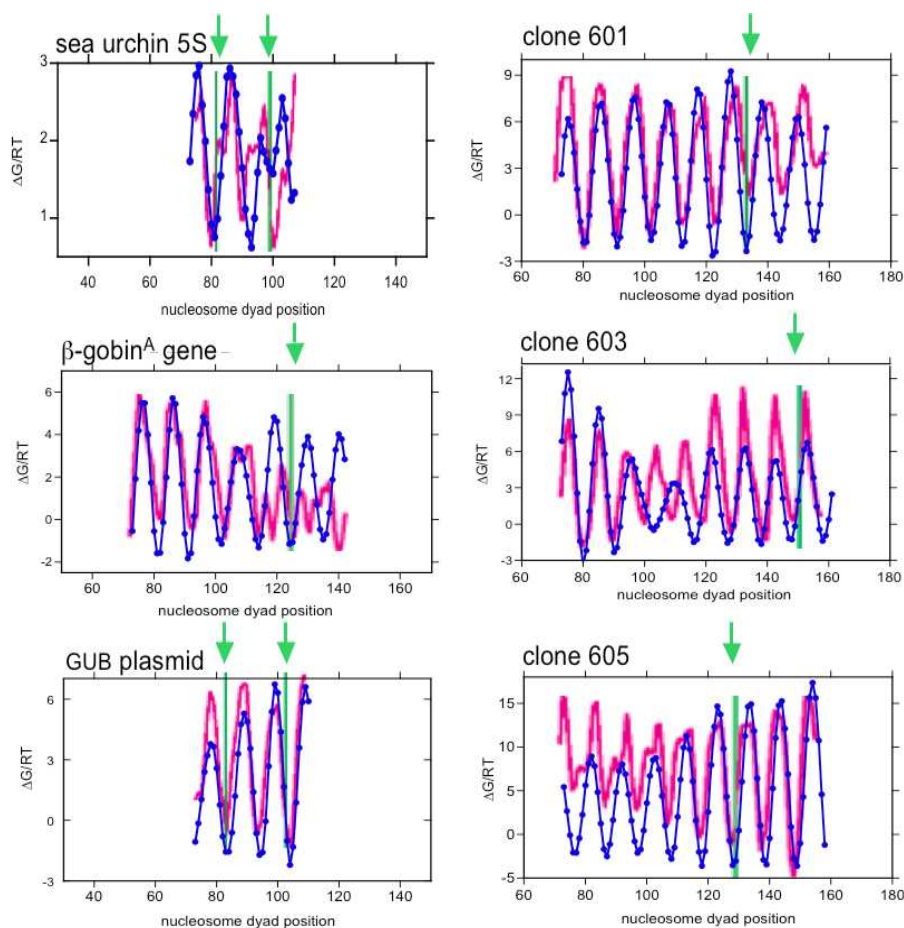
Comparison between the dinucleotide roll angles obtained on the basis of conformational energy calculations proposed by De Santis *et al.* (1988) [50] and later by Packer *et al.* (2000) [53], and those derived from X-ray structures of protein-DNA complexes by Morozov *et al.* (2009) [31].

**Fig. 2. Comparison between experimental and theoretical free energies.**



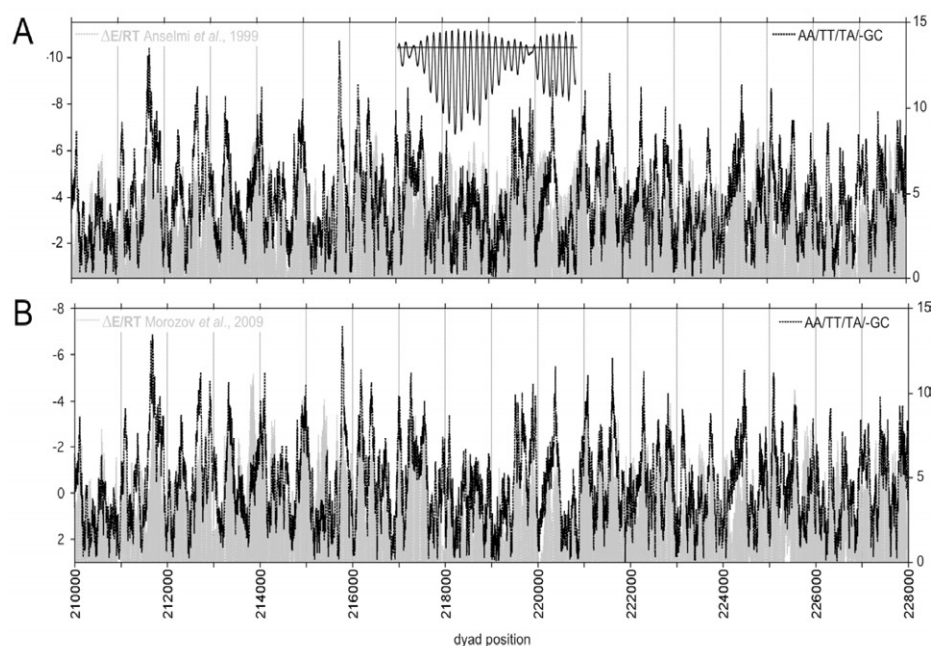
Comparison between experimental [32-49] and theoretical free energies of competitive nucleosome reconstitution of about a hundred DNA tracts different for sequence and length. The free energy values are referred to the TG-pentamer sequence [32], which is 5.0 RT lower than the random sequence (indicated as an arrow). The free energies of the telomeric sequences (■), the lowest histone affinity known so far, are highlighted [38,40,41,45-47].

