



DEPARTMENT of PHYSICS
Doctoral School in Physics

**A CONVERGENT AND MULTISCALE
ASSESSMENT OF DNA DAMAGE BY
PARTICLE RADIATION**

Advisor

Prof. Gianluca Lattanzi

Co-advisor

Dr. Emanuele Scifoni

PhD Candidate

Lorenzo Petrolli

- TRENTO, APRIL 2022 -

Abstract

The mutation/deletion of the hereditary material in the cell nuclei is a chronic biochemical hazard; in fact, nuclear DNA faces tens of lesions from metabolic intermediates, hydrolytic reactions and external vectors a minute. The canonical lesions of DNA involve the DNA backbone as well as the nucleic bases, and are mostly associated with elemental chemical aberrations. However, the radiation field from a beam of accelerated ions accounts for a critical streak of local inelastic collisions in cell nuclei, whereby the radiation massively excites/oxidizes the DNA molecule (*direct* effect) and releases toxic radiolytic radicals (*indirect* effect), and achieves exotic clusters of elemental lesions.

Double strand breaks (DSB), i.e. the twofold cleft of the DNA backbone, are hazardous fractures of the chromatin fold associated with the radiation field and underlie cytotoxic and aberrant chromosomal outcomes. Eukaryotic cells, however, rejoin double strand breaks via an apt enzymatic machinery, or the DDR. Prior to the deployment of enzymatic effectors, host enzyme *sensors* engage the DSB termini in reversible supramolecular assemblies, which requires that the fractured DNA moieties be fully exposed.

Eukaryotic cells stock the nuclear DNA as a hierarchically layered framework, or *chromatin*. At the basis of chromatin are extended arrays of supramolecular assemblies, or **nucleosomes**, where a 147-nucleotides stretch of DNA winds a core histone octamer in 1.67 superhelical gyres. In chromatin, tens of thousands of nucleosome beads are threaded and intercalated by linear stretches of 20 to 70 nucleotides, or DNA *linkers*, and thus folded into mesoscale fibers. The histone octamer of nucleosomes involves a core bundle of interwoven α -helices, which the DNA holds onto, and erratic, misfolded terminal *tails*, which extend from the nucleosome core and between adjacent nucleosomes and chromatin fibers.

Nucleosomes take on a broad mechanical and biochemical role in the chromatin framework; in fact, both the helix fold and the tails describe an active contact network, whereby the access to the DNA and the recruitment of chromatin factors are modulated locally via tailored mutations. However, the earliest mechanical role of nucleosomes in the context of DNA lesions is

controversial.

The microdosimetric theoretical frameworks (on the noxious effects of the radiation field in cells) benefit from the *in silico* assessments of the distribution of lethal DNA lesions at the micro- and nanoscale. In this concern, advanced *Monte-Carlo track structure* (MCTS) codes define as double strand breaks any closely associated, local inelastic collisions of the radiation field with the DNA backbone (to a certain likelihood), by means of coarse distance criteria. However, the diverse DSB *motifs* (i.e. where the twofold cleft of the DNA backbone is distanced by zero to several nucleotides) account for a different dynamics of the DNA. Moreover, we reckon that in the absence of excess external stimuli, short-distanced DSB motifs alone elicit the thermal fracture of the nucleosome fold, as observed earlier *in silico*.

In this thesis elaborate, we tackle the *in silico* assessment of the earliest frames of the thermal dissociation of **blunt** double strand breaks, and of the local strain exerted by DSBs on the DNA helix and on the nucleosome fold. The outline of the thesis is as follows.

Chapter 1 is devoted to a brief introduction to the DNA helix and the chromatin fold. In **Chapter 2**, we broadly describe the lesions from the irradiation of DNA and the microdosimetric theoretical framework of cell survival.

In **Chapter 3**, we carry out a systematic MCTS assessment of the early *spectra* of double strand breaks over a nucleosome tetramer (chromatin-like) framework, enforced by accelerated hydrogen ions (H^+) at a Bragg peak-relevant energy range (500 keV - 5 MeV). We account for the overall *direct* effect of the radiation field, i.e. the local oxidation/excitation of DNA from the inelastic collisions of the radiation carriers. We infer that event-by-event MCTS cannot discern the local morphological variance of the chromatin layout, and that track structure assessments at the nanometric scale are likely biased from statistical artifacts. We thus conclude that a Poisson fit describes the spectrum of double strand break motifs in a chromatin-like framework and, notably, we observe a bias towards short-distanced, *staggered* DSBs.

Chapter 4 is dedicated to the theoretical framework underneath *molecular dynamics* (MD) and the *PyInteraph* workflow, for the dynamical extraction of connected networks across macromolecular frameworks from MD trajectories.

Prior to the MD assessment of blunt double strand breaks, in **Chapter 5** we benchmark an extension of the *PyInteraph* toolkit to treat nucleic acids. We employ the nucleosome fold as a case scenario, and thus broadly infer on the mechanical and allosteric role of the terminal tails of histones in the nucleosome fold.

In **Chapter 6**, we thus enforce *blunt* double strand breaks over both short DNA linkers and the nucleosome fold, and evolve their subsequent dynamics via classical, all-atom molecular dynamics. We observe the fracture of the DNA linkers, as well as a notable variety in the dislocation dynamics of the blunt π -stack interface. Moreover, we assess mild nuances in the behavior of the lesion sites, modulated by the choice of the DNA termini.

Conversely, the nucleosome fold exerts a kinetic barrier towards the thermal dislocation of DSB lesions (which we reckon enforces a “latency lapse” of the enzymatic machinery), in spite of the excess strain of the wound DNA layout, mediated by the contact interface between DNA and the histone fold.

In conclusion, we remark that in the absence of further data from assessments *in vitro* and *in vivo*, the inferences about the thermal and mechanical resilience of the DNA lesions are as reliable as the molecular mechanics force fields underneath. In fact, it is debated whether all-atom force fields and water models overestimate the forces between π -stack contacts and over-stabilize folded layouts.

We reckon that diverse layers of the chromatin fold are apt for MCTS/MD assessments of DNA lesions, and that further efforts shall be attempted to establish a shared, multi-scale framework.

Contents

1	DNA and chromatin	1
1.1	The DNA helix	1
1.2	The chromatin fiber	3
1.3	The nucleosome fold	4
2	DNA lesions and the radiation field	7
2.1	The DDR	8
2.2	The theoretical framework of cell survival	9
2.3	The RBE	11
2.3.1	Microdosimetric kinetic model	12
2.3.2	Local effect model	13
2.4	Geant4-DNA and Monte-Carlo track structure codes	15
3	The radiation track at the nucleosome scale	19
3.1	Geant4-DNA and the nucleosome tetramer: Materials and methods	20
3.2	MCTS assessments at the nanometric scale are biased from statistical artifacts	23
3.2.1	Shannon’s entropy as a bias estimator	25
3.3	The double strand break distance score	27
3.4	Conclusive remarks	28
4	The theoretical framework of molecular dynamics	31
4.1	The thermodynamic ensemble	31
4.2	Molecular dynamics	32
4.2.1	Non-bonded forces	33

4.2.2	Thermostats and barostats	34
4.2.3	The CHARMM force field	36
4.3	The PyInteraph workflow	37
5	MD assessment of the nucleosome interface	39
5.1	The nucleosome framework: Materials and methods	39
5.2	Extension of the PyInteraph workflow to the nucleosome in- terface	41
5.2.1	Hierarchical clusters analysis	43
5.3	Dynamics of the nucleosome interface	44
5.3.1	The erratic dynamics of the histone tails	45
5.3.2	RMS fluctuations of the DNA fold	49
5.3.3	The canonical contact sites are affected by the imbal- ance of the nucleosome fold	52
5.3.4	Histone tails act as reversible allosteric switches of the nucleosome framework	56
5.4	Conclusive remarks	65
6	MD assessment of blunt double strand breaks	67
6.1	The DSB scenario in molecular dynamics	68
6.2	Blunt double strand breaks: Materials and methods	69
6.2.1	Local assessment of the residual contact interface at a DNA linker	70
6.3	The behavior of a blunt DSB interface at a DNA linker	72
6.3.1	Intercalated interfaces are artificial distortions of the DNA helix	74
6.3.2	Blunt DSB interfaces crack at 310 K	76
6.3.3	5' termini modulate the residual forces between blunt DNA moieties	76
6.4	The nucleosome fold is resilient towards the thermal fracture of double strand breaks	79
6.5	Conclusive remarks	87
	Discussion and conclusive remarks	88

Chapter 1

DNA and chromatin

DNA neither cares nor knows. DNA just is. And we dance to its music -

Richard Dawkins, British evolutionary biologist

1.1 The DNA helix

Life spawns across all realms of nature upon a shared, narrow alphabet of elemental blocks - all in all, a few dozens of amino acids, nucleotides, carbohydrates and lipids. Within this “limited” variety, **nucleotides** - and their macromolecular assemblies, or **nucleic acids** (DNA, RNA) - are the backbone of all major (metabolic, structural, hereditary) activities in the life-cycle of cells.

Akin to an amino acid, a nucleotide associates i) a *backbone* moiety, which combines a *carbohydrate*¹ and one to a few *phosphate* beads, and ii) a **nucleic base**, i.e. one of five aromatic heterocycles and derivatives of either *purine* (*adenine*, *guanine*) or *pyrimidine* (*cytosine*, *thymine*, *uracil*).

In nucleic acids, the nucleotides are threaded via twofold, phosphoesteric bonds that join the 5' and 3' hydroxyl termini² of adjacent monomers. In cells, the sequential assemble of nucleotides as “elemental letters” overall defines its hereditary content. Indeed, by the *central dogma of molecular biology*, nucleic acids encode all instructions that are recursively and sub-

¹It is (a derivative of) *ribose* in nucleic acids and metabolic vectors, such as ATP

²A nomenclature is associated to nucleotides whereby atoms of the ribose moiety are indexed by a ' mark.

sequently transcribed from DNA to RNA, and translated into sequences of amino acids, which mediate the balanced biosynthetical framework of all the factors involved with a cell life-cycle.

DNA (*deoxyribonucleic acid*)³ is the helical, supramolecular assemble of two *strands* of nucleotides, where nucleic bases are coupled via a conserved, binary scheme that strictly associates ADE with THY and CYT with GUA - or the *Watson-Crick rules*. This double-helical framework is right-handed in the $5' \Rightarrow 3'$ verse of each strand, and revolves with a stride of ~ 10 nucleotides (3.4 nanometers) about a core helical axis. Moreover, both the ADE \cdots THY and CYT \cdots GUA **base pairs**⁴ (BP) enforce a consistent distance of ~ 1.1 nanometers between C1' ribose atoms of the two strands, and are but slightly tilted ($\sim 6^\circ$) from the helical cross-section, so that the double helix of DNA is associated with a constant width of about 2 nanometers.

About the helical axis of DNA, twofold and threefold hydrogen bonds hold the *Watson-Crick edges* of each of the ADE \cdots THY and CYT \cdots GUA base pairs respectively. Thus, two hollow grooves flank the DNA helix at either side of the core network of hydrogen bonds, and expose the *Hoogsteen* (**major groove**) and *sugar edges* (**minor groove**) of the aromatic nucleic bases (FIG. 1.1). Yet, it is not the supramolecular network of hydrogen bonds alone to elicit the double-helical layout of DNA; after all, the nucleic bases establish hydrogen bonds with nucleotides and water alike. In fact:

- nucleic acids bear a physiological electrostatic imbalance, which is shielded effectively by mono- and divalent salts;
- adjacent nucleotides coalesce by means of **π -stack** contacts, which are mediated by the exclusion of intercalated water molecules between nucleic bases and van der Waals forces.

These synergistic electrostatic and hydrophobic forces, and the steric hin-

³The one that follows is a description of the canonical, or B-conformation, of DNA, and is a virtually idealized depiction of the helical layout. In fact, the local structural variance of DNA is modulated by the sequence of the nucleic bases, the state of hydration, and the class and concentration of the electrolytes - up to the existence of alternate (A, Z) conformations of DNA, not described here.

⁴The \cdots mark is associated with hydrogen bonded moieties.

drance of the nucleotides, restrain the torsional freedom of the double helix, so that DNA is stiff and scarcely flexible overall.

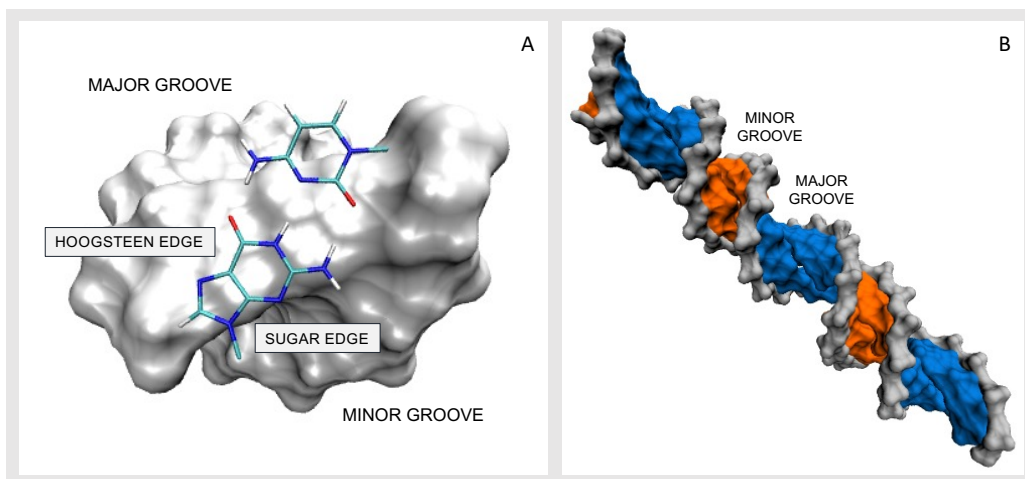


Figure 1.1: (A) The threefold hydrogen bond holds the *Watson-Crick* edges of CYT and GUA about the core helical axis of DNA. The *Hoogsteen* and *sugar* edges of the nucleic bases are exposed at the hollow major and minor grooves of DNA (B) respectively.

1.2 The chromatin fiber

The whole chromosomal DNA in the nucleus of a diploid human cell (3.2 billion BP) unravels to a contour size little short of one meter. Yet, DNA is stocked rationally within micrometer-sized, crowded nuclei via an elaborate, hierarchical and hybrid framework, or **chromatin**.

At the basis of chromatin is the **nucleosome**, a discoid bundle of an octamer of small (102 - 135 amino acids) **histones**, wound by a short DNA stretch (145 - 147 BP) in 1.67 left-handed wraps (section 1.3). In fact, a molecule of DNA threads over dozens of thousands of nucleosome “beads” within a *chromosome*⁵ to achieve a 10-nanometer fiber, where a linear texture of nucleosomes is intercalated with variably sized stretches of DNA, or **linkers**. The layout of the subsequent hierarchical layer of chromatin has been a matter for controversy [1]. In fact, the ever-accepted idea that the 10-nanometer

⁵*Chromosomes* are but *chromatin territories*, which coalesce at cell division.

nucleosome fiber coalesces consistently into a 30-nanometer fiber *in nucleo*⁶ [2] does not tally with a depiction of nuclear chromatin as an active substrate:

- *silent* chromatin has been associated with evenly spaced, ill-located nucleosome beads, whereas *active*, transcribed loci exhibit a chaotic nucleosome layout, which is nevertheless often phased downstream of *transcription start sites* and regulatory sequences (such as transcription *enhancers*). The dynamic modulation of the nucleosome occupancy of DNA is mediated both by sequence affinity and (overridden) by ATP-fueled *chromatin remodelers*.
- chromatin itself describes a continuum of local, reversible functional states, modulated by *histone variants* and chemical modifications (PTM), and *linker histones* - which both enhance/shield inter-nucleosome (electrostatic) contacts and recruit chromatin factors *in situ*.

The actual chromatin fiber *in nucleo* is thus reckoned to be aptly described by a dynamical variety of local motifs and elemental clusters of nucleosomes⁷.

1.3 The nucleosome fold

The **nucleosome core particle** (NCP) is the invariant, elemental block of the hierarchical framework of chromatin, across the eukaryotic domain⁸. In the nucleosome fold, a stretch of double stranded DNA (147 BP) winds about a discoid, helical spool of **histones** in $1\frac{3}{4}$ left-handed rounds. This framework involves four conserved histones (H3, H4, H2A and H2B), which are replicated about a twofold rotational axis that crosses the nucleosome at the *dyad* - or superhelical location⁹ (SHL) 0; thus, two juxtaposed layers of the

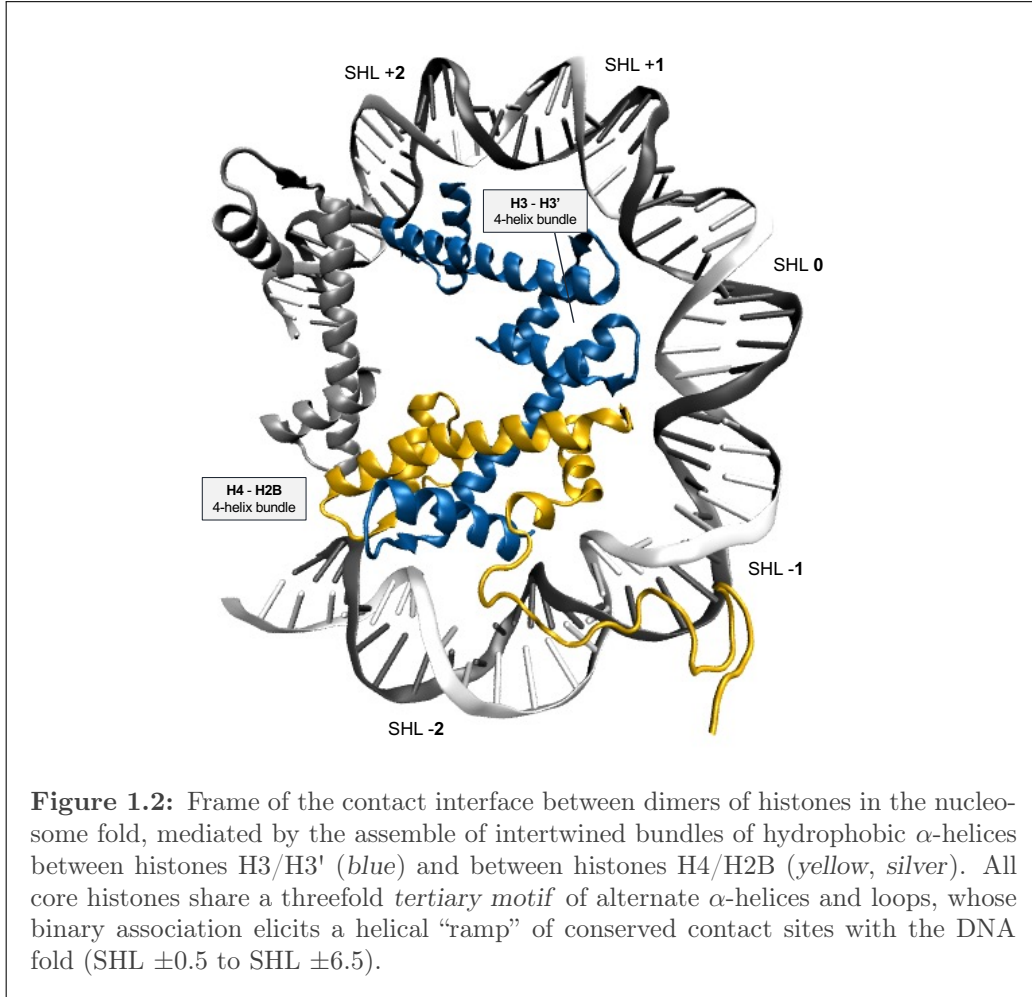
⁶The 30-nanometer fiber layout of chromatin was achieved *in vitro* from reconstituted nucleosomes about tandem 5S or 601 DNA stretches (associated with an elevated nucleosome affinity), or from dilution of native chromatin, at a low concentration of chromatin and of divalent salt.

⁷The subsequent hierarchical levels of the chromatin framework, that is the tertiary fold of the chromatin fiber, will not be dwelt upon here.

⁸We refer the reader to Cutter and Hayes 2015 as an effective review of the structure and dynamics of nucleosomes [3].

⁹The *superhelical location* (SHL) nomenclature of nucleosomal DNA defines the distance from the dyad site in effective units/helical twists. The *dyad*, or SHL 0, lies at

four core histones are pseudo-symmetrically related in the octameric fold of the NCP.



In fact, the supramolecular association of histones at physiological salt activity (~ 150 mM) accounts for the stable assemble of a (H3/H4)₂ tetramer and H2A/H2B dimers, whereas the octameric fold is achieved either via the further association of DNA or at ~ 2 M salt (both of which shield the excess electrostatic imbalance exerted by the basic residues at the amino-termini of histones - *vide infra*). H2A/H2B dimers thus bind either flank of the

an *inward major groove* of the DNA, that is where the groove faces the NCP. Thus, all subsequent inward major/minor grooves are associated with integral/semi-integral SHL intervals respectively (FIG. 1.2).

(H3/H4)₂ tetramer via hydrophobic “bundles” of core α -helical motifs (which stabilize the (H3/H4)₂ tetramer alike - FIG. 1.2), and establish further ancillary contacts therein.

In spite of low sequence homology, all histones in the NCP share a core *tertiary motif*, whereby two short ($\alpha 1$, $\alpha 3$) helices bracket an extended ($\alpha 2$) helix via coiled L1 and L2 loops, which overall define the **helix fold** domain. Moreover, the binary association of $\alpha 1$ -helices and of L1/L2 loops in the four histone dimers elicits a helical “ramp” of conserved, evenly spaced contact sites between the helix fold and DNA, hereafter **canonical sites**, whereby each histone dimer holds a stretch of DNA of about 30 BP (section 5.3.3).

Structure-less amino-termini, enriched in basic ARG/LYS residues, thus leak from the helix fold domain and define the **tails** of the core histones (section 5.3.1). Histone tails amount to 25-30% of the overall mass of the core histones and are hosts to reversible PTMs, whereby nucleosomes actively recruit the diverse chromatin factors *in situ* and modulate the local chromatin framework [4–6].

DNA achieves an overly bent and flattish layout¹⁰, and is overall under-twisted about the NCP (10.2 v. 10.5 BP/helical twist of free DNA), which is accounted for by the recursive texture of conserved canonical sites and by the contact interface with the histone fold.

The thermal dynamics of nucleosomal DNA at the short (microsecond, millisecond) timescale is described by i) the reversible, *impromptu* detachment of the outer DNA stretches, i.e. at the entrance/exit sites of DNA into the NCP, from the histone core, or DNA *breathing*, and ii) the relaxation/translocation of local *twist-defective* states. These modes underlie the *unwrapping* and *sliding* of the nucleosome fold at a slower scale, whereby nuclear DNA is made accessible and/or translocated across loci. Likewise, nucleosomes may be forcibly evicted, relocated and unraveled by means of ATP-fueled enzymes, such as chromatin factors and remodelers.

¹⁰The nucleosomal DNA is associated with excess *roll* and mild excess *slide* of adjacent base steps v. free DNA.

Chapter 2

DNA lesions and the radiation field

The biochemical hazard that is associated with the non-reversible mutation and/or deletion of nucleic material in cells is ubiquitous, and arises from both internal - such as metabolic intermediates - and external vectors.

A radiation field, from a beam of accelerated ions or an electromagnetic (EM) ray, strikes a live cell nucleus and thereby collides stochastically with the water medium and with DNA. Two toxic outcomes are thus bound to follow:

- the electronic excitation or oxidation of DNA, which accounts for a local deterioration of the double helix, or the **direct** effect of radiation;
- about the first solvation shell of DNA, the radiation field spawns reactive species from the radiolytical decay of water molecules (such as hydrogen and hydroxyl radicals, and solvated electrons), which mediate the further chemical lysis/modification of the DNA molecule, or the **indirect** effect of radiation.

The whole of DNA is exposed to the radiation field, so that collectively a plethora of noxious DNA lesions is achieved by i) the extraction/deletion of nucleic bases, ii) the manifold chemical and electrochemical modifications of the nucleic bases as well as of the DNA backbone, iii) the (twofold) covalent fracture of the DNA backbone, or (**double**) **strand break (DSB)**.

Moreover, the radiation field from such vectors as accelerated (heavy) ions at elevated *linear energy transfer* (LET) rates is associated with a massive release of dose¹ over localized volumes of a cell nucleus, as well as a critical amount of electrons and radicals from the subsequent collision cascade, whereby “exotic”, hardly reversible clusters of elemental DNA lesions are achieved (*vide infra*).

2.1 The DDR

Yet, live cells tolerate most classes of noxious DNA modifications and thousands of lesions a nucleus. In fact, (eukaryotic) cells enable substrate-selective recovery cascades, which are set off by an enzymatic framework that senses the DNA lesions and restores the double helical layout - collectively involved with the *DNA damage response*, or **DDR** [7]. Both the activity and the outcome of the DDR are modulated by the status of the cell and of the cell-cycle [8], which may imply that a cell i) is able/unable to recover from a DNA lesion or an aberration of the chromatin framework, and ii) enters an alternate apoptotic (death) or senescence track, or bears toxic DNA mutations.

It has been widely acknowledged that *double strand breaks* of DNA are a detrimental outcome of the radiation field in cells [9]. While DSBs are not inherently critical, the flawed recovery from a DSB lesion underlies the chances that aberration and mutations of the chromatin fold arise, via the deletion of ssDNA ends and/or the misjunction of native non-consecutive DNA termini. In fact, eukaryotic cells deploy two major, competitive cascades to tackle DSBs:

- in **non-homologous end joining** (NHEJ), the two fractured DNA moieties are involved with a short-range synapse and (optionally) subject to manifold takes of synthesis/lysis, whereby the DNA termini are eventually ligatable (this may be established by limited microhomology). NHEJ may thus account for a flawed misjunction of double strand break lesions - yet, it is reckoned to be the default choice in the eukaryotic cells. Auxiliary

¹The *dose* defines the energy absorption in the unit mass [$\text{J} \cdot \text{kg}^{-1}$], whereas the LET is associated with the local release of energy about a radiation track, within a dx layer of the medium [$\text{keV} \cdot \mu\text{m}^{-1}$].

sub-routes that require extended DNA resection (lysis) and microhomology between the annealed DNA strands may take over as the canonical NHEJ track fails [10].

- in **homologous recombination** (HR) - which is restricted to the S and G2 (pre-mitotic) phases of the cell-cycle - majorly resected 3' ssDNA overhangs are loaded with recombinase enzymes and a search for the homologous DNA sequence over a sister chromatid (which new DNA is synthesized upon) is commenced. Provided that the template chromatid is not corrupted, HR faithfully restores the native chromatin sequence.

The balance between NHEJ and HR - and alternate minor routes - is mediated within a faceted framework that involves the cell-line, the status of the cell-cycle, the expression of DDR enzymes and co-factors, the local layout of the chromatin fold and texture of histone PTMs, as well as the amount and class of the DNA lesions.

In spite of - and on account of - their inherent cytotoxicity, radiations have been employed as an alternate and/or in synergy with the chemical therapy of cancerous tissues. The controlled irradiation by accelerated ions is a state-of-the-art and convenient route to convey exceptional spurs of energy to deep cancerous tissues, thereby mostly sparing healthy tissues from the toxic by-effects associated with conventional (x-rays) therapies.

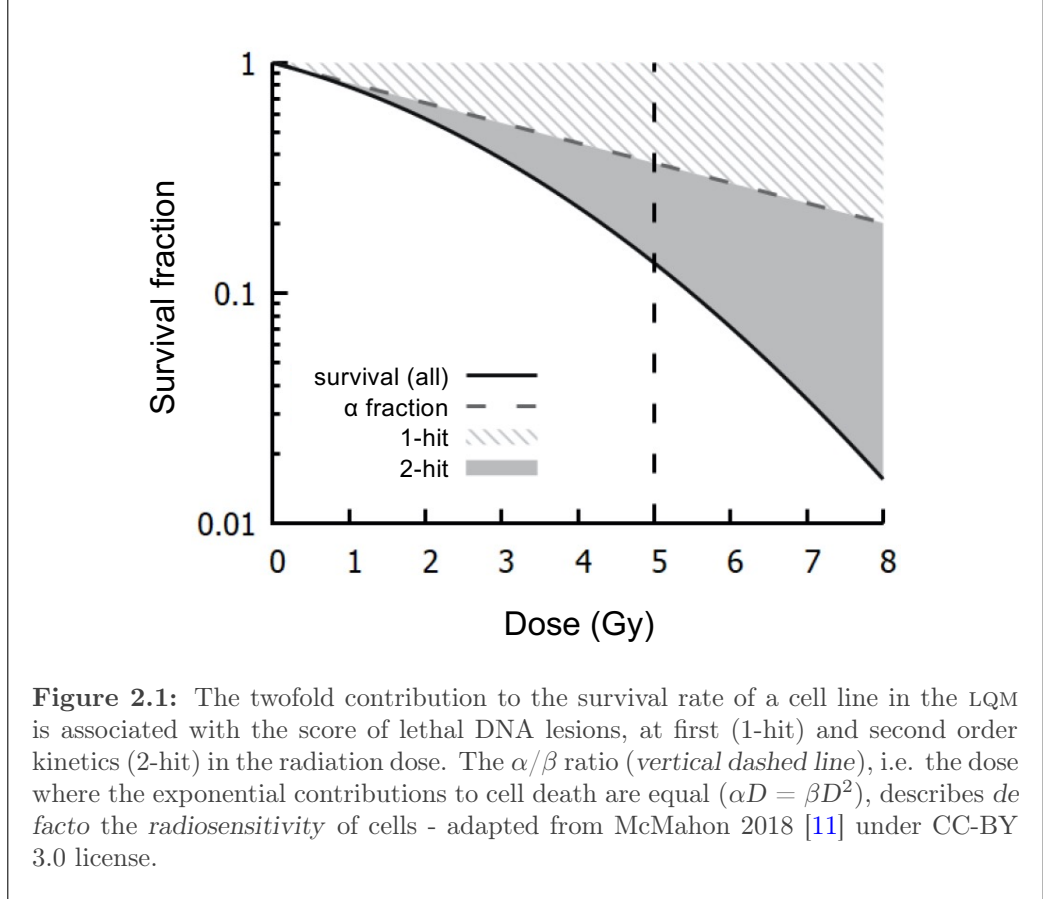
2.2 The theoretical framework of cell survival

The survival rate of cell lines exposed *in vitro* to a radiation field is defined as the ratio of cells to outlive a radiation dose, over the total of irradiated cells. It has been described - over fifty years now - by an exponential fit of the radiation dose as:

$$S = e^{-\alpha D - \beta D^2} \quad (2.1)$$

or the *Linear-Quadratic model* (LQM) [11], with α and β associated with the inherent *radiosensitivity* of a cell line. In fact, the radiosensitivity of cells is described *de facto* by the curvature of the survival fall-off and by the α/β

ratio, i.e. the dose where the twofold exponential contributions to cell death - at first and second order in the dose - are equal. (FIG. 2.1).



While α and β are strictly empirical factors, the theoretical/mechanistic framework that underlies their behavior is faceted [11]. In fact, it was soon established that critical DNA lesions accounted for the basic noxious action of the radiation field in live tissues, such as cell annihilation and the mutations/aberrations of the chromatin fold. A layered scenario thus arose, as it was further observed that a cell radiosensitivity is modulated by both *intrinsic* and *extrinsic* factors, which involve: i) the status of the cell cycle, and thus the access to the HR cascade to rejoin double strand breaks; ii) the (excess) score of non-reversible DNA lesions and/or of non-viable *chromosomal*

*aberrations*²; iii) the balance between alternate cell death routes, and with cell *senescence*; iv) the carriers, LET and fractionation rates of the radiation field; v) the levels of dissolved oxygen in the tissue, and further enhancement factors driven by the cellular micro-environment, such as the *bystander effect*³.

Nevertheless, these effects overall fall loosely within the LQM framework and accounts for its broad use in the clinical scenario [11].

2.3 The RBE

The earliest theoretical framework of the cell survival via the LQM relied upon extensive datasets from x-rays irradiations *in vitro*; thus, a revision of the basic framework was invoked with the advance of accelerated ion beams in the clinical facilities. In fact, the dose is delivered by x-rays and accelerated ions with remarkable differences at the macro- and nanoscale (FIG. 2.2):

- **x-rays** are associated with a diffuse radiation field and a slow, exponential fall-off of the dose within the medium. Hence, a massive flux of x-rays shall strike a cell nucleus to achieve an effective dose release (about the order of 1 Gy/nucleus) and sparse layout of lethal DNA lesions;
- the local release of dose from a beam of **accelerated ions** is coaxial with the core of the radiation *track* and radially correlated within a few nanometers⁴. At the macroscale, accelerated ions unleash a major dose fraction within a localized volume and depth in the cell, or the *Bragg peak*, where critical, exotic DNA lesions are achieved.

²The radiation-enforced aberrations of chromosomes are (mostly) associated with the structural abnormalities that arise via the flawed misjunction of DSBs, such as truly defective (dicentric, acentric) and/or mutated chromosomes, from the translocation/deletion of loci. Notably, “viable” lesions of the chromatin fold (which contribute to the survival rate of the cell line) may bear cancerous DNA mutations [12].

³The *bystander effect* describes the lethal/detrimental outcomes observed in cells that have not been struck by the radiation directly, and “conveyed” from closely associated, irradiated cells [13].

⁴However, *delta* electrons may branch radially, a few micrometers from the core of the radiation track.

It was soon established that the local variance of LET across a radiation track (from the loss of momentum of the radiation in the medium) underlay a modulation from the canonical survival fall-off of the LQM framework - hence, the definition of the *relative biological effectiveness* (RBE) of a radiation carrier r , as:

$$\begin{aligned} \text{RBE} &= \frac{D_x}{D_r} \Big|_{\text{isoeffective}} = \\ &= -\frac{1}{2D_r} \left(\frac{\alpha}{\beta} \right)_x + D_r^{-1} \sqrt{\frac{1}{4} \left(\frac{\alpha}{\beta} \right)_x^2 + \frac{\alpha_r}{\alpha_x} \left(\frac{\alpha}{\beta} \right)_x D_r + \frac{\beta_r}{\beta_x} D_r^2}, \end{aligned} \quad (2.2)$$

where the x and r subscripts define the reference (taken *de facto* as the x-rays) and the new carrier, and the D_x/D_r ratio involves *isoeffective doses*, i.e. doses of the two radiations that achieve equal survival rates. Thus, variably empirical formulations were elaborated that relate the (α, β) radiosensitivity factors of cells and the LET of the radiation [14]. As for light carriers such as hydrogen ions (H^+), it has been remarked that such empirical fits are as reliable as the irradiation datasets underneath, and have overall achieved a variety of RBE estimates⁵. However, to describe the RBE of heavier ions, an advanced theoretical framework was elaborated, and dedicated treatment codes at clinical facilities in Japan and Germany now rely on either of two semi-mechanistic models [18–20].

2.3.1 Microdosimetric kinetic model

The *microdosimetric kinetic model* (MKM) [21, 22] is derived from the formalism of the *dual radiation action* of the clinical ion beam [23], where:

- the radiation field accounts for either *lethal* or *sub-lethal* lesions of sensitive DNA volumes in cell nuclei. Sub-lethal lesions thus either i) drift/evolve, and achieve further lethal lesions, ii) are rejoined, or iii) are conserved (harmless).

⁵In fact, the case of hydrogen ions is controversial: Hydrogen ions have been associated with a fixed, conservative RBE = 1.1 in the clinical scenario, based on assessments *in vivo* and clinical data [15]. However, this accounts for a mild misestimate of the radiation RBE about a tumour volume, and is critical at the distal dose fall-off [16, 17].

- cell nuclei are subdivided into local *domains*, which are inactivated by lethal DNA lesions, to lead to cell death.

The earliest formulations of the MKM thus defined the survival rates of cell lines as:

$$\ln S = -\langle w_n \rangle = -(\alpha_0 + z_{1D}\beta)D - \beta D^2 \quad (2.3)$$

with w_n the score of lethal lesions in the cell nuclei; α_0 and β are LQM-like factors, and are defined via the (kinetic, mechanistic) rates of the dual radiation action framework; and:

$$z_{1D} = \langle z_1^2 \rangle / \langle z_1 \rangle = \frac{\int z_1^2 f(z) dz}{\int z_1 f(z) dz} \quad (2.4)$$

describes the statistical dose release at a cell sub-domain by one irradiation event (i.e. a radiation track), with $z = \epsilon/m$ the microscopic energy release over unit mass of the domain.

In the limit of $D = 0$ and low LET radiation carriers, whereby lethal lesions are Poisson distributed over z , the RBE:

$$\text{RBE} = \frac{\alpha_0 + \beta z_{1D}}{\alpha_x} \quad (2.5)$$

with α_x of the reference radiation, and $\beta(\text{LET}) = \beta_0$.

Subsequent advancements of the MKM have accounted for i) the *overkill* artifacts of the model, by a saturation overhead to z_{1D} [24], and ii) the stochastic distribution of the dose across cell nuclei [25].

2.3.2 Local effect model

All formulations of the *local effect model* (LEM) share the native theoretical framework of LEM I [26], where:

- the radial dose distribution about a radiation track is described via a three-

fold formula, as:

$$D(r) \propto \begin{cases} \text{LET}/r_{min}^2 & r < r_{min} \\ \text{LET}/r^2 & r_{min} \leq r \leq r_{max} \\ 0 & r > r_{max} \end{cases} \quad (2.6)$$

with r_{min} on the order of a few nanometers and $r_{max} \propto E$, i.e. the energy of the radiation carrier;

- an equivalence is established between the local outcome of a radiation track and that of a reference (x-ray) field, upon the basis that lethal/sub-lethal DNA lesions are mediated by secondary electrons, from the inelastic collisions of the radiation with the medium.

Thus, cell survival rates are defined by the (mean) score of lethal events N over the volume V as:

$$N = \int_V dV \ln S_x[d(x, y, z)] V^{-1} \quad (2.7)$$

where $d(x, y, z)$ is the local dose induction. S_x is associated with each of the sensitive sub-voxels of the volume of a cell nucleus, and is described by a LQ-L (*Linear-Quadratic-Linear*) fall-off of the survival rate:

$$S(D) = \begin{cases} e^{-\alpha_x D - \beta_x D^2} & D \leq D_t \\ S_t e^{-s(D-D_t)} & D > D_t \end{cases} \quad (2.8)$$

i.e. a correction to the LQM, where $S_t = S(D_t)$ and D_t account for the overestimate of the cell death rates in the LQM above a dose threshold [27] - with $s = \alpha_x + 2\beta_x D_t$ the *limit slope*, and α_x and β_x of the reference radiation carrier. Thus:

- LEM II [28] refines the late linear fall-off of the survival rates, and accounts for excess lethal events from the stochastic association of sub-lethal SSBs to achieve additional DSBs. Moreover, the formula of the dose in EQ. 2.6 is smeared/convoluted by a two-dimensional Gaussian fit of the radial diffusion of hydroxyl radicals, to involve radical-enforced DNA lesions;

- LEM III [29] rescales the lower r_{min} threshold of the radial dose formula by the relative velocity of the radiation carrier;
- lastly, LEM IV [30] refines the radiation equivalence criteria, defined as the “isoeffective” layout of *isolated* and *clustered* DSBs (rather than the microscopic dose release) from a radiation track and an equivalent dose of reference radiation.

Notably, both (MKM, LEM) frameworks rely upon an analytical formulation of the radial dose, or *amorphous track model*, such as the formalism described in EQ. 2.6.

2.4 Geant4-DNA and Monte-Carlo track structure codes

The mechanistic estimates of the radiosensitivity of cell lines from both the MKM and the LEM benefit from a detailed depiction of the radial dose decay, and of the layout of lethal DNA lesions about a radiation track, down to the nanometric scale. In fact, a manifold scenario modulates the outcome of cell irradiation both *in vivo* and *in vitro*, and this is critical to accelerated ion fields, which are associated with a branched cascade of lethal and sub-lethal events. However, the assessment of the cell-like scenario *in silico* forces a choice in scale and, beside actual limitations, there are diverse accessible layers of the hierarchical chromatin fold that relate to both the micro- and the nanometric scale of the radiation track.

Gold standards of the (micro, nano) dosimetric assessments *in silico*, the **Monte-Carlo track structure** (MCTS) codes evolve the stochastic decay of radiation in a medium⁶, i.e. the cascade collisions of all carriers associated with the radiation field: Particles are attributed initial coordinates and velocities, and their subsequent dynamics is described as an alternate sequence of *free flights* and *events*. The latter are “drawn” stochastically from a collection of (elastic, inelastic) interactions, associated with a (differential)

⁶That is, a water *continuum* in a cell-like scenario.

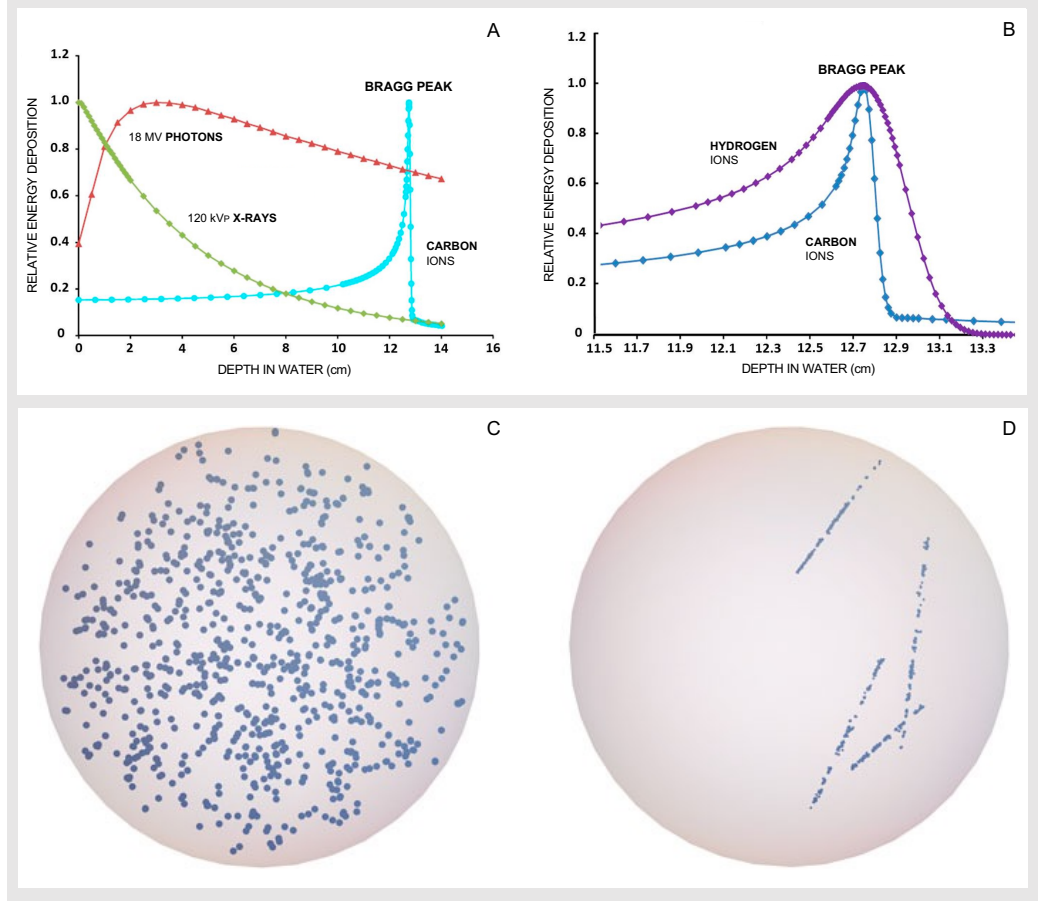


Figure 2.2: (A), (B) Outlines of the (macroscale) dose-depth curves from different radiation carriers. Photons are associated with an exponential dose fall-off, whereas accelerated ions unleash a massive dose fraction within a localized depth in the medium (*Bragg peak*). For the latter, the shape and depth of the curve is modulated by a variety of factors, such as the mass and momentum of the radiation carriers - adapted from Mohamad *et al.* 2017 [31] under CC-BY 4.0 license. **Bottom:** the sparse and localized layouts of ionization events in a 10 micrometer-wide cell nucleus *in silico*, from 1 Gy of (C) x-rays and (D) helium ions (D) respectively - taken from Brzowska *et al.* 2020 [32] under CC-BY 4.0 license.

cross-section ς as:

$$\varsigma(E) = \int_0^E dE \int_0^\pi 2\pi \sin\vartheta d\vartheta \frac{d^2\varsigma(E, W, \vartheta)}{dW d\vartheta} \quad (2.9)$$

with E the kinetic energy, and W and ϑ the subsequent energy loss and scattering angle of the radiation carrier from the event. The new coordinates

and velocities of the carrier⁷ are thus updated via a distribution function of W and ϑ :

$$P(E, W, \vartheta) = \frac{2\pi \sin\vartheta}{\zeta(E)dWd\vartheta} \cdot \frac{d^2\zeta(E, W, \vartheta)}{dWd\vartheta} \quad (2.10)$$

Historically, diverse MCTS codes, such as Geant4 [33–35] and FLUKA [36], have made use of a *condensed history* formalism, whereby the collisions of species below a threshold energy (1 keV) are coalesced into one “collective” event. While this has offered a convenient trade-off to track the whole evolution of the radiation field at the macroscale, it lacks the essential detail of the radial dose distribution at the nuclear and sub-cellular scale [37].

Thus, low-scale, state-of-the-art extensions to MCTS codes were derived to aid in the formulation of the microdosimetric frameworks [38]. Here, an *event-by-event* formalism is adopted, whereby the decay of all “low-rank” carriers is evolved down to a few eV. Advanced codes further depict the subsequent diffusion and reaction of species from the radiolysis of water, so that the rational evolution of a cell-like scenario is detailed within a microsecond of an ion collision⁸. In fact, the inelastic collisions of the radiation account for the excitation/oxidation of water molecules, associated with diverse decay channels and thereby spawning radiolytic radicals and molecules, such as hydrated electrons (e_{aq}^-), OH and H radicals and hydroxyl (OH^-) and hydronium (H_3O^+) ions. An assumption is made that radiolytic species are at thermal equilibrium, thus subject to Brownian diffusion, i.e. $\langle R \rangle \propto (D\Delta t)^{1/2}$; moreover, chemical recombination occurs as two species come into contact, that is, within an effective radius defined upon their reaction rate.

The **Geant4-DNA** toolkit [39–42], and low-scale extension to the Geant4 distribution, thus estimates the earliest layout of (sub)lethal lesions of DNA, upon effective criteria and significant sub-cellular volumes⁹ [43] (*vide infra*).

⁷The depth of the free flight is estimated from the total cross-section, and the azimuthal angle from the collision is drawn from a uniform distribution in $(0, 2\pi)$.

⁸That is about the effective timescale for lethal and sub-lethal DNA lesions to unfold

⁹The layout of events of energy deposition (from the radiation track and the diffusion of radicals) is juxtaposed to a DNA volume depicting a layer of the chromatin framework - such as the DNA double helix or (arrays, fibers of) nucleosomes - and (sub)critical DNA lesions are assessed therein.

Chapter 3

The radiation track at the nucleosome scale

Microdosimetric theoretical frameworks (sections 2.3.1 and 2.3.2), as well as track structure assessments, benefit from arbitrarily “coarse” criteria, which describe as double strand breaks of the DNA backbone the closely associated, twofold events of energy deposition i) beyond a fixed energy threshold, and ii) within a fixed distance, to a certain likelihood [28, 44–46].

The balance of these low-scale criteria is hardly underlain by mechanistical inferences; rather, it matches (indirectly) the outcome of the radiation field at the local scale, i.e. the observed/estimated score of SSBs, DSBs and clustered DNA lesions from assessments *in vivo/in vitro* [47]. Yet, it is beyond the level of theory of the microdosimetric frameworks to discern the likelihood and scores of each DSB *motif*, associated with a variable distance of the twofold cleft of the DNA backbone, which are, in fact, overlooked.

We reckon, however, that the different DSB motifs exert a distinct (kinetic, micromechanical) strain of the DNA molecule, and that short-distanced DSBs alone would crack by means of thermal forces - and effectively within microseconds to milliseconds of an ion collision. This was observed earlier *in silico* [48] and inferred from the *electrophoretic assay* of double strand breaks in plasmid DNA.

In fact, the electrophoretic assay discerns between the supercoiled/intact, circular/SSB- and linear/DSB-cleaved fractions of plasmid DNA, upon mi-

gration in an agarose matrix and subject to an external electric field. To enforce “controlled” DSB motifs, i.e. at a fixed distance and location over a DNA molecule, either i) *restriction endonucleases* (which hydrolyze the phosphoester bonds of the DNA backbone), or ii) *DNA glycolases/AP endonucleases* (which subsequently excise the glycosidic bond of defective nucleic bases and the associated ribose-phosphate moiety) have been employed¹ [49–51]. Collectively, effective fractures of DNA from a DSB lesion have never been achieved beyond a distance of 12 BP. Moreover, it was remarked that the cohesion of the DNA overhangs in-between of the twofold cleft is modulated by the sequence of nucleotides.

In this work, we assess the radiation track structure of hydrogen ions (H^+) of clinically relevant LET^2 *in silico*, at the nanometric scale - i.e. over a *nucleosome tetramer* (chromatin-like) framework - and attempt to relate the low-scale layout of lethal DNA lesions and the local morphology of the DNA fold³.