**Fig. 3. Comparison of experimental and theoretical nucleosome positions for mononucleosomal sequences.**



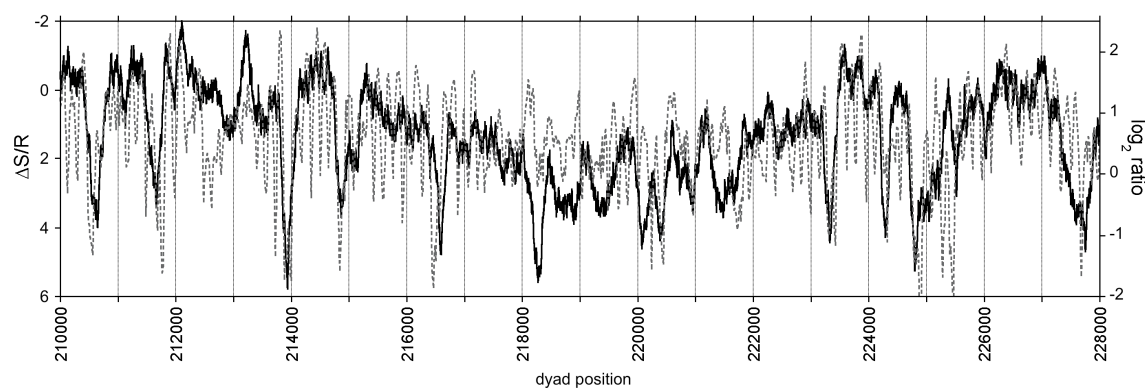
Predictions of *in vitro* nucleosome positions calculated according to the models by Morozov *et al.* [31] (red lines) and Anselmi *et al.* [69] (blue lines) for six mononucleosomal sequences (5S rRNA gene of sea urchin DNA, pGUB plasmid, chicken  $\beta$ -globin<sup>A</sup> gene, and three synthetic high-affinity sequences 601, 603 and 605 [39]). The vertical lines and arrows (green) indicate the experimentally known nucleosome positions. All DNA sequences are downloaded from the web site: <http://nucleosome.rockefeller.edu>.

**Fig. 4. Comparison between rotational distribution of AA/TT/TA/-GC steps vs the energy necessary to distort a DNA tract into the nucleosomal shape.**



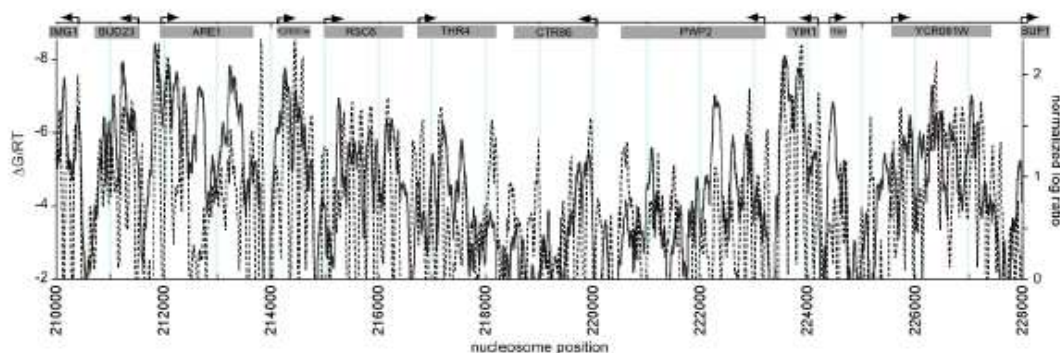
Comparison between rotational distribution of AA/TT/TA/-GC dinucleotide steps per nucleosome along a tract of yeast genome (210000-228000 bp of chromosome III) and the minimum distortion energy needed to transform the intrinsic curvature into the nucleosomal shape adopting Anselmi *et al.* [68] (A) and Morozov *et al.* [31] (B) internal dinucleotide parameters (roll, tilt, twist and elastic constants). The insert in A shows the periodical fluctuations of the energy function coherent with the B-DNA structure. The energy function is reported on a reverse scale.

**Fig. 5. Importance of entropy contribution in ruling nucleosome occupancy.**



Comparison between the profiles of experimental nucleosome occupancy, represented by the logarithm of the ratio between the nucleosomal and genomic DNA hybridization data [20], and the entropy contribution (reported on a reverse scale) calculated according to the last term of Eqs. (6) and (7) for a tract of the yeast genome (210000-228000 bp of chromosome III).