3.1 Geant4-DNA and the nucleosome tetramer: Materials and methods

The PDB4DNA extension [53] of the Geant4-DNA toolkit takes in a set of C++ libraries, whereby users are allowed to define *tailored* nanometric volumes upon the atomic coordinates of DNA (or hybrid) molecules. Moreover, DNA nucleotides are further classified and labeled by a serial index, which codifies their location over the DNA sequence (that is, residues 1 to 694 of the nucleosome tetramer framework referred to in this work - *vide infra*).

The 1zbb entry [54] of the RCSB Protein Data Bank [55] describes the dimeric unit of the nucleosome tetramer [56], which we transformed to achieve the tetrameric framework (694 BP) associated with PDB4DNA⁴.

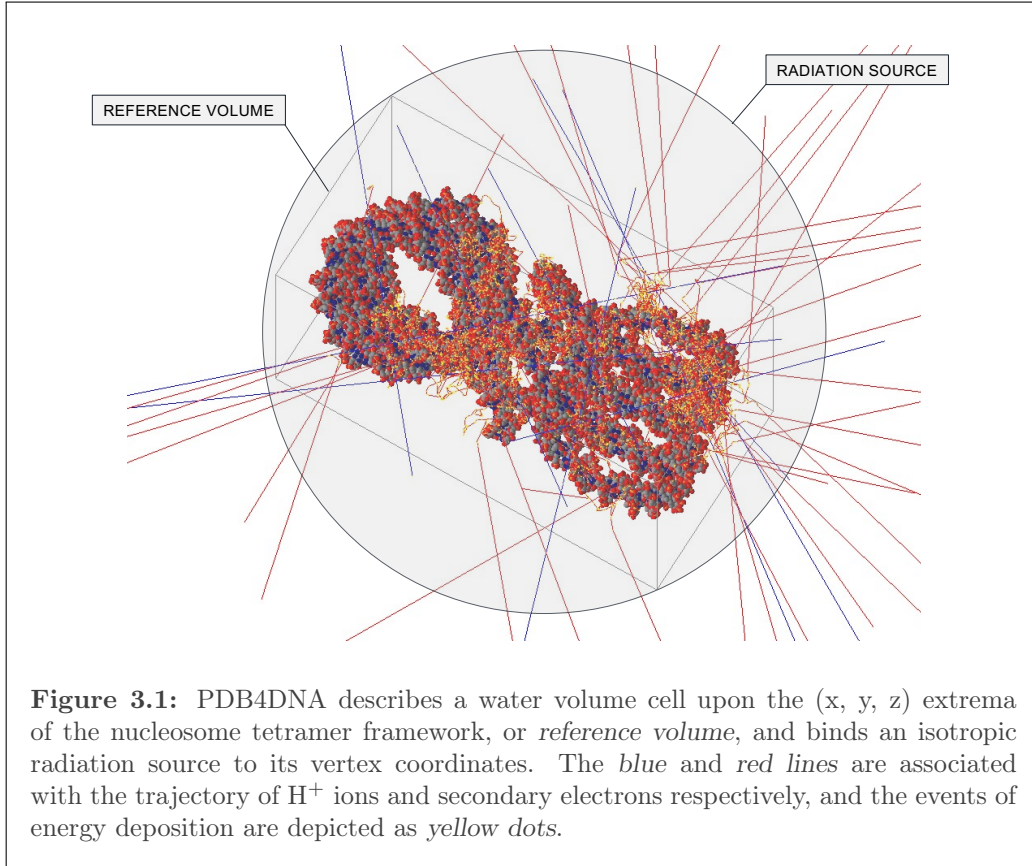
¹In fact, D’souza and Harrison assessed the inhibition of the transcription activity of plasmid DNA from double strand breaks [49].

²That is, at the base and distal edges of the *Bragg peak* (FIG. 2.2)

³This chapter has been adapted from Petrolli *et al.* 2020 [52] under CC-BY 4.0 license.

⁴The records of the edited PDB file were fixed to comply with the format criteria of the *PDBlib* reader - as detailed in Delage *et al.* 2015 [53].

PDB4DNA thus describes a water volume cell upon the extrema of the nucleosome tetramer framework (a $13.0 \times 15.2 \times 25.4$ nanometer box), which we will hereafter refer to as *reference volume* (FIG. 3.1).



The effective features of the nucleosome tetramer - that is, the atomic coordinates and center of mass of all nucleotides - are extracted implicitly and cached as instances of Geant4 classes. The reference volume (a G4Box instance) is made of a continuum medium (G4_WATER, a material defined from the NIST database⁵), and lies within a void environment of “Galactic” vacuum.

The default analysis classes of PDB4DNA score the overall energy deposition within the water cell, and the SSBs and DSBs over the nucleosome tetramer

⁵<https://www.nist.gov/pml/atomic-weights-and-isotopic-compositions-relative-atomic-masses>

fold⁶.

We thus extended the default classes and took in further ROOT [57] scorers of:

- the events of energy deposition (hereafter **hits**) within the coordinates of the reference simulation volume;
- the hits and strand break score over each nucleotide;
- the DSB **distance** score, i.e. the distance between strand breaks that lie at either DNA strand. We define a DSB as the twofold cleft of the DNA backbone within a 10-BP threshold distance.

A strand break is achieved above an overall 8.22 eV energy deposition⁷ to the ribose-phosphate moiety of a nucleotide. In fact, several energy and distance criteria that effectively estimate the scores of (double) strand breaks *in silico* from the direct effect of radiation have been described [58]. However, we remark that our efforts in this concern have been aimed at a critical assessment of the PDB4DNA scenario.

Raw data were collected within the Geant4 10.02-P03 toolkit environment and elaborated via tailored ROOT instances. Each of several *runs* covered 10⁷ radiation tracks and diverse volume choices (*vide infra*). However, all runs were associated with the default G4EmDNAPhysics *list constructor*, which involves a broad set of cross-sections from the electromagnetic interactions of slow electrons⁸ (electronic excitation, ionization, vibrational excitation, dissociative electron attachment and elastic scattering) and ions (excitation, ionization, electron capture) with the water medium⁹.

The default PDB4DNA scenario involves an isotropic, outer radiation source, described over the vertex coordinates of the reference volume (FIG. 3.1). Particles, that is 500 keV to 5 MeV H⁺ in this work, are thus shot towards the water cell. An isotropic source accounts for the statistical effect of the

⁶Track-wise data are collected at runtime.

⁷Such threshold defines the first excitation level of the water molecule in Geant4-DNA.

⁸Each class of interaction is associated with a range of applicability, down to a few eV, as described in the Geant4-DNA documentation (<http://geant4-dna.in2p3.fr/styled-3/styled-8/index.html>)

⁹As well, several choices for the list constructor of cross-sections are available [42].

radiation field that is 4π -distributed over the nucleosome tetramer fold, which we reckon to be an effective depiction of a chromatin fiber scenario. Moreover, the radiation source is bound to the active box and it widens as the volume of water is stretched from the reference coordinates of the nucleosome tetramer (*vide infra*).

3.2 MCTS assessments at the nanometric scale are biased from statistical artifacts

In the default PDB4DNA scenario, a 500 keV H^+ radiation field strikes the reference volume from a fit, isotropic source. The hit score, i.e. the *run-wise* score of energy deposition events at each nucleotide of the nucleosome tetramer fold, is biased explicitly¹⁰ (FIG. 3.2).

We expected that a correlation arose between the hit score and the local morphology of the DNA fold, whereby to associate the likelihood for a nucleotide to be struck/involved with a DSB with its location over the nucleosome tetramer fold.

However, no such correlation arises, as the spikes in the hit score involve nucleotides over the linker and nucleosomal DNA alike. Notably, these spikes describe a cluster of nucleotides at the core of the nucleosome tetramer fold (marked in *red* in FIG. 3.2), which we reckon to be a statistical artifact. In fact, the water volume bound to the reference nucleosome tetramer is over-sampled at its core coordinates and nucleotides (FIG. 3.3).

We will hereafter refer to this effect as the **hit artifact**: As the carriers of the radiation field leak from the water medium (into a vacuum environment), their tracks are cut off *de facto*. Hence, no further collisions/events are detailed about the outer solvation shell (i.e. outside of the reference volume), and we lack track data at the boundaries of the active water box, where a local imbalance arises. Thus, as a minor dose fraction is deposited over nucleotide “outliers”, the overall statistical assessment of the nucleosome tetramer is an oversample of the core of the reference volume, which is taken

¹⁰Likewise, so are the SSB and DSB scores - not shown here.

up (incidentally) by linker DNA chains. Indeed, the hit artifact biases the assessment of the radiation track at the nanometric scale in a context where we lack an effective treatment of the boundaries of the water volume and/or track data at the outer shell.

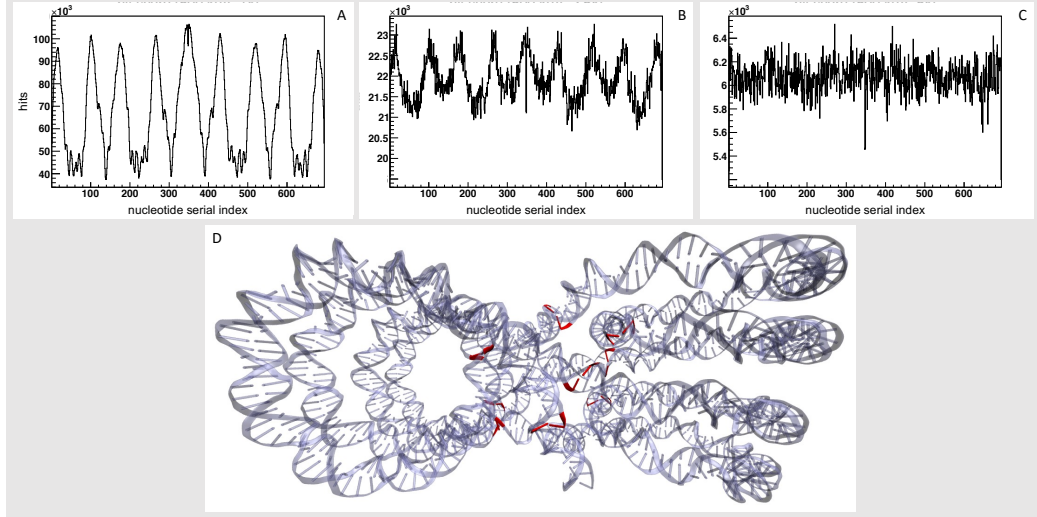


Figure 3.2: Hit scores over the nucleosome tetramer fold from 500 keV hydrogen ions at a (A) 1-fold, (B) 2.5-fold, and (C) 5-fold rescaling of the reference volume. The hit spikes in (A) are associated with a cluster of nucleotides at the core of the nucleosome tetramer (red nucleotides in (D)), and vanish as the active volume is rescaled - adapted from Petrolli *et al.* 2020 [52] under CC-BY 4.0 license.

To circumvent the hit artifact, we scaled/stretched the water cell from the reference volume of the nucleosome tetramer, i.e. we rescaled each size of the active (default) water volume isotropically by one to fivefold. This trivial workaround ensured that all nucleotides are struck evenly, whereby the spikes of the hit score effectively vanish as the water volume is rescaled (FIG. 3.2). Thus, while we reckon that the hit artifact applies to all DNA frameworks at the nanometric scale alike, a *solvation shell* ensures that the radiation field is detailed about the outer layers/nucleotides of the nucleosome tetramer - which now lies within a wider water cell.

However, this raises a further issue, i.e. to establish a convenient tradeoff, where to describe the radiation field with exhaustive detail and not overscale the active water volume.

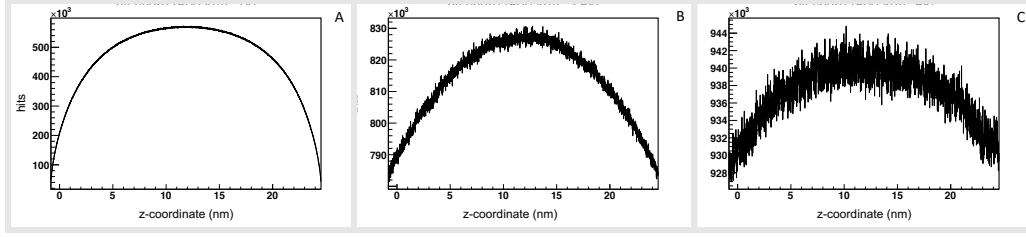


Figure 3.3: Hit scores over the reference volume frame of the nucleosome tetramer (coordinates are arbitrary and described in the PDB file), from 500 keV hydrogen ions at a (A) 1-fold, (B) 2.5-fold, and (C) 5-fold rescaling of the reference volume. The imbalance between the core volume and the outer layers of the nucleosome tetramer underlies a statistical artifact - adapted from Petrolli *et al.* 2020 [52] under CC-BY 4.0 license.

3.2.1 Shannon’s entropy as a bias estimator

To estimate the share of effective data we collect or waste about the nucleosome tetramer as we rescale the active water box, we define:

- the **volume hit score** (VHS) as the overall score of energy deposition events within the reference volume cell (i.e. defined upon the coordinates and extrema of the nucleosome tetramer - as described in section 3.1);
- the **DNA hit score** (DHS) as the subset of all VHS events that fall over the backbone of the nucleotides of the nucleosome tetramer.

The isotropic rescaling of the active water cell is associated with an overall increase in the VHS (FIG. 3.4). In fact, the solvation shell ensures that the radiation field is detailed at the outer layers of the reference volume - whereby we achieve a statistically unbiased strike of both the reference volume (FIG. 3.3) and the nucleosome tetramer (FIG. 3.2).

Yet, we scored fewer hits over the DNA backbone, i.e. the DHS decreases via the rescaling of the water cell (FIG. 3.4), which is controversial. In fact, the DNA backbone effectively takes up a minor fraction of a nucleosome volume, and in the default PDB4DNA scenario we overstrike the “crowded” core of the nucleosome tetramer - where the radiation field is effective. We thus expect this imbalance to be underlain by the uneven morphology of the nucleosome tetramer, where the outer layers are associated with a thinner distribution of nucleotides than the core of DNA linkers.

To estimate the bias of the DNA hit score, and to establish a convenient VHS–DHS tradeoff, we referred to a normalized formula of *Shannon’s entropy*, defined as:

$$S = -\frac{1}{\log N} \sum_{i=1}^N P_i \log P_i \quad (3.1)$$

where the index i runs over the $N = 694$ nucleotides, and P_i is the hit score over the i -th BP, divided by the overall DNA hit score. Thus, S varies between 0 (maximal bias) and 1 (unbiased distribution).

Shannon’s entropy rises steeply, and about its maximal value, via the isotropic rescaling of the active cell (FIG. 3.4). Moreover, no effective increase in S is further observed beyond a 2.5-fold rescaling of the default reference volume, which we thus established to be the minimal size of the active water cell to achieve an unbiased strike of the nucleosome tetramer fold.

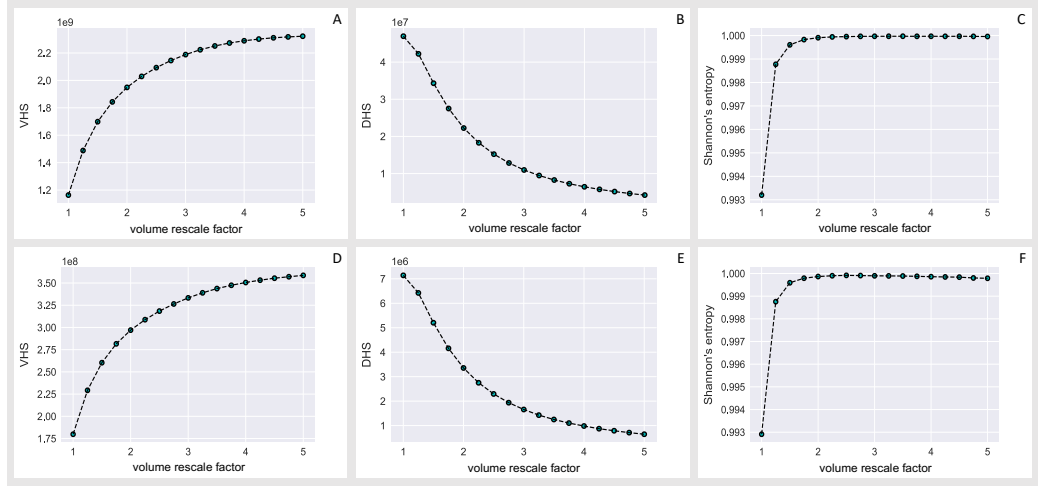


Figure 3.4: The increase in the hit score over the reference volume (VHS v. rescaling factor) is associated with a decrease in the hit score over the DNA backbone (DHS) at (A), (B) 500 keV and (D), (E) 5 MeV alike, which is controversial (section 3.2.1). The bias in the hit distribution over the nucleosome tetramer (FIG. 3.2) is estimated by a normalized *Shannon’s entropy* formula (EQ. 3.1) at (C) 500 keV and (F) 5 MeV. We achieve no further statistical benefit beyond a 2.5-fold rescaling of the reference water cell - adapted from Petrolli *et al.* 2020 [52] under CC-BY 4.0 license.

However, such threshold is tailored to the 500 keV scenario; hence, we extended the assessment further to 5 MeV hydrogen ions. Notably, both the volume and the DNA hit scores at 500 keV and 5 MeV behave alike, within a sevenfold decrease of all VHS and DHS values in the latter case (FIG. 3.4).

This is in line with the effective decrease in the LET, i.e. with fewer energy deposition events at 5 MeV than at 500 keV.

Likewise, the 500-keV steep rise in Shannon’s entropy is mirrored at 5 MeV; as well, S is maximal at a 2.5-fold isotropic rescaling of the default reference volume. We thus finally established that a **2.5-fold** rescaling of the active water cell i) achieves an unbiased and effective strike of the nucleosome tetramer, at all scenarios between 500 keV and 5 MeV, and ii) is associated with a convenient tradeoff between the loss of data and a satisfactory depiction of the core of the radiation track at the nanoscale.

3.3 The double strand break distance score

In view of these outcomes, we assessed further scenarios within the 500 keV to 5 MeV H^+ interval (with a 0.25 MeV stride), at a fixed 2.5-fold rescaling of the reference volume, and took in a further scorer, which we will hereafter refer to as the **double strand break distance score** (DDS).

It has been widely acknowledged that the class of DNA double strand break shall be defined by a threshold distance criterion between the twofold cleft of the DNA backbone. Such threshold is rational to a microdosimetric level of theory, where the sub-lethal lesions of the DNA molecule are clustered upon coarse criteria, and driven by both observations *in vitro* and theoretical inferences - as mentioned earlier in the text. However, each DSB *motif* (which is associated with a variable distance between strand breaks) exerts a local, unique strain of the DNA backbone, and the micromechanical behavior of the lesion interface *in silico* is modulated by the DSB distance [48, 59].

We thus defined and scripted a further ROOT scorer that tracked each DSB motif (zero to 10 BP) over the nucleosome tetramer fold. Notably, all DDS distributions between 500 keV and 5 MeV share a Poisson fit, and are biased towards short-distanced DSBs (FIG. 3.5). Moreover, the mean values of the DDS decrease (mildly) between 500 keV and 5 MeV, whereby we reckon that a correlation exists between a DNA morphology, the radiation track and an effective DSB motif - yet, we lack an exhaustive dataset.

We further observed that the mean DDS fluctuates as we rescale the active

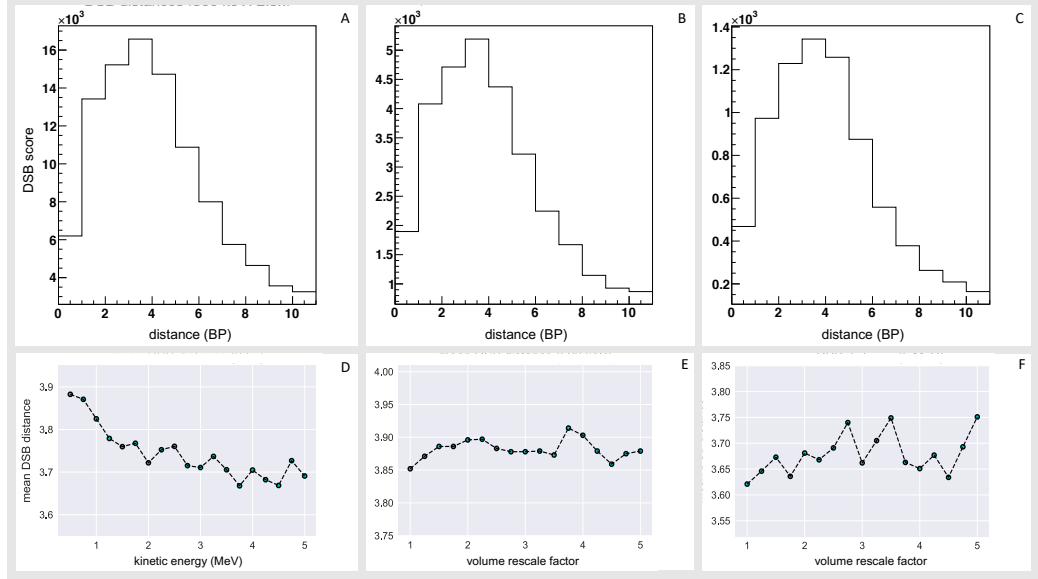


Figure 3.5: DSB distance scores (DDS) at fixed 2.5-fold isotropic rescaling of the reference volume, at (A) 500 keV, (B) 1.5 MeV, and (C) 5 MeV. All distributions share a Poisson fit and are biased towards short-distanced DSBs. Moreover, (D) the mean values of the DDS decrease between 500 keV and 5 MeV (at fixed 2.5-fold rescaling), and fluctuate mildly v. the rescaling factor of the active water volume at both (E) 500 keV and (F) 5 MeV - adapted from Petrolli *et al.* 2020 [52] under CC-BY 4.0 license.

cell (FIG. 3.5). However, as we overscale the water volume beyond a 2.5-fold factor (i.e. the threshold where we achieve an effective solvation shell and moderate the data loss), we overstrike the outer, blank layers of the water cell and collect no significant details of the core of the radiation track - we thus expect the mean DDS to converge eventually.

3.4 Conclusive remarks

In a brief assessment of the low-scale layout of direct DNA lesions, associated with the core of the radiation track of hydrogen ions at a Bragg peak-relevant energy range (500 keV - 5 MeV)[60], we infer that:

- the *in silico* assessments of track structure at the nanometric scale (i.e. below an *effective* radial dose cutoff) are biased by statistical *hit artifacts*, whereby volumes at the core coordinates of the active water box are over-sampled, unless the behavior of the radiation track at the outer solvation

shells of a DNA framework is accounted for;

- a bias score of the stochastic distribution of hits based on Shannon's entropy let us establish a convenient tradeoff between the effective cutoff of the radial dose distribution at the "core" of a radiation track and the size rescaling of the solvation shell about the sensitive water volume;
- short-distanced DSB *motifs* (i.e. where the twofold cleft of the DNA backbone is distanced by 1 to 5 nucleotides) are favored by the direct effect of hydrogen ions at a Bragg peak-relevant energy range (500 keV - 5 MeV). We remark, however, that a statistical assessment of DSBs shall not overlook the excess, indirect contribution from radiolytic radicals.

Chapter 4

The theoretical framework of molecular dynamics

4.1 The thermodynamic ensemble

The core concept of the theoretical framework underneath *molecular dynamics* is the **thermodynamic ensemble**, that is the exhaustive collection of the *microstates* accessible to a macroscopic system of defined and fixed *control variables* (such as the amount of matter, the volume and the enthalpy). The microstates (\mathbf{x}) are defined from the coordinates q_α and (conjugate) momenta p_α of the atoms of the system, and their dynamical behavior across the *phase space*¹ is described by *Hamilton's equations*:

$$\dot{q}_\alpha = \frac{\partial \mathcal{H}}{\partial p_\alpha} \quad \dot{p}_\alpha = -\frac{\partial \mathcal{H}}{\partial q_\alpha} \quad (4.1)$$

where \mathcal{H} is the system **hamiltonian**, or the (kinetic, potential) energy of the system - from classical mechanics²:

$$\mathcal{H}(\mathbf{q}_1, \dots, \mathbf{q}_N, \mathbf{p}_1, \dots, \mathbf{p}_N) = \sum_i \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{q}_1, \dots, \mathbf{q}_N) \quad (4.2)$$

¹That is, the $6N$ -dimensional space of the atomic coordinates and momenta.

²Here, forces are conservative, i.e., $\mathbf{F}_i(\mathbf{q}_1, \dots, \mathbf{q}_N) = -\nabla U(\mathbf{q}_1, \dots, \mathbf{q}_N)$

Thermodynamic ensembles are associated with a *density distribution function* of the microstates over the phase space $f(\mathbf{x},t)$, which is conserved at thermodynamic equilibrium:

$$\{f(\mathbf{x},t), \mathcal{H}(\mathbf{x})\} = \frac{\partial f(\mathbf{x},t)}{\partial q_\alpha} \frac{\partial \mathcal{H}(\mathbf{x})}{\partial p_\alpha} - \frac{\partial f(\mathbf{x},t)}{\partial p_\alpha} \frac{\partial \mathcal{H}(\mathbf{x})}{\partial q_\alpha} = 0 \quad (4.3)$$

and thus defined upon the system hamiltonian, i.e. $f(\mathbf{x},t) = F[\mathcal{H}(\mathbf{x})]$, up to a normalization factor:

$$f(\mathbf{x},t) = \frac{F[\mathcal{H}(\mathbf{x})]}{\int d\mathbf{x} F[\mathcal{H}(\mathbf{x})]} = \frac{F[\mathcal{H}(\mathbf{x})]}{Z} \quad (4.4)$$

Z defines the **ensemble partition function**, whereby all thermodynamic variables of the ensemble, as well as equilibrium observables, are derived:

$$\mathcal{A} = \langle a(\mathbf{x}) \rangle = Z^{-1} \int d\mathbf{x} a(\mathbf{x}) \cdot F[\mathcal{H}(\mathbf{x})] \quad (4.5)$$

4.2 Molecular dynamics

Yet, Z is far too elaborate an object to handle analytically; in fact, it escalates with N , and forces of $\mathcal{H}(\mathbf{x})$ are strictly non-linear. Thus, numerical, discrete methods were devised to estimate the thermodynamic observables of an ensemble, upon the equivalence:

$$\begin{aligned} \mathcal{A} = \langle a(\mathbf{x}) \rangle &= Z^{-1} \int d\mathbf{x} a(\mathbf{x}) \cdot F[\mathcal{H}(\mathbf{x})] \\ &= \lim_{t \rightarrow \infty} t^{-1} \int_0^t dt a(\mathbf{x}_t) \approx M^{-1} \sum_{n=1}^M a(\mathbf{x}_{n\Delta t}) \end{aligned} \quad (4.6)$$

or that the system is *ergodic*, i.e. the microstates of an ensemble are sampled exhaustively in an infinite time³.

In **molecular dynamics** (MD), systems of atoms are evolved from initial coordinates and velocities, via a numerical, step-wise *solver* of Hamilton's

³Taken that no infinite trajectory is achievable *in silico*, the estimates of ensemble observables are as reliable as they are statistically unbiased.

equations, such as the *velocity-Verlet* routine⁴:

$$\begin{cases} \mathbf{r}_i(t + \Delta t) \sim \mathbf{r}_i(t) + \mathbf{v}_i(t)\Delta t + \frac{\mathbf{F}_i(t)\Delta t^2}{2m_i} \\ \mathbf{v}_i(t + \Delta t) \sim \mathbf{v}_i(t) + \frac{[\mathbf{F}_i(t + \Delta t) + \mathbf{F}_i(t)]\Delta t}{2m_i} \end{cases} \quad (4.7)$$

which arises from the discrete formulation of the *classic propagator* of the phase space vector.

The dynamics of biomolecules is thus assessed within neutral cells of water/salt media. *Periodic boundary conditions* (PBC) are enforced to avoid surface artifacts, whereby cells interface isotropically with infinite self-replicas. However, a *minimal image convention* is adopted, i.e. the closest of the binary interactions between infinitely replicated atoms is accounted for (once)⁵. Even so, further effective routines have been devised to decrease the $\sim N^2$ overhead of the calculation of forces.

4.2.1 Non-bonded forces

The binary *non-bonded* (NB) *forces* are associated with intermolecular as well as intramolecular⁶ interactions, and described by the *Lennard-Jones* and classic *electrostatic* formulations as:

$$\mathcal{U}_{nb}(\mathbf{r}_1, \dots, \mathbf{r}_N) \approx \sum_{i>j \in nb} \left\{ \varepsilon_{ij} \left[\left(\frac{\varsigma_{ij}}{\mathbf{r}_{ij}} \right)^{12} - \left(\frac{\varsigma_{ij}}{\mathbf{r}_{ij}} \right)^6 \right] + \frac{q_i q_j}{\mathbf{r}_{ij}} \right\}, \quad (4.8)$$

where ς_{ij} , ε_{ij} , as well as the (fractional) charges q_i , q_j are defined within the chosen model, or **force field**.

Lennard-Jones forces are short-ranged, and negligible beyond an effective radius, whereas electrostatic forces are associated with a soft tail-off. A *cutoff* threshold \mathbf{r}_k is thus established, and the latter are further treated as:

⁴A *time step* $\Delta t = 1$ fs is defined as a convenient dynamical tradeoff, about 1/10 of the vibration of X-H bonds, in all-atom classical models.

⁵Nevertheless, the self-interaction of the solutes, i.e. across virtual replicas, shall be avoided - this is balanced by force cutoffs (*vide infra*).

⁶That is, between atoms beyond three covalent bonds.

- *short-range* interactions in the real space below 10-12 Å (or about 3ζ). Both Lennard-Jones and short-range electrostatic forces are thus (up)shifted and applied a *switch* function, whereby \mathcal{U}_{short} vanish about \mathbf{r}_k ($\mathcal{U}_{short}(\mathbf{r}_k) = 0$) and an artificial disruption of the forces is avoided. Moreover, *Verlet lists* of the close neighbors (that is, within $\mathbf{r}_k + \delta(n\Delta t)$) of all atoms are kept and updated at effective intervals of $n\Delta t$, thus enhancing the search for binary forces that satisfy $\mathbf{r}_{ij} \leq \mathbf{r}_k$.
- *extended-range* electrostatic forces beyond the cutoff radius, and evaluated in the reciprocal Fourier space via a discrete formulation of the *Ewald sum* (*smooth Particle Mesh Ewald*).

4.2.2 Thermostats and barostats

The numerical solutions to Hamilton's equations enforce the time conservation of the system hamiltonian:

$$\frac{d\mathcal{H}}{dt} = \frac{\partial\mathcal{H}}{\partial\mathbf{q}_\alpha}\dot{\mathbf{q}}_\alpha + \frac{\partial\mathcal{H}}{\partial\mathbf{p}_\alpha}\dot{\mathbf{p}}_\alpha = \frac{\partial\mathcal{H}}{\partial\mathbf{q}_\alpha}\frac{\partial\mathcal{H}}{\partial\mathbf{p}_\alpha} - \frac{\partial\mathcal{H}}{\partial\mathbf{p}_\alpha}\frac{\partial\mathcal{H}}{\partial\mathbf{q}_\alpha} = 0 \quad (4.9)$$

thus describing the trajectory of the phase space vector over an $\mathcal{H}(\mathbf{x}) = E$ hypersurface - or the *microcanonical ensemble* of the microstates of an isolated system. Hence, to assess thermodynamic observables associated with (realistic) *isothermal* and/or *isobaric* ensembles of microstates, one shall deviate from a strictly hamiltonian formalism and devise alternate expedients that would balance the fluctuations of momenta (fixed NVT, or the *canonical ensemble*) and of the volume (fixed NPT, or the *isothermal-isobaric ensemble*).

The **Langevin thermostat** modulates the velocities of the atoms in a system via both a viscous coefficient γ , which accounts for an exponential decay of $\dot{\mathbf{r}}_i$, and a stochastic force η_i , as:

$$\dot{\mathbf{r}}_i(t) = \frac{\mathbf{F}_i[\mathbf{r}_1(t), \dots, \mathbf{r}_N(t)]}{m_i\gamma} + \eta_i(t) \quad (4.10)$$

Thus, all atoms are subject to instantaneous collisions from an artificial bath, where the stochastic forces $\eta_i(t)$ are Gaussian-distributed about $\langle \eta_i(t) \rangle = 0$ and delta-correlated:

$$\langle \eta_i(t) \cdot \eta_j(t') \rangle = 6D\delta_{ij}\delta(t - t') \quad D = \frac{kT}{m\gamma} \quad (4.11)$$

The discrete solver for the atomic coordinates:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \frac{\mathbf{F}_i(t)}{m_i\gamma}\Delta(t) + \xi_i(t)\sqrt{\frac{2kT \cdot \Delta t}{m_i\gamma}} \quad (4.12)$$

involves $\xi_i(t_k) \propto \eta_i(t_k)$ - where $\langle \xi_i(t) \rangle = 0$ and $\langle \xi_i(t_k)\xi_j(t_l) \rangle = \delta_{ij}\delta_{kl}$.

The **Langevin barostat** enforces a dynamical control of the volume, which is therefore an effective variable of an extended phase-space hamiltonian. The dynamics of the volume is mediated from a stochastic “force”, via a *Nosé-Hoover chain thermostat* (not described here), which synchronously scales the atomic coordinates/momenta and the volume:

$$\left\{ \begin{array}{l} \dot{\mathbf{r}}_i = \frac{\dot{V}}{3V}\mathbf{r}_i + \frac{\mathbf{p}_i}{m_i} \\ \dot{\mathbf{p}}_i = -\frac{\partial \mathcal{U}}{\partial \mathbf{r}_i} - \frac{\dot{V}}{3V}\mathbf{p}_i - \xi_1\mathbf{p}_i \\ \dot{V} = \frac{\mathbf{p}_V}{W} \\ \dot{\mathbf{p}}_V = P(t) - P - \gamma\dot{V} + R(t) \\ \langle R(0) \cdot R(t) \rangle = \frac{2\gamma kT\delta(t)}{W} + \text{Nosé-Hoover chain} \end{array} \right.$$

Here, the friction coefficient ξ_1 is associated with the Nosé-Hoover chain, W defines an *effective mass* to the volume momentum, and:

$$P(t) = \frac{1}{3V} \sum_i \left[\frac{\mathbf{p}_i^2}{m_i} - \frac{\partial \mathcal{U}}{\partial \mathbf{r}_i} \mathbf{r}_i \right] \quad (4.13)$$

estimates the *internal pressure* of the system, whereby the dynamics of the volume is modulated by the balance between $P(t)$ and the pressure enforced via the barostat.

4.2.3 The CHARMM force field

The *solver* associated with a set of control variables ensures that the exact statistical distribution of microstates of the ensemble underneath is achieved, in an infinite time. However, it is the forces \mathbf{F}_i that underlie the hamiltonian to describe the equilibrium and dynamical behavior of the system.

In classical *molecular mechanics*, it is theorized that the atomic nuclei are subject to the self-consistent field of the electronic distribution at equilibrium, whereby atoms are attributed a fixed (fractional) charge and van der Waals radius. Moreover, covalent bonds cannot be created nor broken, i.e. molecular topologies are fixed, so that atoms behave as a framework of harmonic oscillators about a manifold equilibrium. The **force fields** thus describe the effective rules to dictate the local behavior of molecules and atoms, and the forces underneath.

The formulation of the classical **CHARMM** (Chemistry at Harvard Macromolecular Mechanics) all-atom force field:

$$\begin{aligned}
\mathcal{U} = & \sum_{\text{bonds}} K_b(b - b_{eq})^2 + \sum_{\text{angles}} K_{\vartheta}(\vartheta - \vartheta_{eq})^2 + \sum_{\substack{\text{improper} \\ \text{dihedrals}}} K_{\varphi}(\varphi - \varphi_{eq})^2 \\
& + \sum_{\text{dihedrals}} \sum_{n=1}^6 K_{\phi,n}(1 + \cos(n\phi - \delta_n)) \\
& + \sum_{\substack{\text{nonbonded} \\ i>j}} \left\{ \varepsilon_{ij} \left[\left(\frac{S_{ij}}{\mathbf{r}_{ij}} \right)^{12} - 2 \left(\frac{S_{ij}}{\mathbf{r}_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 \mathbf{r}_{ij}} \right\} \\
& + \sum_{\substack{\text{Urey-Bradley} \\ \text{angles}}} K_{\text{UB}}(r_{1-3} - r_{1-3,eq})^2 + \text{CMAP}
\end{aligned} \tag{4.14}$$

has undertaken subsequent refinements upon quantum-chemical calculations and experimental benchmarks [61, 62], and is an advanced force field to depict

canonical and derivative amino acids, nucleic acids, carbohydrates and lipids - and macromolecular assemblies thereof - as well as functional biomolecules and cofactors. It involves:

- four *bonded* classes of forces/torques, which account for local equilibrium distances and torsional conformers, and the chemical hybridization of the covalent framework;
- the Lennard-Jones and electrostatic *nonbonded* forces - as described earlier in the text (section 4.2.1);
- the bonded *Urey-Bradley* (mild) refinements of the vibrational motions between 1-3 atoms;
- the two-dimensional CMAP cross-corrections to the distributions of adjacent (ϕ , ψ) backbone dihedrals, and to the affinity to secondary motifs of all amino acids [63].

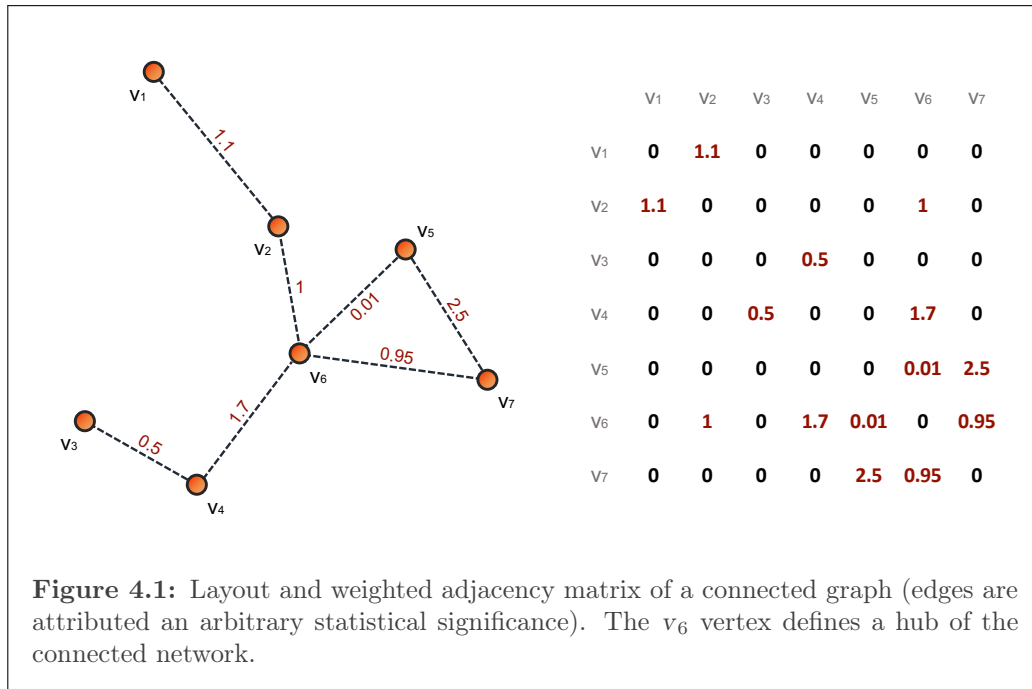
4.3 The PyInterph workflow

Macromolecular assemblies, such as the nucleosome fold, describe active frameworks of **connected nodes** (amino acids, nucleotides), able to convey mechanical strains via weak intermolecular forces. The formalism of *graph theory* has thus been exploited to reframe the dynamics of biomolecules as that of a connected network. In graphs, *vertices* (nodes) are said to be adjacent as an *edge* connects them; such connections may be described on a binary basis (on/off) or be attributed a statistical significance by weighing the edges (FIG. 4.1).

Within a connected graph, all vertices are bound by continuous sequences of contacts; thus:

- the *degree* of a vertex defines the amount of edges branching from a node; a vertex associated with a high degree is referred to as *hub*;
- the *distance* accounts for the shortest route between (remote) vertices and is infinite between nodes of disconnected subgraphs.

A convenient dataframe to depict connected networks is via **adjacency** (or *contact*) **matrices**, whereby the (weighted/unweighted) connections between all vertices of a graph are collected (FIG. 4.1).



The **PyIntergraph** workflow [64] exploits the theoretical formalism of graph theory and the flexible framework of *MDAnalysis* [65, 66] to extract the contact matrices between amino acids (nodes) of a macromolecular assembly from an MD trajectory. It involves a recursive, frame-wise assessment of persistence⁷ of three classes of binary intermolecular interactions, i.e. *electrostatic* and *hydrophobic contacts*, and *hydrogen bonds*, to elaborate weighted adjacency matrices. These matrices are thus filtered upon a criterion of minimal contact threshold (*vide infra*) and merged into a unique dataframe, thereby allowing for the assessment of significant subgraphs and shortest routes between connected nodes (section 5.3.4).

⁷That is, the fraction of frames which a contact (edge) is active for.

Chapter 5

MD assessment of the nucleosome interface

The *in silico* assessments of nucleosomes have flanked the observation of chromatin and of nuclear factors *in vitro* about two decades now [67–69]. In fact, atomistic molecular dynamics has detailed i) the erratic tumble of histone tails (*vide infra*), ii) the thermal, salt-driven and forceful disruption of the nucleosome fold, iii) the contact interface between DNA and the core histone octamer, and between nucleosomes, iv) the modulations of the local dynamics of nucleosomes from PTMs of the histone tails, histone variants and defective DNA lesions.

We hereby benchmark an extension of the *PyInteraph* toolkit (section 4.3) to nucleic acids (section 5.2), upon the acknowledged dynamics of the nucleosome fold at equilibrium, and assess the manifold mechanical role of histone tails in the NCP.

5.1 The nucleosome framework: Materials and methods

The nucleosome framework has been modeled after the 1kx5 entry [70] in the RCSB Protein Data Bank [55]. It describes a wild-type nucleosome core particle at 1.9 Å-resolution, crystallized from *Xenopus laevis* core histones and a 147 BP nucleic stretch extracted from human α -satellite DNA. It is re-

marked that the 1kx5 nucleosome framework lacks the $^1\text{PEP}^3$ distal residues of the N-terminal tails of both histones H2B, which we described via the MODELLER toolkit [71]. Moreover, we extended the outer (entrance, exit) DNA termini by a 20 BP (GACT)₅ “linker” stretch in the B-conformation, via the *Avogadro* [72] and VMD [73] interfaces, as done elsewhere [59, 74]. The 187-BP, full nucleosome framework, hereafter *reference nucleosome*, was therefore solvated within a 1.5 nm-wide water box shell - that is, 91.297 TIP3P water molecules overall - and neutralized by an excess 150 mM KCl (484 K⁺ and 258 Cl⁻ ions).

All trajectories, that is i) one 1-microsecond trajectory of the *reference nucleosome* and ii) three 500-nanosecond trajectories of *fractured nucleosome* frameworks (chapter 6) were collected in NAMD 2.13 [75], via the CHARMM36 all-atom force field [62, 76] and a refinement of electrostatic and Lennard-Jones forces (Yoo and Aksimentiev [77]).

We followed a thermalization routine on all nucleosome frameworks (with either frozen or mildly restrained solute coordinates), that involves i) 5.000 conjugate-gradient minimization steps, ii) 0.5-ns thermal equilibration in the canonical NVT ensemble, and iii) 10-ns equilibration of the solvent in the NPT ensemble. Fixed $T = 310$ K and $P = 1$ atm were enforced via the Langevin thermostat and Nosé-Hoover/Langevin barostat.

Isotropic PBC were associated with the Particle Mesh Ewald calculation of the full-system electrostatics, and a 12 Å cutoff was established to both Van der Waals and electrostatic forces. SHAKE constraints were applied to all hydrogen atoms, thus allowing a timestep of 2 fs for the whole system. To moderate the fluctuations of the linker DNA termini, as in a chromatin-like scenario, atoms at the outermost nucleotides were mildly restrained, as done by Cleri and coworkers [59]. As for the reference nucleosome, we discarded the earliest 100 nanoseconds worth of raw data, upon the convergence of the all v. all RMSD of the core helix fold domain.

5.2 Extension of the PyInteraph workflow to the nucleosome interface

To describe the contact interface between DNA and the core histone fold in the NCP (hereafter **nucleosome interface**), we extended the *PyInteraph* workflow with active nodes from the *nucleic acids* as follows:

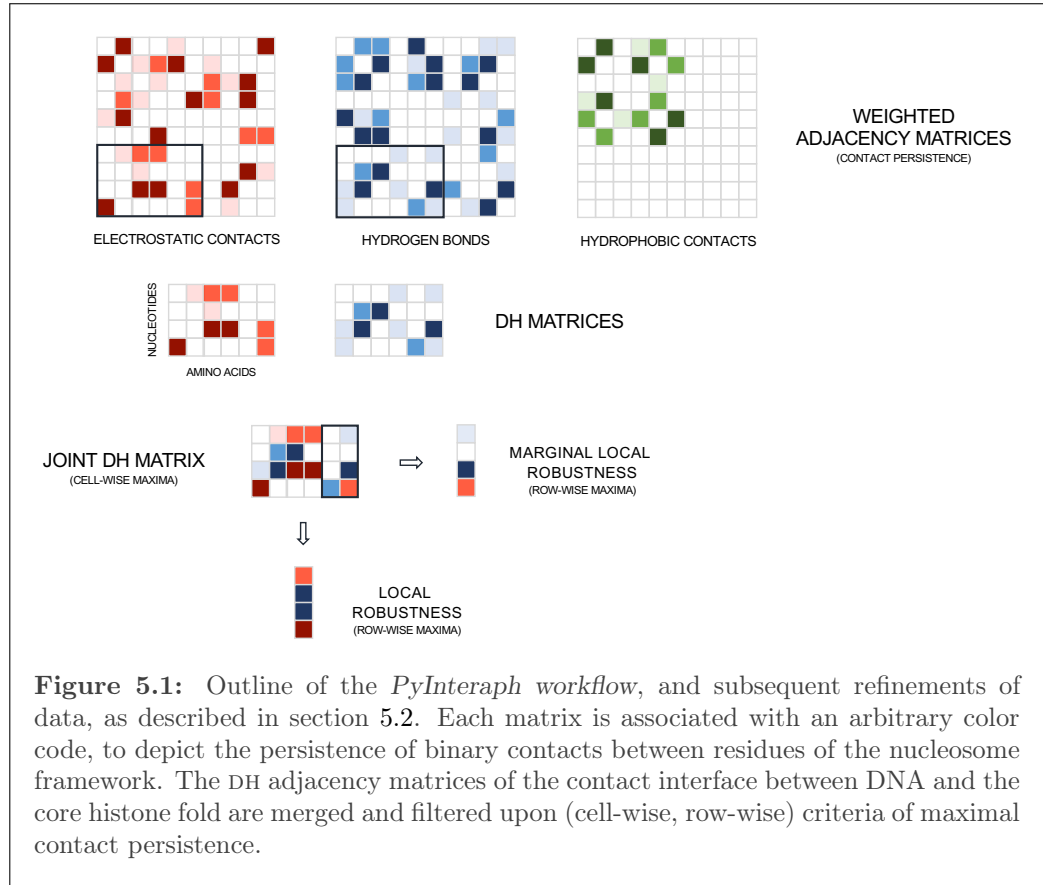
- *acidic* residues, i.e. the carboxylate moieties of ASP and GLU involved with the class of electrostatic contacts, were added with the charged phosphate moiety of DNA nucleotides;
- *hydrogen bond acceptors* were added with all oxygen atoms of the DNA backbone, as well as with all heteroatoms of the nucleic bases whose lone electrons were not shared with the aromatic moiety - that is, *adenine* N3, N1 and N7; *guanine* N3, O6 and N7; *thymine* O2 and O4; *cytosine* N3 and O2;
- *hydrogen bond donors* were added with O5' and O3' hydroxyl termini of the DNA backbone and NH sites of the nucleic bases.

Electrostatic contacts in *PyInteraph* are established between acidic/basic moieties as their atom-wise distance lies within 4.5 Å; *hydrophobic* contacts are established between hydrophobic side chains within a 5 Å threshold distance of their center of mass; lastly, *hydrogen bond* criteria are satisfied as the distance and tilt between the donor and acceptor moieties lie within 3.5 Å and beyond 150° respectively.

Binary, inter-residual contact data have been collected from all nucleosome scenarios in 0.5-ns steps and stored as ($N \times N$) weighted adjacency matrices¹ - one matrix associated with each class of intermolecular interactions - with $N = 1354$ residues of the nucleosome framework. As for the dynamics of the nucleosome interface, however, we isolated the sub-blocks that hold the ($ND \times NH$) data of the contacts between DNA and the core histone fold, hereafter *DH matrices*, with $ND = 374$ nucleotides and $NH = 980$ histone residues (FIG. 5.1).

¹Here, weights describe the persistence of binary contacts.