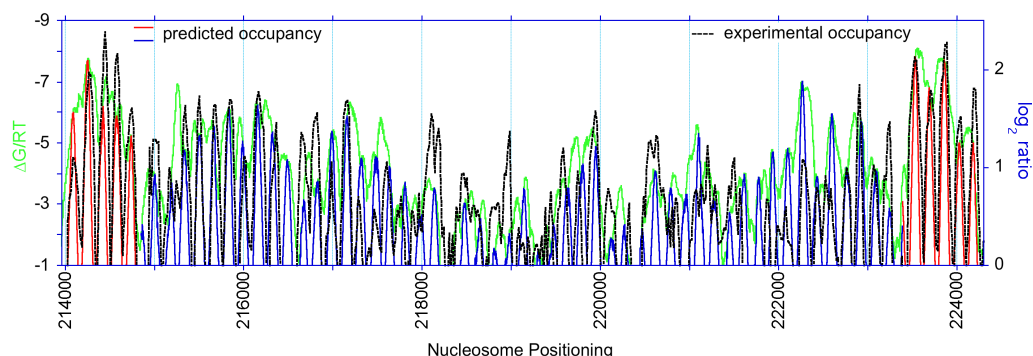
**Fig. 6. Free energy profiles versus the experimental nucleosome occupancy.**



Comparison between the free energy profile (reported on a reverse scale) and that obtained from the Boltzmann averaging according to Eq. (8), with that of the corresponding nucleosome occupancy ( $\log_2$  of the ratio between the nucleosomal and genomic DNA hybridization data) [20], for a tract of the yeast genome (210000-228000 bp of chromosome III). Genes along the sequence are identified in the grey boxes.

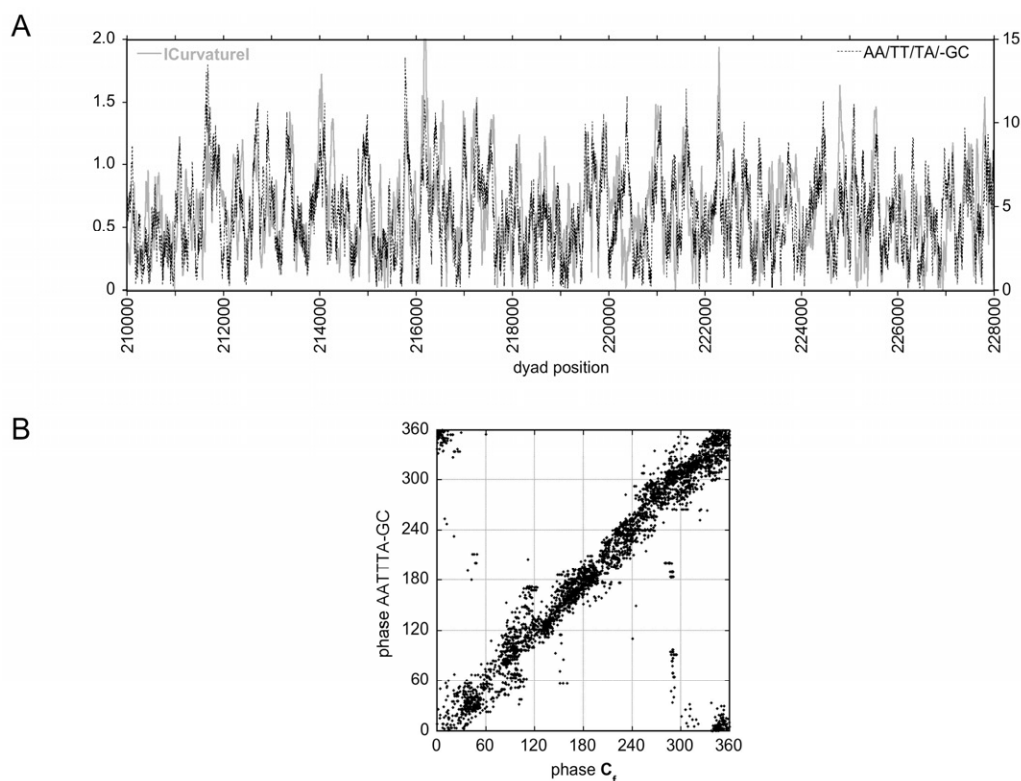


**Fig. 7. Free energy modulated with a periodical function versus experimental nucleosome occupancy.**



Comparison between the free energy profile (green) and those modulated with a periodical function according to Eq. (9) (red and blue) and the experimental nucleosome occupancy *in vivo* [20] (dashed line) along a tract of yeast genome (214000-225000 bp of chromosome III). The free energy profiles are reported on a reverse scale. The different colors indicate opposite phases adopted. The profiles are characterized by successive blocks of nucleosomes with opposite phases as plausibly related to their compaction in chromatin fibers.

**Fig. 8. Modulus and phase of the intrinsic curvature compared with rotational recurrence of AA/TT/TA/-GC steps.**



(A) Comparison between the intrinsic curvature modulus and the value of rotational distribution of AA/TT/TA in phase and GC out of phase in recurrent 146-bp sequences along a yeast genome tract (210000-228000 bp of chromosome III). (B) Comparison between the phases of curvature ( $C_f$ ) and those relative to the rotational distribution of AA/TT/TA/-GC steps.