We thus merged the two DH matrices that are significant to the nucleosome interface (that is, the sub-blocks of the electrostatic contacts and hydrogen bonds), over a cell-wise criterion of maximal contact persistence, and achieved a unique dataframe, which we will refer to as **joint DH matrix** (FIG. 5.1). The joint DH matrix describes the role of each nucleotide and amino acid as active/inactive vertex of the nucleosome interface, whereby all binary contacts between DNA and the histone fold are accounted for.



Via a further, low-dimensional reduction effort, we define the *local robustness* of the nucleosome interface as the ND-dimensional array of the row-wise maxima of the joint DH matrix (FIG. 5.1), i.e. the maximal contact persistence associated with each nucleotide. By a similar criterion, we define the **marginal local robustness** by extracting the row-wise maxima from isolated sub-blocks of the joint DH matrix (FIG. 5.1), thereby achieving tai-

lored ND-dimensional arrays to account for the maximal contact persistence between each nucleotide and i) the helix fold domain (section 5.3.3), ii) the helix fold extensions and histone tails domain (section 5.3.1) or iii) isolated histones.

As for the dynamics of the reference nucleosome, we located stretches of nucleotides that are latched onto the core histone bundle, as well as “blind” nucleotides, which exhibit low to no direct contact with the core histones (*vide infra*).

5.2.1 Hierarchical clusters analysis

Matrices, sub-blocks and arrays of maximal contact persistence are flexible dataframes to define the overall behavior of the nucleosome fold. As well, these objects are “discretizable” over time, so that we may describe the collective behavior of a sub-network or subset interface at discrete intervals of the nucleosome trajectory.

To describe the erratic tumble of the N-terminal tail of histone H4' over the nucleosome fold (section 5.3.4), we subdivided the trajectory of the reference nucleosome into discrete timeframes (20-nanosecond blocks) and associated each timeframe with the joint DH sub-matrix of the maximal contact persistence between the DNA and the H4' N-tail².

We thus calculate the *Frobenius distance* between 20-nanosecond blocks as:

$$d_{AB} = N^{-1} \sum_{i,j} \frac{|A_{ij} - B_{ij}|}{|A_{ij} + B_{ij}|} \quad (5.1)$$

where A , B are joint DH sub-matrices from the trajectory timeframes; the summation runs over $i = \{1 \dots 374\}$ nucleotides and $j = \{1 \dots 26\}$ amino acids of the N-terminal tail of histone H4'; N is a normalization factor, to account for all binary contacts between DNA and the H4' N-tail.

The two-dimensional matrix of Frobenius distances d_{AB} thus achieved describes the affinity between the sub-matrices associated with the residue-wise

²Each joint DH sub-matrix thus defines the maximal persistence of all binary contacts between the nucleotides and the twenty-six distal amino acids of the N-terminal tail of histone H4'- extracted recursively within 20-nanosecond timeframes of the trajectory of the reference nucleosome.

persistence of a contact interface, which is assessed at discrete intervals in time.

Moreover, as all d_{AB} elements of equation 5.1 satisfy i) $x_{ii} = 0$, and ii) $x_{ij} = x_{ji} \geq 0$, the matrix of Frobenius distances describes an effective affinity metrics between trajectory blocks, whereby we clustered the trajectory of the reference nucleosome upon the erratic tumble of the N-terminal tail of histone H4'. As for the assessment of the **hierarchical clusters**, we employed the SCIKIT-LEARN [78] and SCIPY [79] *ad hoc* libraries and adopted an *unweighted average linkage criterion*³ (UPGMA) to relate the newly-achieved clusters.

Lastly, the RMSF routines of *MDAnalysis* [80] and the DO_X3DNA [81, 82] toolkit were employed to assess local DNA fluctuations.

5.3 Dynamics of the nucleosome interface

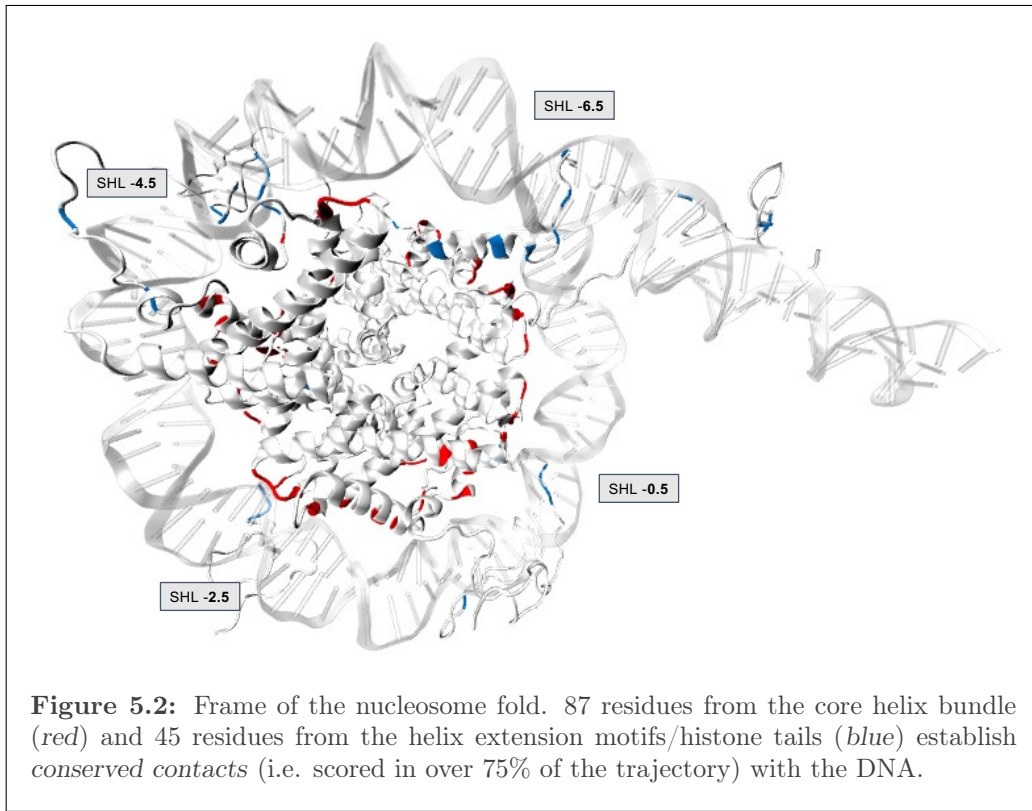
The nucleosome framework is associated with a conserved network of contacts at the interface between DNA and the core histone bundle [74, 83–86], which we refer to as the **nucleosome interface**. Out of two thousand residual interactions, 132 binary contacts⁴ are scored in over 75% of the trajectory of the reference nucleosome, hereafter referred to as **conserved contacts** (FIG. 5.2).

The subset of conserved contacts overall involves 88 amino acids, 75% of which are basic ARG and LYS residues, and 99 nucleotides. Thus, we observe that a major fraction of nodes at the nucleosome interface act as manifold *hubs* of electrostatic contacts and are located:

- at the fourteen *canonical* sites of the nucleosome framework, that is where inward DNA minor grooves host the side chain of basic ARG residues. This manifold layout of contacts holds the nucleosome fold, and is underneath

³The *unweighted average linkage criterion* defines the binary distance between newly-achieved clusters as average of the metrics of all elements of the cluster: If a system of three clusters (A, B, C) is described by arbitrary metrics as $A \Rightarrow B = 0.05$, $A \Rightarrow C = 0.15$, $B \Rightarrow C = 0.45$, the distance between the newly-achieved AB cluster and C amounts to 0.3.

⁴Shared between the helix fold core (87 contacts) and the helix fold extensions and histone tails domain (45 contacts).



the consensus mechanism whereby DNA *breathes* and/or is forcibly unraveled from the core histone bundle in discrete *tranches* of five nucleotides [85, 87, 88];

- at the exit channels, or *roots*, of the N-terminal tails of histones H3/H3' and H2B/H2B', where ARG/LYS-enriched stretches of amino acids interlock adjacent DNA gyres of the nucleosome fold [86, 89];
- where phosphate moieties of the DNA backbone act as reversible electrostatic links in-between basic amino acids, as observed by Chakraborty and Loverde [90].

5.3.1 The erratic dynamics of the histone tails

The domain of the *helix fold extensions* and *histone tails* of the NCP (FIG. 5.3) involves i) structure-less stretches of amino acids at the distal amino (all hi-

stones) and carboxyl-termini (histones H2A/H2A') of the core histones, or **tails**, which extend from the core helix bundle, and ii) unique α C and α N helical *extension* motifs of histones H2B/H2B' and H3/H3' respectively, at each surface of the NCP [83].

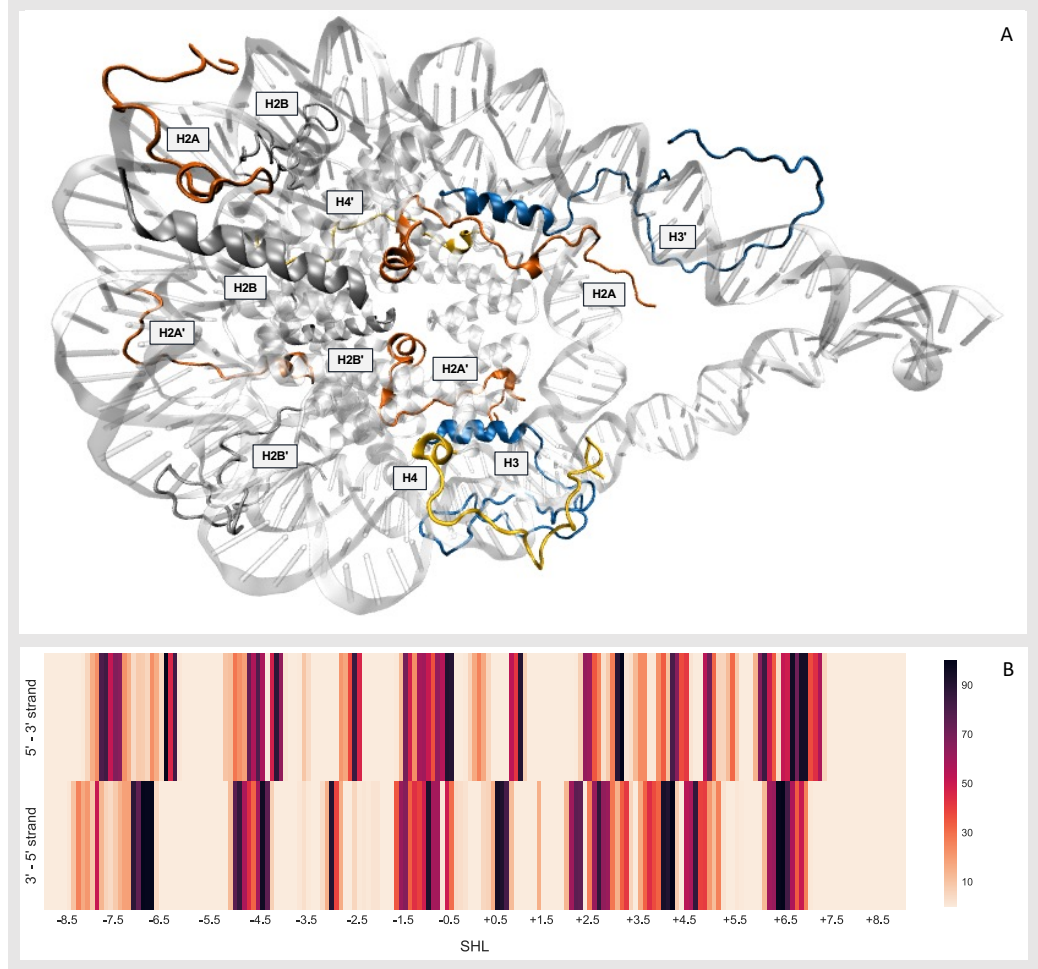


Figure 5.3: (A) Frame of the *helix fold extensions/histone tails* domain: the α N-helix motifs and N-terminal tails of histones H3/H3' (*blue*), the N-terminal tails of histones H4/H4' (*yellow*), the N-terminal (leftmost) and C-terminal tails of histones H2A/H2A' (*red*), and the α C-helix motifs and N-terminal tails of histones H2B/H2B' (*silver*). (B) Marginal local robustness (section 5.2) of the contact interface between DNA and the *helix fold extensions/histone tails* domain. Conserved contacts (*darker shades*) are associated with the roots of the H3/H3' and H2B/H2B' N-terminal tails, and the α N-helical motifs of histones H3/H3'.

While the helix fold core and helical extensions describe a semi-static framework for the DNA molecule to hold onto (section 5.3.3), the dynamics of the

histone tails is elusive and accounts for their blurrier role in the nucleosome fold. In fact, histone tails “tumble” about the NCP, and their motion is erratic and modulated by both the nucleosome core and the medium⁵.

Here, we observe that all histone tails adsorb reversibly onto the DNA, and collectively describe an asymmetrical and unbalanced layout of contacts at the two nucleosome moieties, i.e. across the twofold dyad axis (FIG. 5.3 (B)). Notably, it is reckoned that the erratic dynamics of symmetrically-related, or “twin”, tails accounts for the asymmetric breathing and dislocation of DNA from the nucleosome fold [74, 90, 91].

Moreover, no barrier-less, collective transitions of the histone tails are observed at the sub-microsecond scale, but a local reshuffle of the electrostatic contacts at the nucleosome interface [74, 85, 92]. Therefore, the initial, metastable layout of the histone tails drives the subsequent dynamics of the nucleosome fold.

Histones H3/H3'

The amino-terminal tails of histones H3/H3' mediate the breathing and access to nucleosomal DNA, as inferred by FRET/SAXS [93, 94] as well as *in silico* assessments⁶ [89, 92].

The twin H3/H3' N-tails stretch from the helix fold domain via narrow channels between DNA gyres, about the outward minor grooves at SHL ± 1 and the DNA entrance/exit sites at SHL ± 6.5 . The *roots* of both tails establish a conserved electrostatic interface with the DNA, via the basic ⁴⁰R_{YR}⁴² stretch of amino acids. The H3 tail thus curls between DNA gyres, whereas the twin H3' tail reversibly adsorbs onto the outer linkers.

Histones H4/H4'

The roots of the H4/H4' N-terminal tails lie at either surface of the NCP, about the inward DNA minor grooves at SHL ± 1.5 , and project the tails outwards of the nucleosome fold. In fact, these tails host manifold sites of

⁵Yet, it is controversial whether histone tails adsorb/fold about the acidic DNA backbone (consensus behavior) rather than leak from the NCP [3, 6].

⁶We refer the reader to Ghoneim *et al.* 2021 [6] as an exhaustive review on histone tails.

effective inter-nucleosome contacts [95–99], whereas their role in the NCP framework is fuzzier [93, 94].

Here, the H4' tail is lifted from the lower nucleosome surface and lands onto the DNA between SHL +2.5 and SHL +3.5, whereas the twin H4 tail is rotated about its root and towards the nucleosome dyad, and flanks the DNA between SHL -1.5 and SHL -0.5.

Histones H2A/H2A'

The short helical motifs at the roots of the N-terminal tails of histones H2A/H2A' lie about the inward DNA minor grooves at SHL ± 4.5 , at either surface of the NCP. These short tails - as well as the N-terminal tails of histones H2B/H2B' - exert a major kinetic barrier towards the forceful detachment of DNA from the nucleosome fold at SHLs ± 4.5 [74, 85, 87, 95, 100].

The N-tail of histone H2A thus lies over the outward DNA major groove at SHL -5, whereas the tail of the twin histone H2A' winds about the DNA between SHL +3.5 and SHL +4.5.

The tails at the carboxyl-termini of histones H2A/H2A' are buried effectively beneath the surfaces of the NCP and docked at the core helix bundle of the H3-H4 tetramer. As observed with histones H3/H3', these tails are projected towards the DNA entrance/exit sites and the dyad, and have been associated with the breathing of the outer DNA linkers [89, 90, 92, 101].

Here, the short (12 residues) distal stretches of the tails, which lie ahead of a brief ¹¹³SVLL¹¹⁶ helical motif, bind at different locations of the NCP. In fact, the H2A C-tail wobbles about the nucleosome dyad/DNA linkers, whereas the tail of histone H2A' establishes a conserved electrostatic interface with the DNA about the inward minor groove at SHL +6.5.

Notably, this twofold behavior of the H2A and H2A' carboxyl-tails accounts for two alternate layouts of the mirror canonical contact sites at SHLs ± 6.5 . In fact, the basic side chain of H2A'-K124 forces H3-R49 (a residue of the H3 α N-helix extension) out of the inward DNA minor groove at SHL +6.5 (FIG. 5.4), whereas H3'-R49 is embedded deeply within the mirror canonical

site at SHL -6.5⁷.

Histones H2B/H2B'

Lastly, the N-terminal tails of twin histones H2B/H2B' exit the helix fold via narrow channels, enclosed between the inward DNA minor grooves at SHL -4.5/+2.5 and SHL +4.5/-2.5 respectively. Both tails are locked between DNA gyres via a conserved electrostatic interface of the ²⁸KRRK³¹ basic *roots*. Close to the exit channel, K11, G13 and K16 force the distal stretch of the H2B N-tail into a metastable coiled layout (FIG. 5.4), whereas the twin H2B' N-tail alternately adsorbs and drifts from the NCP.

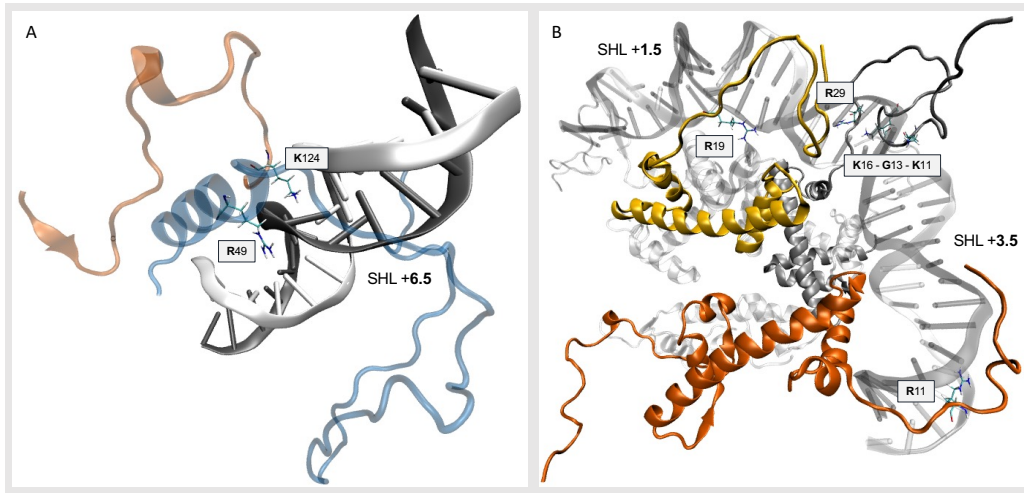


Figure 5.4: (A) Frame of the canonical contact site at SHL +6.5. The carboxyl-tail of histone H2A' (red) steers towards the DNA exit site at SHL +6.5, whereby H2A'-K124 forces H3-R49 out of the inward DNA minor groove. (B) Frame of the nucleosome interface between SHL +2/+4. The N-terminal tails of histones H4' (yellow) and H2B (silver) fold about the canonical contact site at SHL +2.5, and the N-tail of histone H2A' (red) is locked at the outward DNA minor groove at SHL +4 via the basic side chain of R11.

5.3.2 RMS fluctuations of the DNA fold

The asymmetric dynamics of twin histone tails accounts for an overall breakdown of the twofold rotational pseudo-symmetry of the NCP, which reflects on the behavior of DNA. The DNA overall fluctuates mildly about the core

⁷Shaytan and coworkers have made a similar observation [74].

helix bundle. The RMS fluctuations of the backbone ribose moieties (C1') exhibit alternate spikes about the canonical contact sites (FIG. 5.5), as observed by Shaytan and coworkers [74] and in line with the dynamics of the thermal/forceful detachment of DNA from the nucleosome fold, which is associated with low kinetic barriers that recur with a 5-10 nucleotides stride.

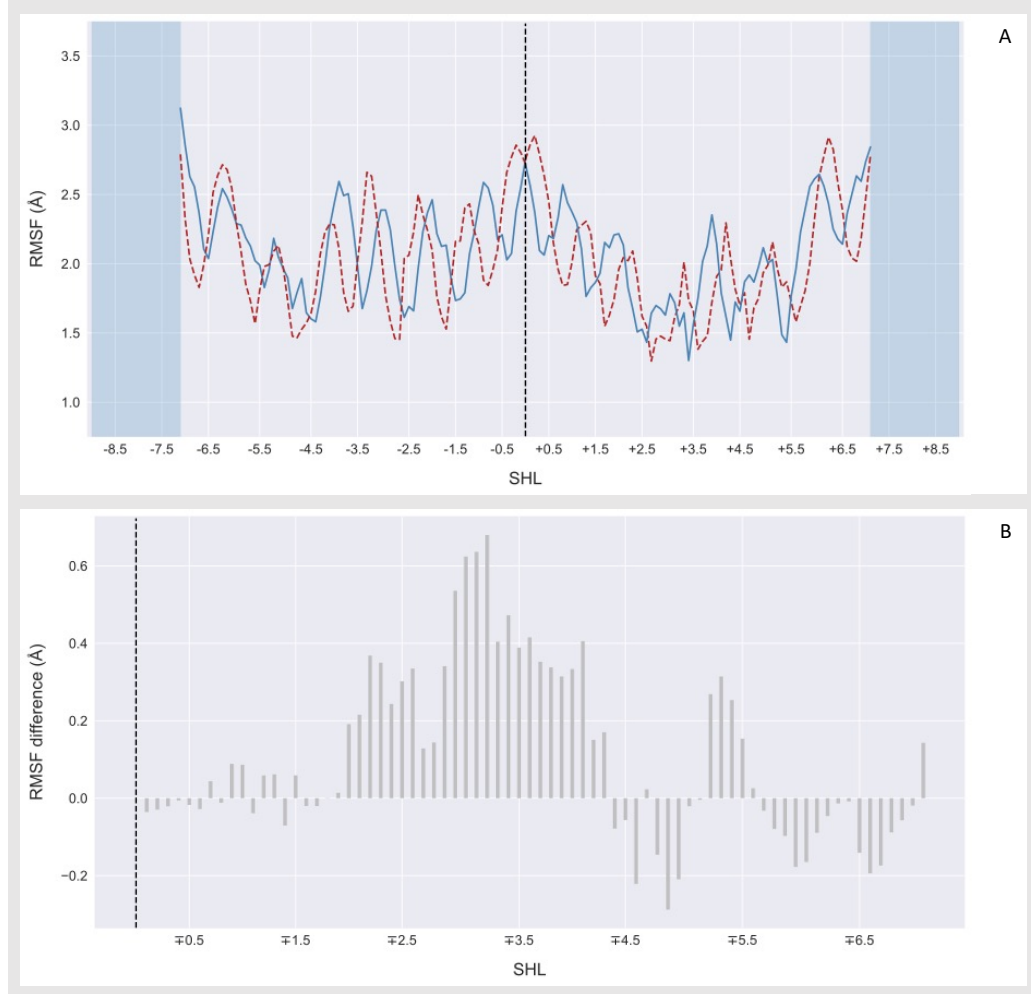


Figure 5.5: (A) RMS fluctuations of the ribose C1' atoms; the textures/colors are associated with mirror 5'-3' (*solid blue* lines) and 3'-5' (*dashed red* lines) DNA stretches across the dyad axis. Local minima in the RMSF characterize the recursive contact interface between DNA and the core histone fold - defined by fourteen *canonical* sites. (B) Relative difference of the mean RMSF between symmetrical nucleosome moieties, i.e. across the dyad axis - *dashed black* line in (A). The RMSF imbalance at SHL $\pm 2/\pm 4$ is mediated by the dynamics of the terminal tails of twin histones about the NCP.

Moreover, the DNA is looser about the entrance/exit sites [74, 86–89, 100],

in spite of the constraints enforced to the outer linker moieties. In fact, it is reckoned that the transcription machinery rectifies the *impromptu* dislocation of the DNA from the nucleosome fold, to circumvent the kinetic barrier exerted by the nucleosome interface [102, 103].

Notably, the RMS fluctuations of the DNA at SHL $\pm 2/\pm 4$ differ between nucleosome moieties (i.e. across the dyad axis - FIG. 5.5), which we associated to the local dynamics of twin histone tails. In fact, both DNA stretches (between SHL +2/+4 and between SHL -2/-4) lie within reach of the close N-terminal tails of histones H2A/H2A' and H2B/H2B', and of the extended H4/H4' N-tails. However, no tail hovers between SHL -2/-4 (but the roots of histone H2B'), whereas three N-terminal tails are adsorbed onto the DNA at SHL +2/+4 (FIG. 5.4):

- K16, K11 and G13 lock the N-terminal tail of histone H2B between SHL +2.5 and SHL +3;
- a conserved contact between H4'-R19 and the inward DNA minor groove at SHL +2.5 steers the H4' N-tail towards the rear nucleosome sites (SHL +2.5/+3.5);
- likewise, the basic side chain of H2A'-R11 is embedded deeply within the outward DNA minor groove at SHL +4, whereby the short N-terminal tail of histone H2A' is driven towards SHL +3.5.

This threefold (reversible) layout buffers the local fluctuations of the DNA between SHL +2/+4, and accounts for a 0.2 - 0.6 Å RMSF imbalance between the nucleosome moieties (FIG. 5.5).

Moreover, this mechanical role mirrors the asymmetric control and accessibility to the DNA in the NCP exerted by the histone tails, as well as earlier observations by Shaytan and coworkers, who have associated the lack of an effective interface between DNA at the entrance/exit sites and the tails of histones H3/H2A with local, defective states of the nucleosome fold [74].

5.3.3 The canonical contact sites are affected by the imbalance of the nucleosome fold

The core helix fold is a bundle of interwoven α -helices and elicits a recursive interface of conserved contacts, or **canonical sites**, whereby the DNA holds onto the nucleosome framework (FIG. 5.6).

These twelve canonical sites involve either of two contact *motifs* [84, 95], which hold the walls of the inward DNA minor grooves via either i) L1/L2 loops, at the interface of histones H3/H4 and of histones H2A/H2B (SHLs ± 0.5 , ± 2.5 , ± 3.5 and ± 5.5), or ii) the cross N-termini of the core $\alpha 1$ -helices (SHL ± 1.5 and ± 4.5). At all canonical sites, the side chain of a basic amino acid is embedded deeply within the minor groove “niche”, flanked by ancillary and/or water-mediated hydrogen bonds and electrostatic contacts that involve both the DNA backbone and the nucleic bases. This recursive interface enforces mechanical stalls on the transcriptional activity of the RNAP machinery in the NCP [104–106], and has been observed in nucleosomes and sub-nucleosomal assemblies alike [107–110].

As observed with earlier *in silico* assessments of the NCP [85, 86], the network of contacts at the interface between DNA and the core helix bundle is replicated by each DNA strand (FIG. 5.6), on account of the twofold rotational axis of the NCP across the dyad, whereby twin amino acids and nucleotides are involved at both nucleosome moieties. However, we assessed subtle discrepancies at twin contact sites, as described below.

SHL ± 4.5

Three ARG residues (H2B-R33, H2A-R29 and H2A-R32) lock the inward DNA minor groove at SHL -4.5 by a metastable “arc-like” layout (FIG. 5.7), whereby H2A-R29 is embedded deeply within the minor groove niche. Such layout is seldom observed at SHL +4.5, and H2A'-R29 and H2A'-R32 are rather driven from the contact site and towards the acidic H2B'-E35 carboxylate moiety (FIG. 5.7). In fact, we observed twice the score of conserved contacts between DNA and the core helix bundle at SHL -4.5 than at SHL +4.5.

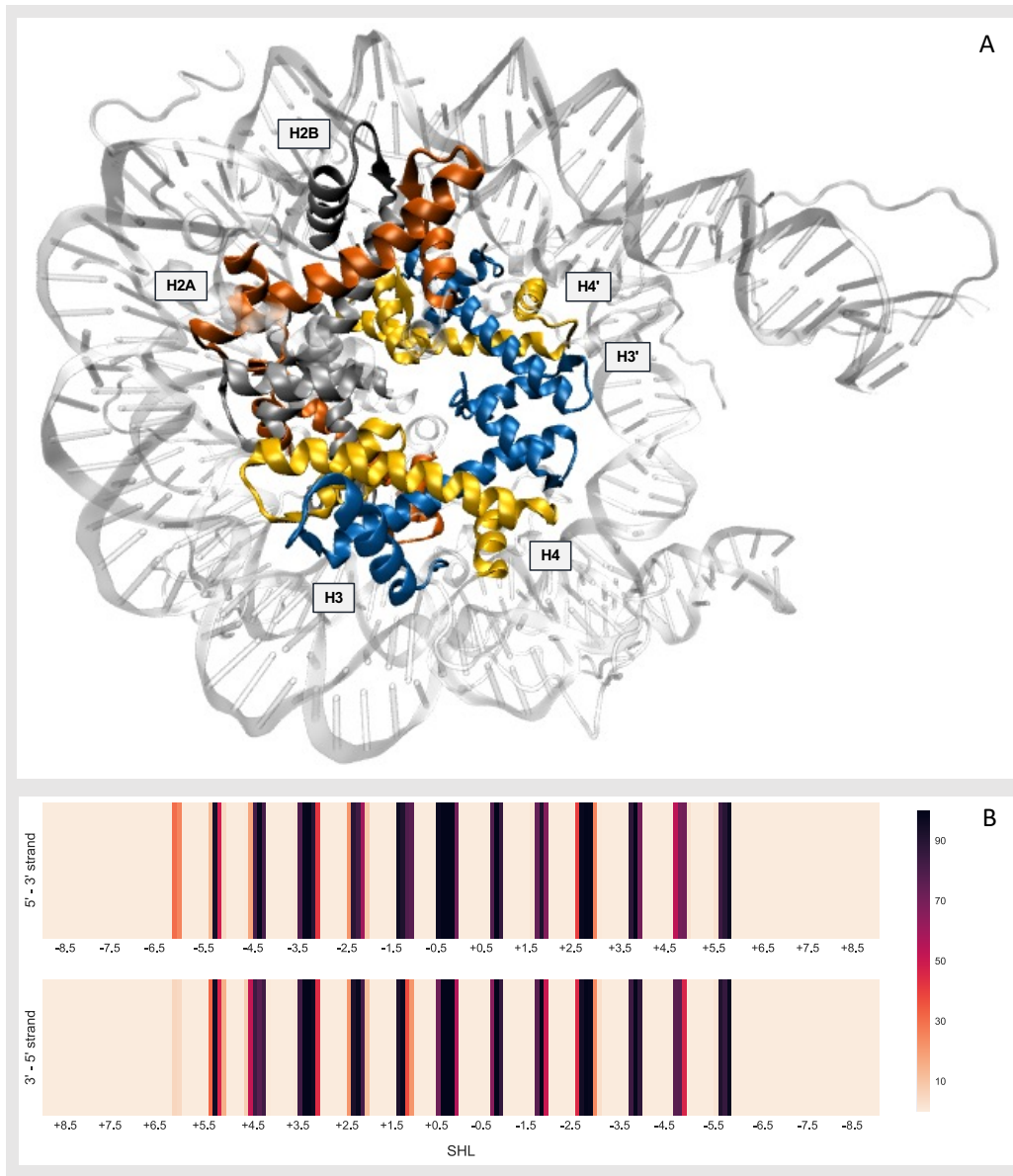


Figure 5.6: (A) Frame of the *helix fold* domain: histones H3/H3' (blue), H4/H4' (yellow), H2A/H2A' (red) and H2B/H2B' (silver). (B) Marginal local robustness (section 5.2) of the contact interface between DNA and the *helix fold* domain. The recursive interface is replicated between the DNA strands; however, subtle discrepancies arise at SHL ± 1.5 and ± 4.5 , modulated by the local dynamics of the histone tails.

SHL ± 1.5

The contact interface about the twin SHL ± 1.5 canonical sites is overall alike. The side chain of R63 of both histones H3 and H3' is embedded within the

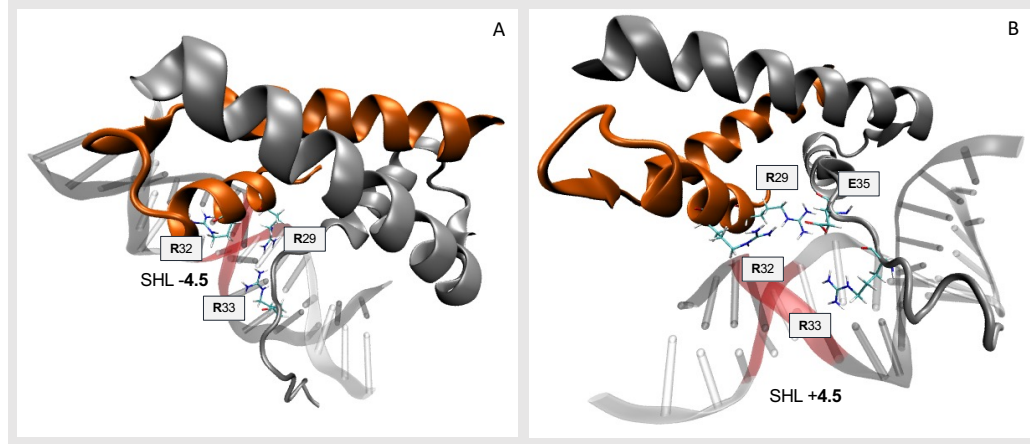


Figure 5.7: Frame of the contact interfaces between DNA and the helix fold about the inward DNA minor grooves at SHL ± 4.5 . Both interfaces are defined by two alternate layouts: (A) an “arc-like” lock of the basic side chains of H2B-R33, H2A-R29 and H2A-R32 (major layout at SHL -4.5), and (B) a loose layout, where H2A'-R29 and H2A'-R32 are steered towards the acidic H2B'-E35 carboxylate moiety (major layout at SHL +4.5).

inward DNA minor grooves at SHL ± 1.5 , flanked by i) conserved hydrogen bonds between the backbone amide moieties of H3 and H3'-L65 - which lie at the N-termini of the $\alpha 1$ -helices of histones H3 and H3' - and the walls of the minor grooves, and ii) electrostatic contacts that involve the basic R36 residues at the $\alpha 1$ -helices of both histones H4 and H4'.

Three further amino acids, H4/H4'-R55, H4/H4'-K31 and H3/H3'-R69, hover about the inward DNA minor grooves at SHL ± 1.5 . All residues establish conserved electrostatic contacts with the walls of the minor groove niche at SHL -1.5 (FIG. 5.8), whereas no effective contact with the DNA that involved H4'-R55, H4'-K31 or H3'-R69 is observed at SHL +1.5.

Here, we reckon that the dynamics of the N-terminal tails about the two canonical sites accounts for the local contact imbalance observed at SHL +1.5/-1.5 (FIG. 5.8). In fact, the N-terminal tails of both histones H3 and H4 are adsorbed onto the DNA between SHL -1.5 and -0.5, whereas their twin H3' and H4' N-tails stretch over and towards the linker DNA moieties and rear nucleosome sites (SHL +2.5/+3.5) respectively. Thus, the DNA is looser about the core helix bundle at SHL +1.5 than at SHL -1.5, and out of reach of the basic moieties of H4'-R55, H4'-K31 and H3'-R69.

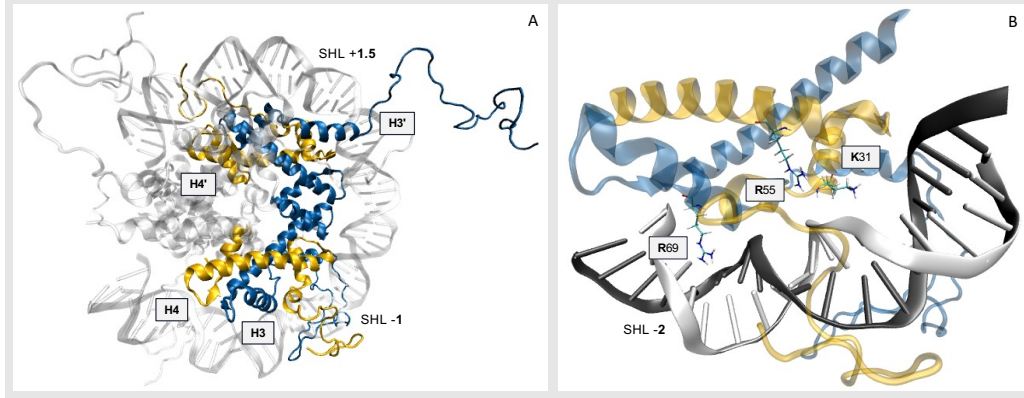


Figure 5.8: (A) Frame of the nucleosome interface about the canonical contact sites at SHL ± 1.5 . The observed local imbalance in the score of resilient contacts at SHL $+1.5/-1.5$ (FIG. 5.6) is accounted for by the behavior of the N-terminal tails of twin histones H3/H3' and H4/H4'. (B) Frame of the contact layout between DNA and the helix fold at the inward minor grooves at SHL ± 1.5 . As the DNA molecule is retained by the N-terminal tails of histones H3 and H4, ancillary electrostatic contacts are established at SHL -1.5 (H3-R69, H4-K31 and H4-R55), which are not observed at SHL $+1.5$.

SHL ± 0.5

Likewise, the local dynamics of twin histone tails underlies minor discrepancies about the dyad site (SHLs ± 0.5). Both canonical sites are crowded with conserved contacts at the interface with the core helix bundle, which tallies with the traces of major stalls of the transcriptional machineries about the dyad [87, 104, 106].

However, we assessed that H4'-K44 establishes a conserved electrostatic contact with the inward DNA minor groove at SHL $+0.5$, whereas H4-K44 does not. Both basic residues lie at the L1 loop of histones H4' and H4, and bind the backbone moieties of the close H2A/H2A'-L115 residues, about the brief $^{113}\text{SVLL}^{116} 3_{10}$ helical motif of the H2A/H2A' carboxyl-tails (FIG. 5.9). However, while the distal tail of histone H2A adsorbs over the DNA about the dyad site, thereby steering the basic side chain of H4'-K44 with, the twin H2A' carboxyl-tail folds back and drives H4-K44 from the canonical site at SHL -0.5 (FIG. 5.9).

We observed no further (major) discrepancies between the contact interface of twin canonical sites. Shaytan and coworkers [74] have corroborated

in silico earlier inferences of a hierarchy between the canonical sites of the nucleosome fold [95], upon the local score of conserved contacts and residual contact surface between the DNA and the core helix bundle. Here, we infer an enhanced sensitivity of $\alpha 1\alpha 1$ sites of the NCP ($\text{SHL} \pm 1.5$ and ± 4.5) towards the dynamical imbalance of the nucleosome fold, mediated by the local tumble of the histone tails.

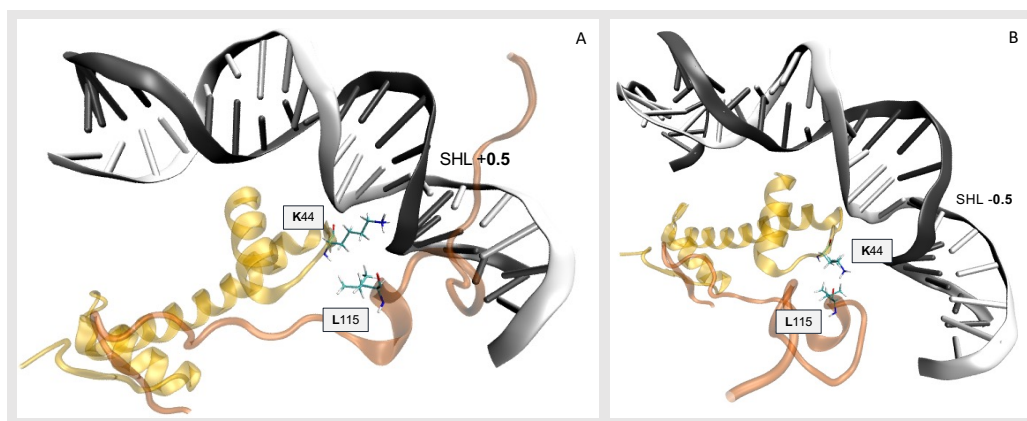


Figure 5.9: Frame of the contact layout between DNA and the helix fold at the inward minor grooves at $\text{SHL} \pm 0.5$. The carboxyl-terminal tails of histones H2A and H2A' steer H4'/H4-K44 towards and from the dyad at $\text{SHL} +0.5$ (A) and $\text{SHL} -0.5$ (B) respectively, via a conserved contact with the backbone moieties of H2A/H2A'-L115.

5.3.4 Histone tails act as reversible allosteric switches of the nucleosome framework

We corroborated the idea that the erratic dynamics of the histone tails biases the micromechanical behavior of the nucleosome fold via the assessment of the allosteric routes across the core helix fold⁸. In fact, it is reckoned that the helix fold bundle may convey a negatively-correlated breathing motion of the linker DNA moieties [90, 109, 112]; likewise, the thermal/forceful relocation of DNA about the nucleosome fold is mediated by a concerted strain of the NCP [89, 113–116].

The *PyInteraph* workflow (section 4.3) extracts the *shortest allosteric routes*

⁸Here, we relate to the nucleosome fold as an allosteric framework (as done in Bowerman and Wereszczynski 2016 [111]), not as an allosteric *effector* (via PTMs of the histone tails, for instance).

between actively connected nodes of a macromolecular framework via a step-wise routine [64]:

- the threefold adjacency matrices of (hydrophobic and electrostatic interactions, and hydrogen bonds) intermolecular contacts are filtered upon a threshold criterion of contact significance⁹;
- the filtered matrices are thus merged, to achieve a boolean (unweighted) dataframe of actively connected nodes, where the *shortest allosteric routes* between remote nodes are defined.

We assessed two “mirror” tracks across the helix fold, hereafter **TR-1** and **TR-2**. These tracks are related via the twofold rotational axis of the nucleosome fold, and connect the DNA entrance/exit sites (SHL -6.5/+6.5) and the roots of the N-terminal tails of histones H2B'/H2B (SHL -2.5/+2.5), across the core helix bundle (FIG. 5.10).

Both the twin H2B'-R29 and H2B-R29 residues establish a conserved interface with the DNA at SHL -2.5/+2.5 and retain the N-terminal tails of histones H2B'/H2B between DNA gyres. TR-1 and TR-2 differ, however, at the DNA entrance/exit sites, as H3'-R49 and H2A'-K124 are embedded within the inward DNA minor grooves at SHL -6.5 and SHL +6.5 respectively.

Track 1

Seven *shortest* routes (19 nodes) link the termini of track TR-1 - that is, H3'-R49 (SHL -6.5) and H2B'-R29 (SHL -2.5). These routes cross the core helix fold via the H3-H4 tetramer and describe a branched allosteric network (FIG. 5.11). Such network relies on conserved *intra*- and *interhelix* contact lines, at the core of the helix fold bundle, except for labile nodes/edges at the interface of histones H3/H4 and of histones H3'/H4'.

In fact, five out of seven routes at TR-1 (*red* routes in FIG. 5.11) involve the distal A135 residue at either of the short H3/H3' carboxyl-termini. These twin terminal amino acids behave as both hydrophobic and electrostatic

⁹Such threshold lies about the flex of curves of *maximal extent of connected sub-graphs* v. *minimal contact persistence* (FIG. 5.10), whereby all binary interactions above this threshold satisfy a criterion of contact significance. It has been set to 15% of the trajectory timeframes in this work.

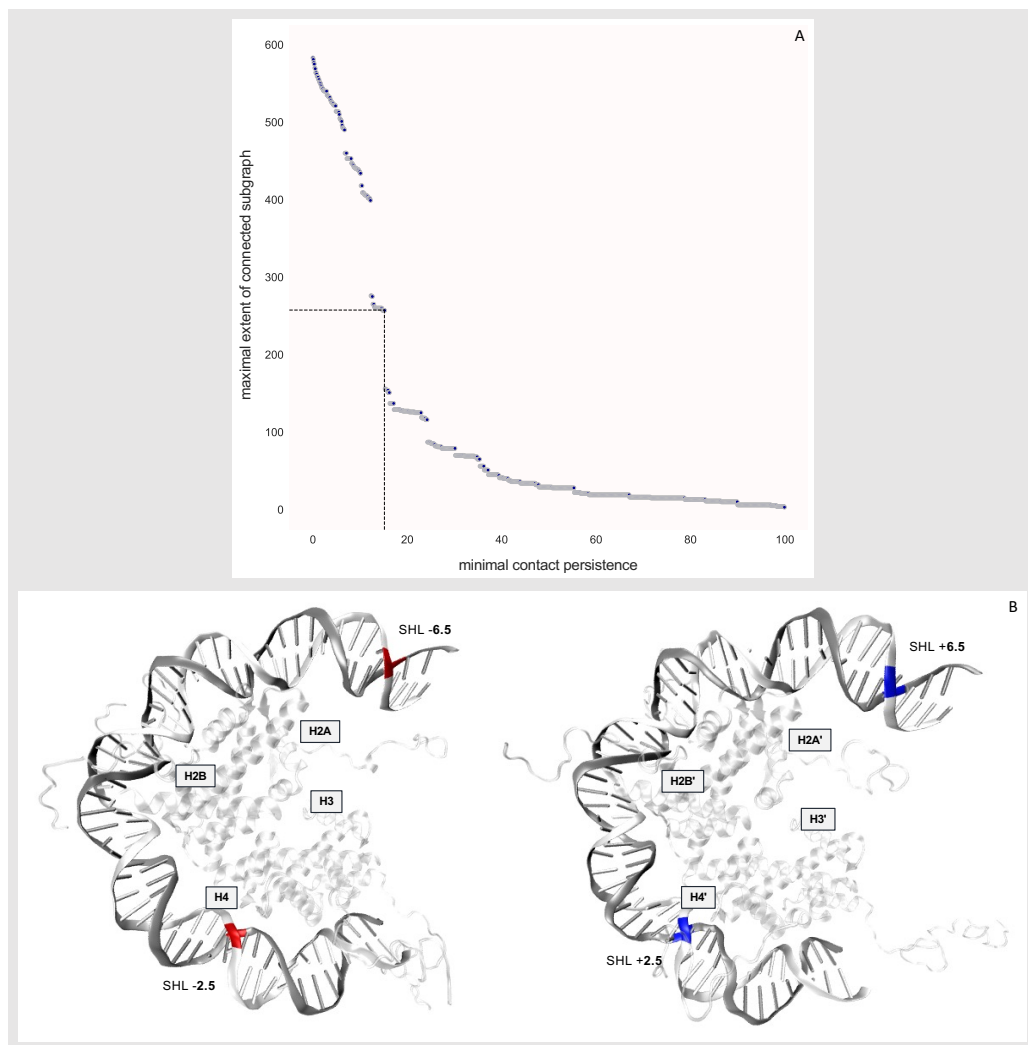


Figure 5.10: (A) Curve of maximal extent of connected sub-graphs v. minimal contact persistence - here associated with the class of electrostatic contacts. About the flex of the curve (i.e. at a minimal contact persistence of 15% of the trajectory timeframes) has been established the threshold for contact significance between connected nodes. (B) Termini of the “mirror” TR-1 (red) and TR-2 (blue) allosteric tracks across the core helix bundle of the NCP.

(acidic) contact nodes, at the bundled interfaces of dimers H3/H4 and H3'/H4'. However, both H3/H3'-A135 residues rather “shuttle” between H3/H3'-R129 and H4/H4'-V60 (yellow nodes in FIG. 5.11) than close the allosteric lines effectively.

Two alternate routes exist at TR-1, nevertheless, where the allosteric lines

are diverted towards conserved H3'/H4' inter-helix edges (*silver* routes in FIG. 5.11), which circumvent the H3/H3'-A135 carboxyl-termini via the direct interface between the H3'- α 3 and H4'- α 2 helices (FIG. 5.11).

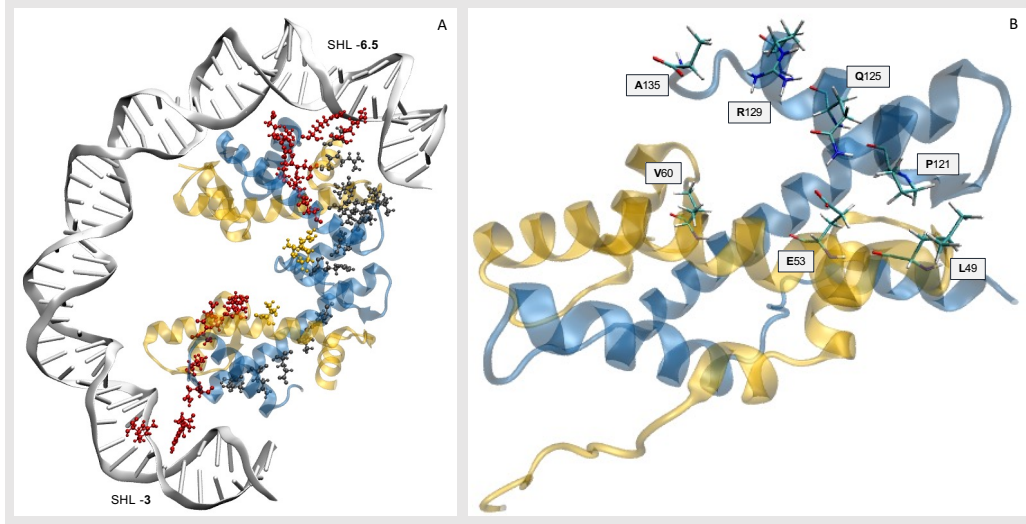


Figure 5.11: (A) Shortest allosteric routes (19 nodes) at track TR-1, i.e. between the root of the N-terminal tail of histone H2B' (SHL -2.5) and the DNA entrance site (SHL -6.5). Five out of seven routes (*red* nodes) share a labile edge, at the contact interface of either dimer H3/H4 or dimer H3'/H4' (*yellow* nodes, and the R129-A135-V60 residues in (B)), associated with the motile H3-A135 and H3'-A135 termini of the short H3 and H3' carboxyl-tails. Two alternate routes (*silver* routes) circumvent this “liability” via the direct contact of the H3'- α 3 and H4'- α 2 helices (B).

Track 2

An artifact similar to that observed at the interface of both dimers H3/H4 and H3'/H4' unfolds at TR-2, that is, the mirror track between H2A'-K124 (SHL +6.5) and H2B'-R29 (SHL +2.5). Here, the two shortest routes (18 nodes) wind across the H3 α N-helix and towards the helical core of histone H2B', via the docking domain of the H2A' carboxyl-tail (*red* routes in FIG. 5.14 (A)). We found, however, that an artificial link between H2B'-E93 and GUA66 at SHL +2.5¹⁰ thus closes the two routes, mediated by R3 at the distal terminus of the H4' N-tail (*yellow* nodes in FIG. 5.14 (A)).

Such artifact arises from the criteria underneath the definition of the con-

¹⁰H2B'-E93 and GUA66 lie at about 30 Å

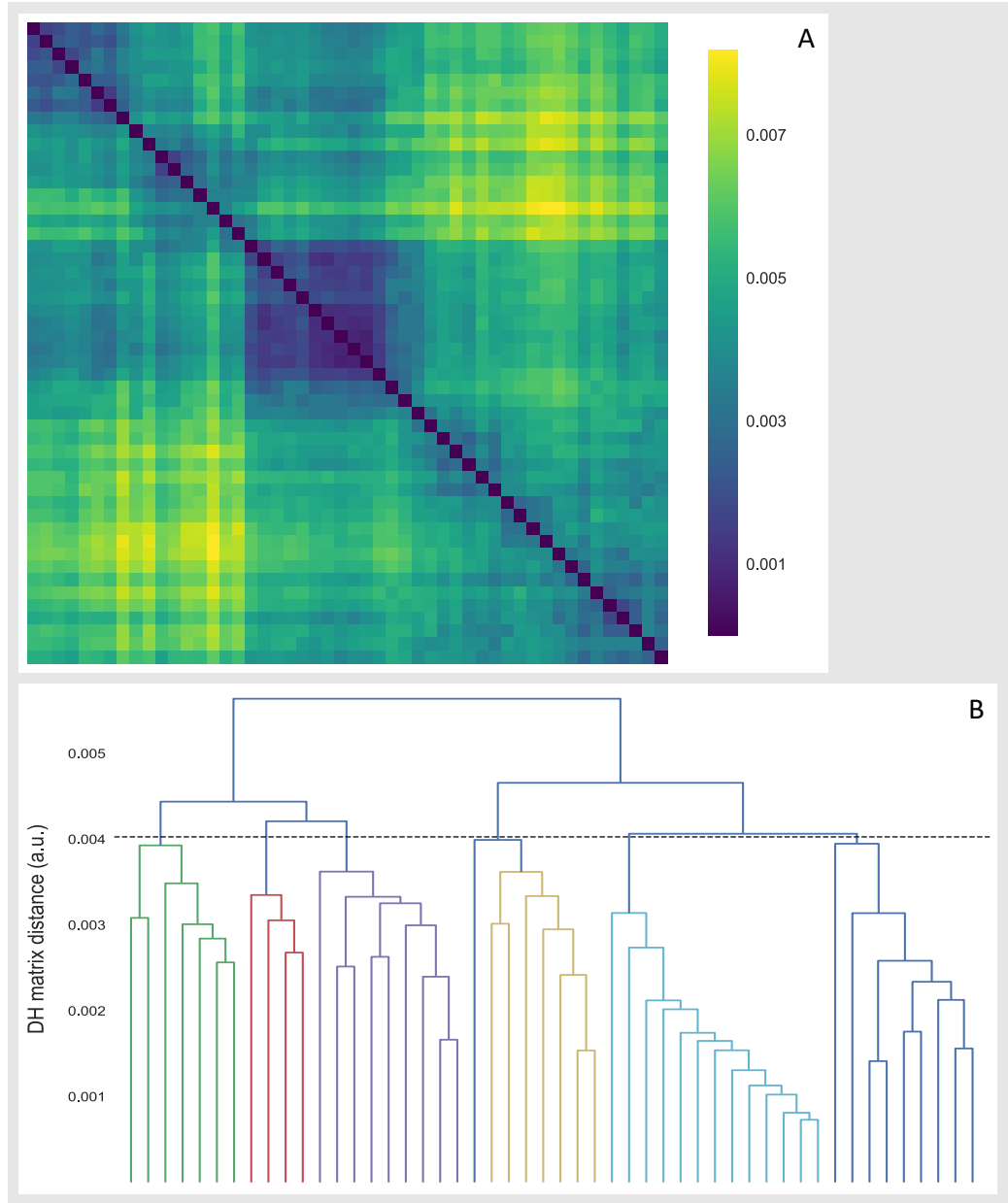


Figure 5.12: (A) Outline of the all v. all Frobenius distance between joint DH matrices of the contact interface between DNA and the N-terminal tail of histone H4', associated with each of the 20-nanosecond blocks of the trajectory of the reference nucleosome. (B) Hierarchical cluster assessment of the trajectory of the reference nucleosome, upon the contact interface between DNA and the H4' N-terminal tail - as described in section 5.3.4.

nected network, as well as from the erratic tumble of the histone tails. In fact, the *PyInteraph* workflow [64] establishes that a minimal contact persistence

between nodes be exceeded whereby a binary edge is a *significant contact* - which we adopted as 15% of the trajectory timeframes (FIG. 5.10 (A)). As the N-terminal tail of histone H4' wobbles about the nucleosome fold, i.e. H4'-R3 hovers over both H2B'-E93 and GUA66 at subsequent stages in the trajectory of the reference nucleosome (*vide infra*), *PyInteraph* mistakes H4'-R3 for an effective link between histone H2B' and the DNA.

We thus further assessed the allosteric bias from the tumble of the N-terminal tail of histone H4' as follows:

- the trajectory of the reference nucleosome was divided into 20-nanosecond blocks, and each block was associated with the joint DH matrix of the contact interface between DNA and the twenty-six distal amino acids of the N-terminal tail of histone H4';
- the trajectory was thus clustered upon the all-v-all Frobenius distance between joint DH matrices (section 5.2.1 and FIG. 5.12), whereby we extracted six conformational clusters to describe the behavior of the H4' N-tail - as outlined in TABLE 5.1 (FIG. 5.13).

At first, the N-terminal tail of histone H4' is folded over a short stretch of nucleotides, between SHL +2/+2.5, and the basic side chain of H4'-R3 is embedded therein (cluster **2**, and the *red* tail layout in FIG. 5.13).

cluster	timeframe (ns)	E105-R129	E93-R3-E74
2	< 180	on	off
0	180 - 340	on	off
5	340 - 600	off	on
1	600 - 740	on	off
4	740 - 820	off	off
3	> 820	on	off

Table 5.1: Outline of the six conformational clusters of the N-terminal tail of histone H4' and status of the E105-R129 and E93-R3-E74 allosteric *switches*.

However, the tail loosens its hold onto the DNA between 180 and 335 nanoseconds (cluster **0**), and the distal stretch of amino acids is released and steered towards the core helix fold (cluster **5**, and the *blue* tail layout in FIG. 5.13). Thus, H4'-R3 lands between the acidic H2B'-E93 and H4'-E74 moieties, at

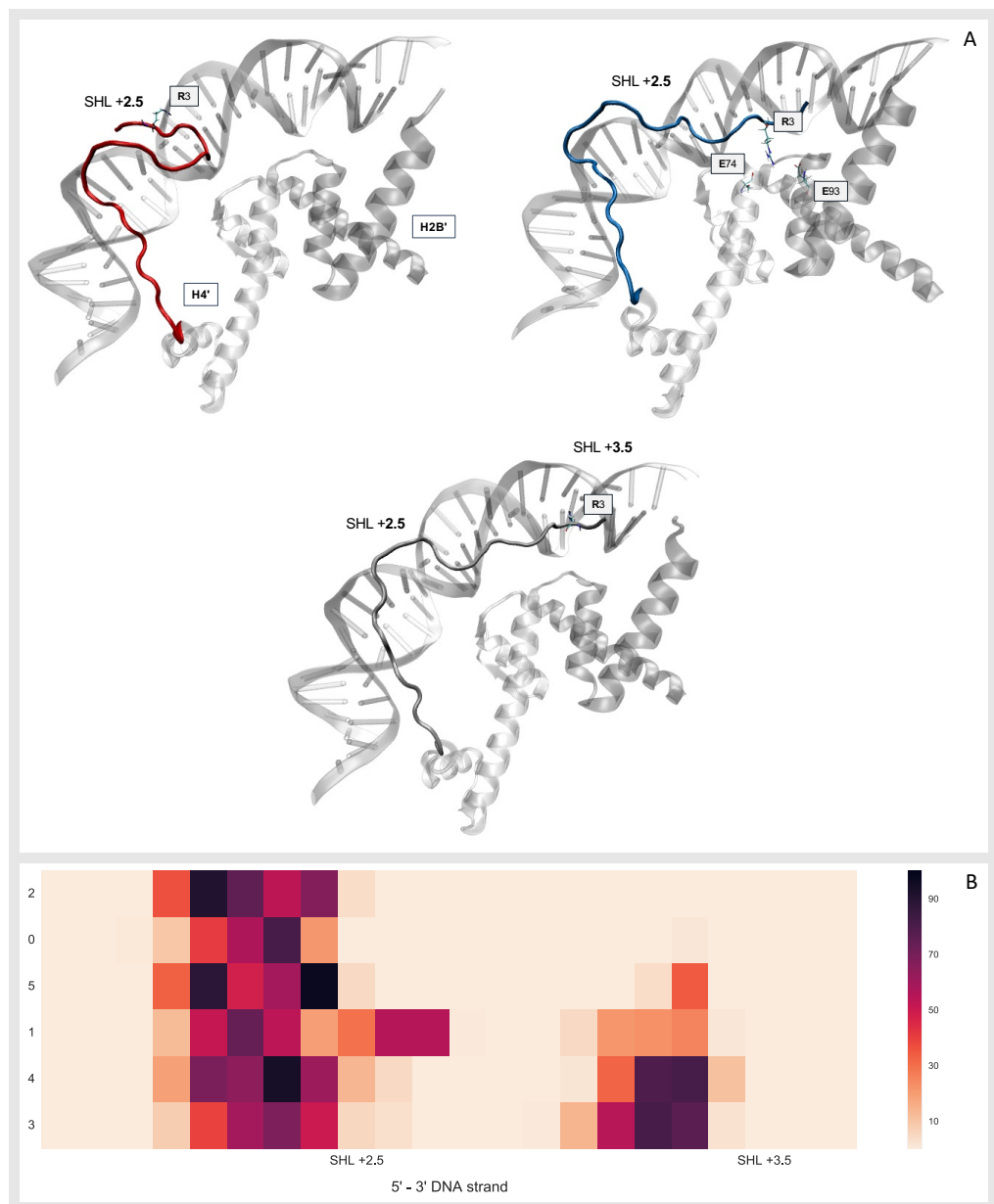


Figure 5.13: (A) Mean layout of the N-terminal tail of histone H4' at conformational clusters **2** (red tail), **5** (blue tail) and **4** (silver tail) respectively, and location of the H4'-R3 allosteric switch. (B) Marginal local robustness of the contact interface between DNA and the N-terminal tail of histone H4', associated with each of the conformational clusters of the H4' N-tail. The reversible adsorption of basic ARG/LYS residues onto the DNA molecule accounts for the elusive dynamics of the histone tail domain. Moreover, the H4' N-tail loosens its hold onto the DNA between local tumbles (clusters **0** and **1**), whereby its contact interface with the DNA is enhanced steadily.

the four-helix bundle interface of histones H4' and H2B'.

The twofold E93-R3-E74 contact overall holds for a few hundred nanoseconds; it is broken abruptly, however, as the N-terminal tail of histone H4' bends aside and towards the DNA, and reaches over the inward DNA minor groove at SHL +3.5 (clusters **1**, **4** and **3**)¹¹.

This behavior of the N-terminal tail of histone H4' tallies with the observations by Armeev and coworkers [89], whereby histone tails enhance the contact interface with DNA over time (FIG. 5.13), as well as with the observed sub-microsecond tumble imbalance between the static/inner (residues 16-24) and motile/outer (residues 1-15) stretches of the H4 N-tails [6, 117, 118]. Moreover, Rabdano and coworkers [118] have inferred that H4-R3 behaves as a “latch” of the H4 N-tail, via both NMR and molecular dynamics assessments.

The shortest allosteric routes associated with TR-2 wind across the H3-H4 tetramer fold in all conformational clusters of the H4' N-tail but cluster 5. Moreover, all major routes assessed at clusters **2**, **0**, **1** and **3** (20 nodes) outline a robust allosteric network, across the inter-helix contacts that flank the extended α 2-helices of histones H3 and H4' (*red* routes in FIG. 5.14 (B)).

As observed with TR-1, however, all routes share a labile electrostatic edge between H3-E105 and H3'-R129, i.e. at the four-helix bundle interface of histones H3 and H3' (*yellow* nodes in FIG. 5.14 (B)), which behaves as a binary allosteric **switch**. In fact, the electrostatic contact between E105 and R129 is *idle* (or active in less than 15% of the timeframes) between 740 and 820 nanoseconds, and the allosteric network underneath is “silent” at cluster **4**. Thus, all routes are diverted to a 23-node line across DNA gyres, which involves a forwarded link between SHL +6.5 and SHL -0.5 via the embedded root of the N-terminal tail of histone H3 (*silver* routes in FIG. 5.14 (B)).

A different scenario unfolds at cluster **5** - that is, between 340 and 600 nanoseconds. In fact:

- the core H3-H3' four-helix bundle lacks the active contact between H3-

¹¹These clusters overall share a similar layout of the H4' N-tail (the *silver* tail layout in FIG. 5.13) within local, minor distortions.

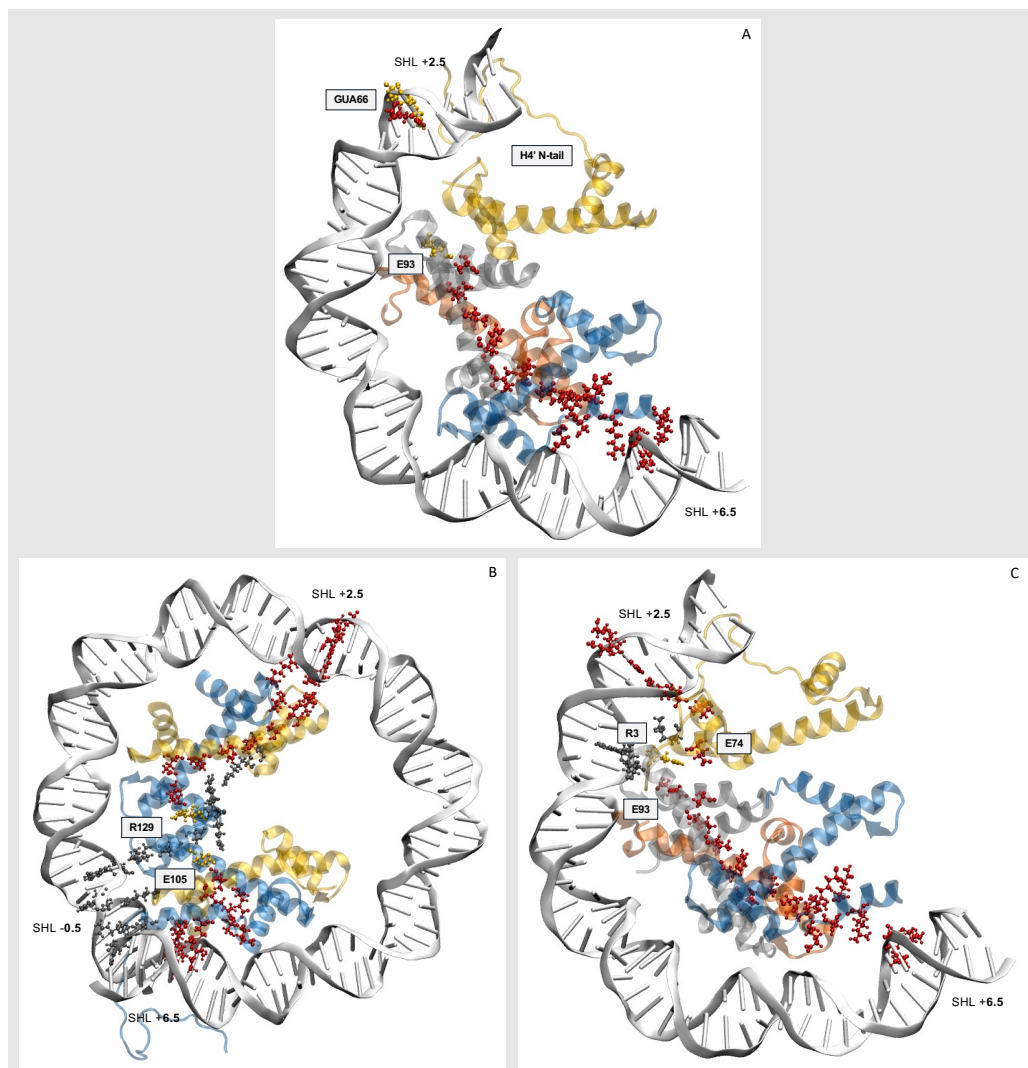


Figure 5.14: (A) Shortest allosteric routes (18 nodes) at track TR-2, that is, between the root of the N-terminal tail of histone H2B (SHL +2.5) and the DNA exit site (SHL +6.5). An artificial link is established between H2B'-E93 and GUA66 (yellow nodes) via H4'-R3 (at the distal terminus of the H4' N-tail), which satisfies the criteria of *PyInteraph* for contact significance. (B) Shortest allosteric routes (20 nodes) associated with the (2, 0, 1, 3) conformational clusters of the N-terminal tail of histone H4' (red routes). At cluster 4, the E105-R129 switch (yellow nodes), at the H3-H3' four-helix bundle, is idle and allosteric routes are diverted towards a 23-node line across DNA gyres. (C) Shortest allosteric routes (21 nodes) associated with conformational cluster 5 of the N-terminal tail of histone H4' (red routes). The H4'-R3 switch (yellow node) enables an alternate route in-between H2B'-E93 and H4'-E74, across the H4'-H2B' four-helix bundle interface. The silver nodes divert the allosteric routes towards a side 23-node branch, as both E105-R129 and E93-R3-E74 switches are idle.

E105 and H3'-R129;

- the tumble of the distal stretch of amino acids at the N-terminal tail of histone H4' drives H4'-R3 at the interface of histones H4' and H2B' (*yellow* node in FIG. 5.14 (C)).

This layout accounts for a radical diversion of the allosteric routes from the H3-H4 tetramer. In fact, the twofold electrostatic contact of H4'-R3 between the acidic moieties of H2B'-E93 and H4'-E74 leads a backup 21-node route across the core helix fold, via the docking domain of the H2A' carboxyl-tail and the H4'-H2B' four-helix bundle interface (*red* routes in FIG. 5.14 (C)).

Notably, we found alternate, minor routes at cluster 4 (23 nodes), where the silent contact between H2B'-E93 and H4'-E74 is circumvented via the L2 motif of histone H2B' (*silver* nodes in FIG. 5.14 (C)).

5.4 Conclusive remarks

The nucleosome interface, that is, the contact interface between DNA and the core histone fold, has been assessed via all-atom molecular dynamics and an extension of the *PyInteraph* workflow [64] (which is dedicated to the dynamical extraction of connected graphs and allosteric networks), where we involved the nucleic acids of the nucleosome fold and detailed the manifold mechanical role of the histone tails in the NCP. We thus infer that:

- the erratic dynamics of the histone tails account for a loss of the twofold rotational pseudo-symmetry of the nucleosome fold, whereby i) the RMS fluctuations of DNA are modulated by the reversible adsorption of the tails onto the NCP, and ii) subtle discrepancies arise between twin canonical sites;
- histone tails are associated with a moderate tumble at the sub-microsecond scale: Labile, distal nodes (such as basic ARG/LYS residues) are landed stochastically about the nucleosome fold, which may reversibly enable or silence alternate allosteric routes between connected nodes.

Chapter 6

MD assessment of blunt double strand breaks

For decades, DNA has been at the core of a faceted, wide research effort; yet, to tackle the assessment of DNA *in silico* is non-trivial even at the lowest ranks of the chromatin framework. In fact, DNA is associated with a multi-framed dynamics, mediated by a balance of intermolecular forces and variable stimuli from the nuclear *milieu*, whereby DNA breathes, twists and slides over the nucleosome beads. Moreover, the tails of the core histones drift between nucleosomes and between chromatin fibers, and modulate the slow, mesoscale dynamics of the chromatin fold. As such, chromatin and DNA have been depicted *in silico* by a manifold theoretical framework, between the atomistic and the sub-cellular scale [119].

A wide share of works on DNA has been devoted to (reversible) **modifications** of canonical nucleotides, such as those associated with epigenetic modifications, oxidative and metabolic intermediates, and external noxious vectors. In this context, we take in the vast amount of *in silico* assessments of DNA lesions from radiations and radiolytic radicals, such as thymine glycols [120–122] and the 7,8-dihydro-8-oxoguanine [123–125], to cite but a few [126].

6.1 The DSB scenario in molecular dynamics

The microdosimetric coarse criteria that define the class of double strand breaks take in all DSB motifs within an arbitrary distance, and are thus weak in classical MD. However, the variety of DNA lesions enforced by the radiation field is stochastic [127, 128], so that a simplified scenario is invoked. In a twofold work, Cleri and coworkers have described the thermal and mechanical strain exerted by **staggered** double strand breaks (i.e. distanced by ≥ 1 BP), via all-atom MD assessments of short DNA linkers and of the nucleosome fold [48, 59], and have collectively inferred that:

- short-distanced DSBs alone may elicit the thermal fracture of the broken DNA moieties, within the effective timescales of the DDR machinery;
- the nucleosome fold delays the dislocation of DSB lesions by twenty-fold, so that the (lower bound) estimate of the thermal fracture of *staggered* DSBs in nucleosomes is about 100 microseconds;
- the active dynamics of the nucleosome fold about a lesion of a canonical contact site contrasts with the mild mechanical feedback from a lesion of an inward DNA major groove, of the dyad and of the DNA entrance/exit sites;
- as the electrostatic forces between DNA and the core histone fold overwhelm the release of mechanical strain (from the overly bent DNA layout of the NCP), a DNA moiety that has been drawn forcefully from the nucleosome core may fold back (quasi-)reversibly.

In this context, we reframe the scenario of Cleri and coworkers and assess a weaker, **blunt** DSB motif¹ over both a short DNA linker and the nucleosome fold, via manifold takes of all-atom, classical molecular dynamics. We thus describe the strain of the contact interface between the DNA moieties via their residual, elemental interactions.

¹In a *blunt* double strand break, no canonical forces but the aromatic π -stack of nucleic bases hold the DNA moieties.

6.2 Blunt double strand breaks: Materials and methods

The all-atom coordinates of a 30-BP double stranded DNA chain² in the B-conformation (hereafter *reference linker*) were extracted via the *Avogadro* toolkit [72]. *Blunt* double strand breaks of the DNA backbone were enforced between the fifteenth and sixteenth BP of the reference linker as either of two “flavors” (FIG. 6.1):

- **class 1:** 3'-OH and 5'-PHO termini are applied at the lesion site, i.e. the ester backbone is cleaved at the 3'-OH end of GUA15, over both DNA strands;
- **class 2:** 3'-OH and 5'-OH termini are applied at the lesion site, i.e., the ester backbone is hydrolyzed at both the 3'-OH end of GUA15 and the 5'-OH end of CYT16, and a phosphate moiety is released at both DNA strands.

Blunt DNA moieties are thus achieved (via either DSB class), held by the effective π -stack interface between the twofold (GUA15-CYT16)₂ base step.

In all scenarios, the DNA linker was solvated via a 2.5 nm-wide TIP3P water shell box and neutralized by an excess 0.15 M KCl, and thus subject to a minimization/thermalization routine that involved i) a 5000-step static minimization, ii) a 0.5 ns-NVT thermalization of the water/salt shell at fixed DNA coordinates, and iii) a 10 ns-NPT thermalization, where the DNA was subject to mild constraints.

1000 and 250 ns of unrestrained MD have been collected for the fractured and reference DNA linkers respectively, via NAMD 2.13 [75]; each DSB class has been simulated in three distinct MD instances, to an overall six microsecond trajectory. Fixed $T = 310$ K and $P = 1$ atm were enforced via the Langevin thermostat and Nosé-Hoover/Langevin barostat. Moreover, the outermost terminal nucleotides were applied soft harmonic constraints ($k = 0.03$ kcal/mol/Å², to allow $\Delta \mathbf{r} \sim 4.5$ Å fluctuations at 310 K), as done similarly by Landuzzi and coworkers [48].

²5' - 3' strand: ACGTAAAATTCGATGCCTAAGCATGCAGGA

The CHARMM36 all-atom force field [62, 76] and refinements to the Lennard-Jones and electrostatic forces (Yoo and Aksimentiev [77]) were employed. Isotropic PBC were associated with the Particle Mesh Ewald calculation of full-system electrostatics and a 12 Å cutoff was fixed to both the van der Waals and electrostatic forces. SHAKE constraints were enforced to hydrogen atoms, thus allowing a 2 fs timestep for the whole framework.

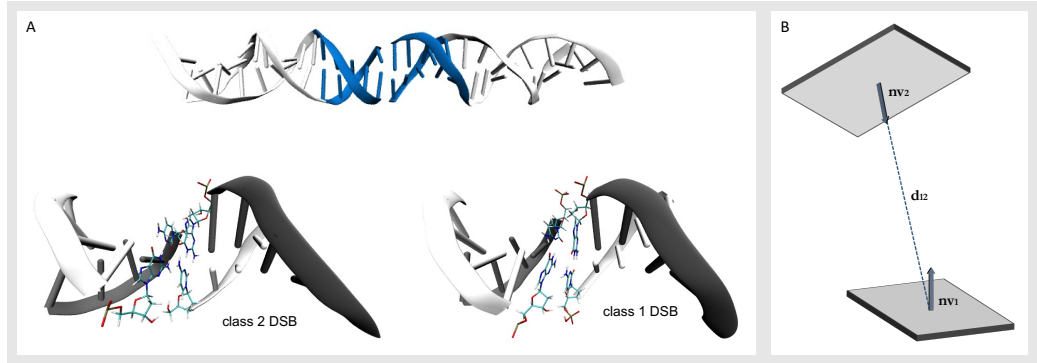


Figure 6.1: (A) Fractured DNA linker and DSB interface (in blue); the double strand break classes differ at the blunt π -stack interface by the 5' termini. (B) Non-canonical, transient π -stack contacts between DNA moieties are assessed via distance, tilt and shear criteria between adjacent nucleic bases (see section 6.2.1).

Likewise, fractures of the nucleosome fold were enforced as **class 1**, *blunt* double strand breaks, i.e. 3'-OH and 5'-PHO termini were applied at three locations of the NCP³ (FIG. 6.2). Thus, all fracture scenarios were solvated/thermalized via the routine described in section 5.1, and their subsequent dynamics evolved in 500 nanoseconds MD trajectories. The choice of the lesion sites followed from the assessment of the contact interface between DNA and the core histone fold in the reference nucleosome (section 6.4).

6.2.1 Local assessment of the residual contact interface at a DNA linker

The twofold cleft of the DNA backbone releases the nucleotides at the blunt π -stack interface from the constraints of the helical framework, and the residual contacts between DNA moieties are mediated by weak, metastable inter-

³The coordinates from the last frame of the trajectory of the reference nucleosome were employed as basic framework of all fracture scenarios.

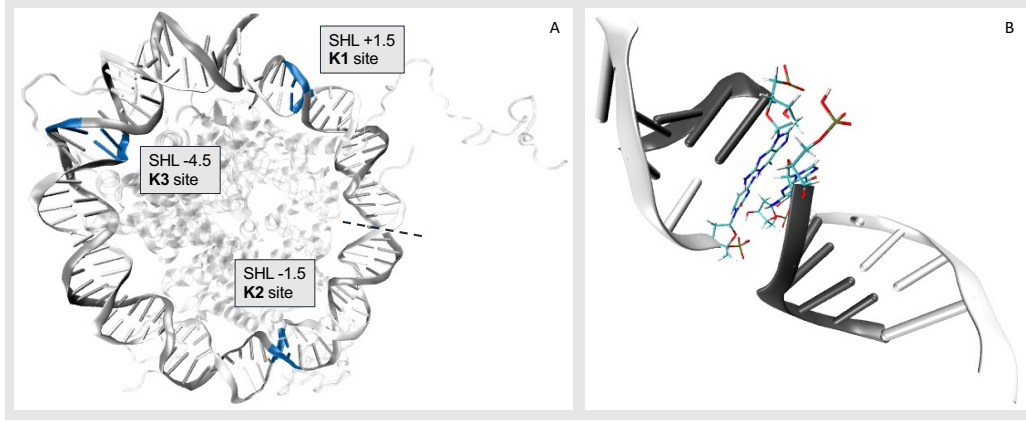


Figure 6.2: (A) Sites of *blunt* double strand breaks over the nucleosome fold (blue): DSBs were enforced at sites of either low- to no contact (**K1** site) or conserved contacts (**K3** site) between DNA and the core histone fold. The lesion at **K2** is symmetrical to K1, via the nucleosome rotational axis, or *dyad* axis (dashed black line). (B) Blunt double strand breaks of the nucleosome fold were enforced as class 1 DSBs, i.e. via 3'-OH and 5'-PHO termini.

actions. We thus assessed transient and non-canonical hydrogen bonds and π -stack contacts about the DSB sites⁴ by means of the classes of *MDAnalysis* [65, 66], whereby we classified each scenario by its overall dynamics (*vide infra*). We hereby define the *frame-wise persistence* of a (π -stack, hydrogen bond) contact as the fraction of frames which a contact is active for.

Hydrogen bonds were defined by a maximal **D** \cdots **A** distance of 3.5 Å and **DH** \cdots **A** tilt of 30°. Likewise, we described the π -stack contact of nucleic bases via three *ad hoc* criteria (FIG. 6.1):

- a threshold **distance** between the centroids of the nucleic bases of 4.83 Å;
- a threshold “**tilt**” between the aromatic surfaces of each nucleic base (or the normal vectors thereof) of 0.52 radians;
- a threshold “**slide**” of the aromatic surfaces of 0.68 radians, defined as the mean tilt value between either of the surface normal vectors (**nv**_{1,2}) and the distance vector **d**₁₂ between the centroids of the nucleic bases.

⁴As an effective DSB interface, we have taken into account a layer of ten BP about the lesion site, or nucleotides 11 to 20 (FIG. 6.1)

The maximal threshold distance, tilt and shear values were extracted from the 230 ns trajectory of the reference DNA linker⁵, as both ensemble averages and means of all values between stacked nucleotides (but the outermost termini):

$$thresh. \ distance/tilt/slide = \sum_{j=1}^N \frac{\langle \mu \rangle + 2\langle \vartheta \rangle}{N} \quad (6.1)$$

with $\langle \mu \rangle$, $\langle \vartheta \rangle$ ensemble average and standard deviation, and N the amount of base steps.

Lastly, helical degrees of freedom (*twist*, *rise*) at the fracture sites were assessed via the DO_X3DNA toolkit [81, 82].

6.3 The behavior of a blunt DSB interface at a DNA linker

The forces that hold the blunt interface of π -stacked DNA moieties are weak and estimated in about a few kcal/mol [129]. Yet, π -stack forces are reckoned to drive the assembly of liquid crystal lattices from isotropic blends of DNA oligomers and nucleotides [130–132].

While it is debated whether classical force fields misestimate the association of blunt DNA moieties [133–135], earlier molecular dynamics assessments have inferred that a major kinetic barrier to the dislocation of π -stack contacts is exerted by the exclusion of water molecules from the contact volume between DNA termini [133, 136].

Here, each DSB class has been simulated via three distinct instances of classical (1-microsecond) molecular dynamics: All runs within a DSB class share the atomic coordinates from the initial frame (that is the last frame from the DNA thermalization routine), but each DNA scenario is seeded different velocities from a Maxwell-Boltzmann distribution. Notably, this accounts for a radical variety in the subsequent dynamics of each fracture scenario, which we classify as:

⁵20 nanoseconds were excluded, based on the all-to-all RMSD of the ribose C1' atoms.

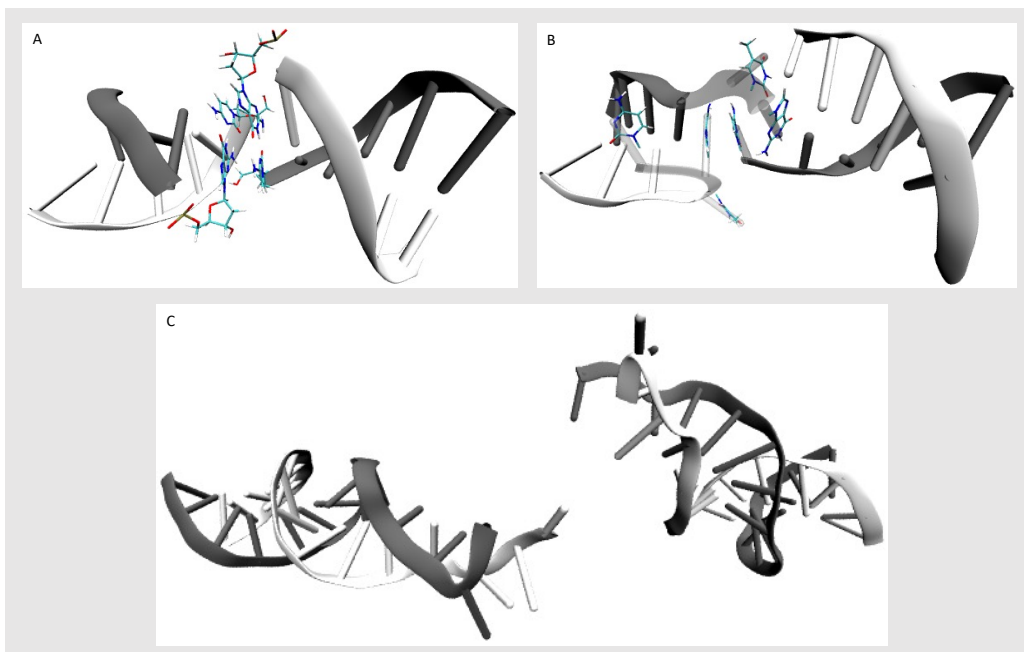


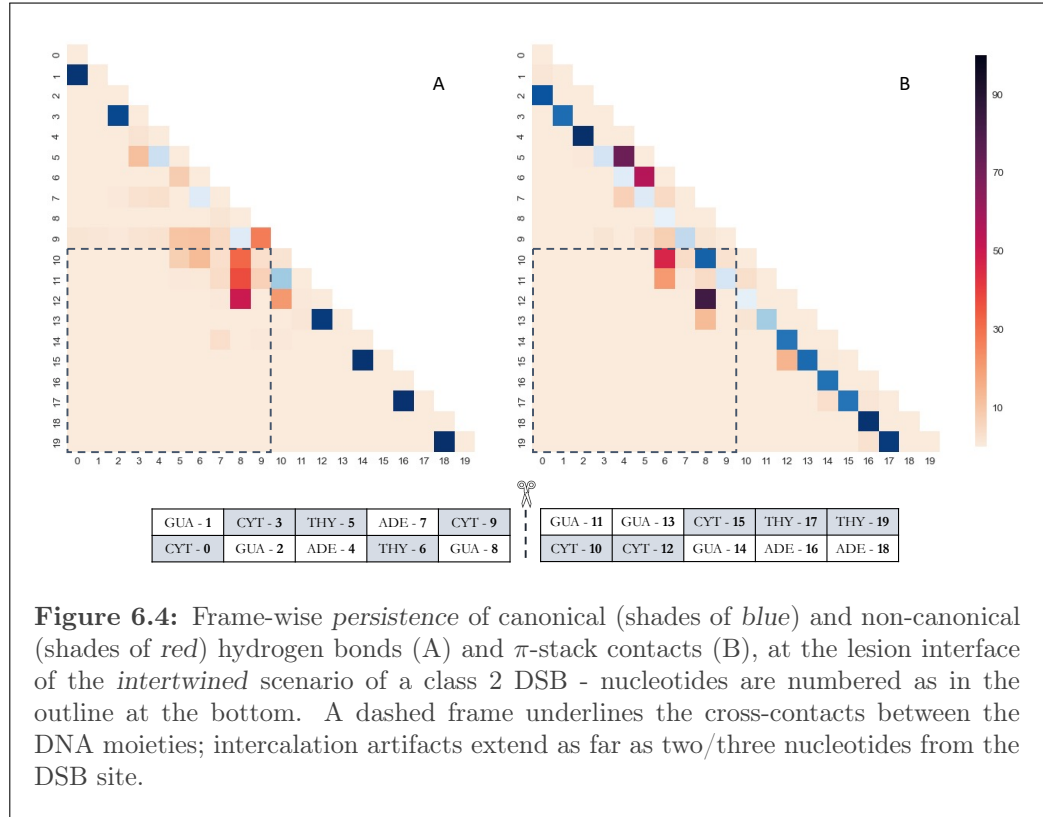
Figure 6.3: Blunt DSB scenarios are classified based on their overall dynamics as (A) resilient, (B) intertwined (artificial), or (C) cracked.

- **cracked:** no residual forces hold the lesion interface, i.e. the helical fold is dislocated and both DNA moieties drift off;
- **intertwined:** both DNA moieties fray at the contact interface and a metastable layout is achieved whereby the aromatic nucleic bases at the DNA overhangs are intercalated - both intra- and intermoiety;
- **resilient:** a mildly distorted helical layout overall holds the blunt contact interface, via canonical π -stack contact and ancillary hydrogen bonds.

Both DSB classes exhibit traits from all three dynamical scenarios, i.e. each MD instance is associated with one of the three dynamics of the lesion interface (FIG. 6.3). Moreover, the dislocation of the π -stack layout between DNA moieties, via either the intercalation of nucleic bases or the fracture of the DNA helix, is non-reversible.

6.3.1 Intercalated interfaces are artificial distortions of the DNA helix

Intertwined interfaces involve the cross intercalation of aromatic nucleic bases between DNA moieties, as far as two/three nucleotides from the π -stack contact (FIG. 6.4).



These are artificial layouts and follow from a non-reversible streak of local cracks of the DNA helix, mediated by the over-fraying of the DNA termini at the DSB interface and the subsequent loss of the Watson-Crick and π -stack order between nucleotides. Nucleic bases are freed from the helical constraints and thus intercalate in-between adjacent nucleotides of a host DNA strand, whereby the backbone of the host is mildly strained.

We overlook the further assessment of intertwined interfaces, however, as these arise from artifacts of the (CHARMM36) force field [137, 138]. In fact,

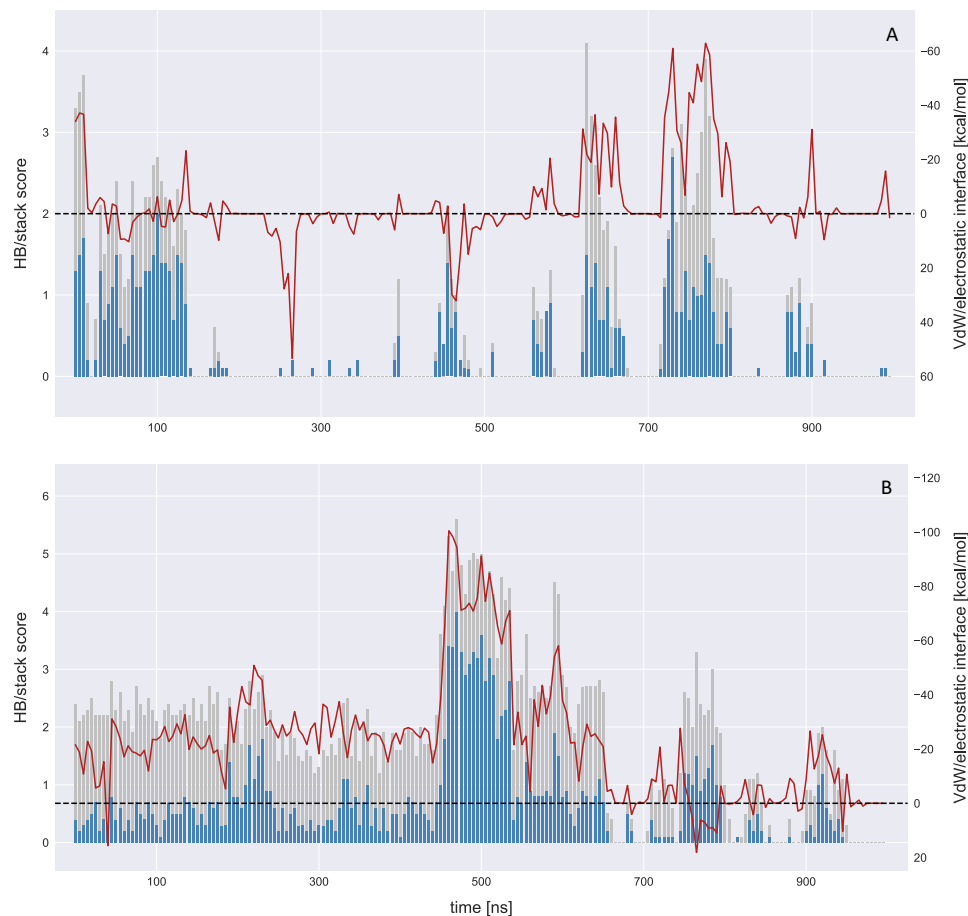


Figure 6.5: Frame-wise score of hydrogen bonds (*blue bars*) and π -stack contacts (*silver bars*), and residual electrostatic/Van der Waals forces (*red lines*) at the lesion interface of the *cracked* scenarios associated with the (A) class 1 and (B) class 2 DSB. Fractures of the DNA helix are achieved as all residual contacts between DNA moieties are cleared.

both the cross-intercalation of nucleic bases and the excess fraying of the DNA termini are reckoned to be aberrant *in silico* distortions of the DNA helix⁶.

⁶Imeddourene and coworkers have established that a mild fraying of the DNA termini shall involve the outermost nucleotides [139].

6.3.2 Blunt DSB interfaces crack at 310 K

In an earlier *in silico* assessment of double strand breaks, Landuzzi and coworkers inferred that *close* DSBs alone, i.e. where the twofold cleft of the DNA backbone lies at 1 BP, elicit the fracture of a short DNA linker, within 400 nanoseconds and below $T = 400\text{ K}$ ⁷ [48].

Here, we observe that the DNA moieties drift from the blunt DSB interface in two out of six scenarios (that is one occurrence from each DSB class), whereby all residual contacts and electrostatic/Van der Waals forces are cleared (FIG. 6.5).

The fracture of the DNA helix at class 1 DSB is abrupt (it takes about 17 ns) and driven by the over-twisted, screw-like dislocation of the contact nucleotides.

However, the failure of the contact interface at class 2 DSB is artificial to a certain extent. In fact, the frayed DNA termini are held by a π -stack contact from an isolated GUA residue, between 190.5 and 445 ns - thus, the DNA moieties intertwine artificially at the DSB interface before all residual contacts are lost (656.5 ns).

6.3.3 5' termini modulate the residual forces between blunt DNA moieties

Two *resilient* scenarios⁸ describe the residual, effective π -stack contact between blunt DNA termini. Here, the DSB lies at a CG:GC site, associated with the maximal cohesion between adjacent nucleotides [131, 140, 141].

At the lesion site, both DSB classes adjust to an over-twisted, broader π -stack interface (TABLE 6.1). However, the different flavors of the 5' DNA termini account for the nuances in the behavior of the two DSB classes.

At a class 2 DSB, both 5'-ends lack the *sticky* phosphate termini and a void volume lies between the DNA moieties. In fact, while the contact nucleotides fluctuate mildly about the metastable π -stack “equilibrium” at both *resilient*

⁷It is unclear, however, whether a *close* DSB interface fractures by thermal forces below $T = 350\text{ K}$.

⁸In fact, the canonical π -stack interface holds at class 2 DSB, whereas it twitches at class 1 DSB (about 639 ns) and adjusts to a 1:1 stack layout.

scenarios, we scored an excess 8.25° and 0.2 \AA in the mean twist and rise of the lesion site at class 2 DSB respectively (TABLE 6.1).

Likewise, the blunt DSB interface is held by a (mean) mixed 0.52 HB/1.88 π -stack contact at class 2 DSB and by a 1.15 HB/1.87 π -stack contact at class 1 DSB, and the overall electrostatic and van der Waals forces between the DNA moieties⁹ amount to $-23.08 \pm 6.22 \text{ kcal/mol}$ and $-30.88 \pm 12.69 \text{ kcal/mol}$ in the former and latter scenario respectively.

	<i>intact linker</i>	<i>class 1 DSB*</i>	<i>class 2 DSB</i>
twist [$^\circ$]	36.39 ± 5.75	47.27 ± 14.38	55.52 ± 14.08
rise [\AA]	3.30 ± 0.37	3.47 ± 0.38	3.67 ± 0.40
electrostatics/VdW [kcal/mol]	-	-30.88 ± 12.69	-23.08 ± 6.22
hydrogen bonds	-	1.15	0.52
π -stack	-	1.87	1.88

Table 6.1: Mean helical dofs, and residual electrostatic/VdW forces and contacts at the fractured and intact CG:GC sites. Data from the *resilient* scenario of class 1 DSB account for the behavior of the metastable, bound DNA before the blunt π -stack interface twitches (about 639 ns).

Furthermore, the hydrogen bond donors/acceptors involved at the DSB interface of class 1 DSB shift towards the phosphate hub. In fact:

- the (fewer) hydrogen bonds at class 2 DSB involve the anilinic amine (N2) donor at both GUA termini, and are overall distributed between the ribose moieties of the 5' CYT termini and the *sugar edges* of the 3' GUA nucleic bases. In fact, the lack of the 5'-phosphate moieties accounts for scarce backbone-backbone contacts between the blunt DNA moieties.
- the O3'-hydroxyl donors of the 3' GUA termini and the sticky phosphate moieties at the 5' CYT termini take on a major role at class 1 DSB and account for over 90% of all hydrogen bonds at the π -stack interface - whereas the aromatic nucleic bases are involved to a lesser extent.

These observations tally with earlier assessments, where it was inferred that phosphate termini (mildly) stabilize the contact interface of π -stacked DNA moieties [132, 133].

⁹Electrostatic and van der Waals forces are calculated in vacuum. Here, we account for all binary forces that involve the atoms of the 5' and 3' terminal nucleotides of the blunt DNA moieties.



6.4 The nucleosome fold is resilient towards the thermal fracture of double strand breaks

In chapter 5, we have dwelled on the subtleties of the dynamics of the nucleosome fold mediated by the erratic tumble of the histone tails. We have remarked the role of ARG/LYS electrostatic nodes and of the nucleosome interface, and corroborated earlier observations that the tails of histones cushion the local distortions of the DNA in the NCP. In fact, DNA behaves as a sticky substrate of the ARG/LYS basic moieties, whereby histone tails are allowed a moderate tumble at the sub-microsecond scale. Moreover, we have observed that the tumble of the distal termini of histone tails unfolds within several hundred nanoseconds, so that in a microsecond trajectory of the nucleosome fold all tails overall hover about their initial layout.

Notably, we located several *blind* nucleotides, i.e. which make labile or no contacts with the core histone bundle (FIG. 6.7) and which we expected to underlie effective mechanical liabilities of the NCP. Thus, we cleaved the DNA backbone at one such site, in-between of the twofold $\alpha 1 \alpha 1$ contact at SHL +1.5, via a *blunt* double strand break (section 5.1 and FIGG. 6.2 and 6.7), and evolved the subsequent trajectory of the nucleosome fold (0.5 microseconds). We will hereafter refer to this site/scenario as **K1**.

A second site, or **K2**, was chosen to mirror K1, i.e. the DNA was cleaved between GUA79...CYT109 and GUA78...CYT110, at the inward DNA minor groove at SHL -1.5 (FIG. 6.7). The sequences of nucleotides at K1 and K2 are symmetrical; at K2, however, the N-terminal tails of both histones H3 and H4 wind about the *innermost* DNA moiety (that is, the closest to the dyad site), between SHL -0.5 and -1.5 (FIG. 5.8).

Lastly, at a third site, or **K3**, the DSB lies between ADE48...THY140 and ADE49...THY139, and in-between of the twofold $\alpha 1 \alpha 1$ contact at SHL -4.5. Unlike K1, K3 is crowded with conserved electrostatic contacts to the core histone fold (FIG. 6.7), which we expected to buffer the local dynamics of the DNA.

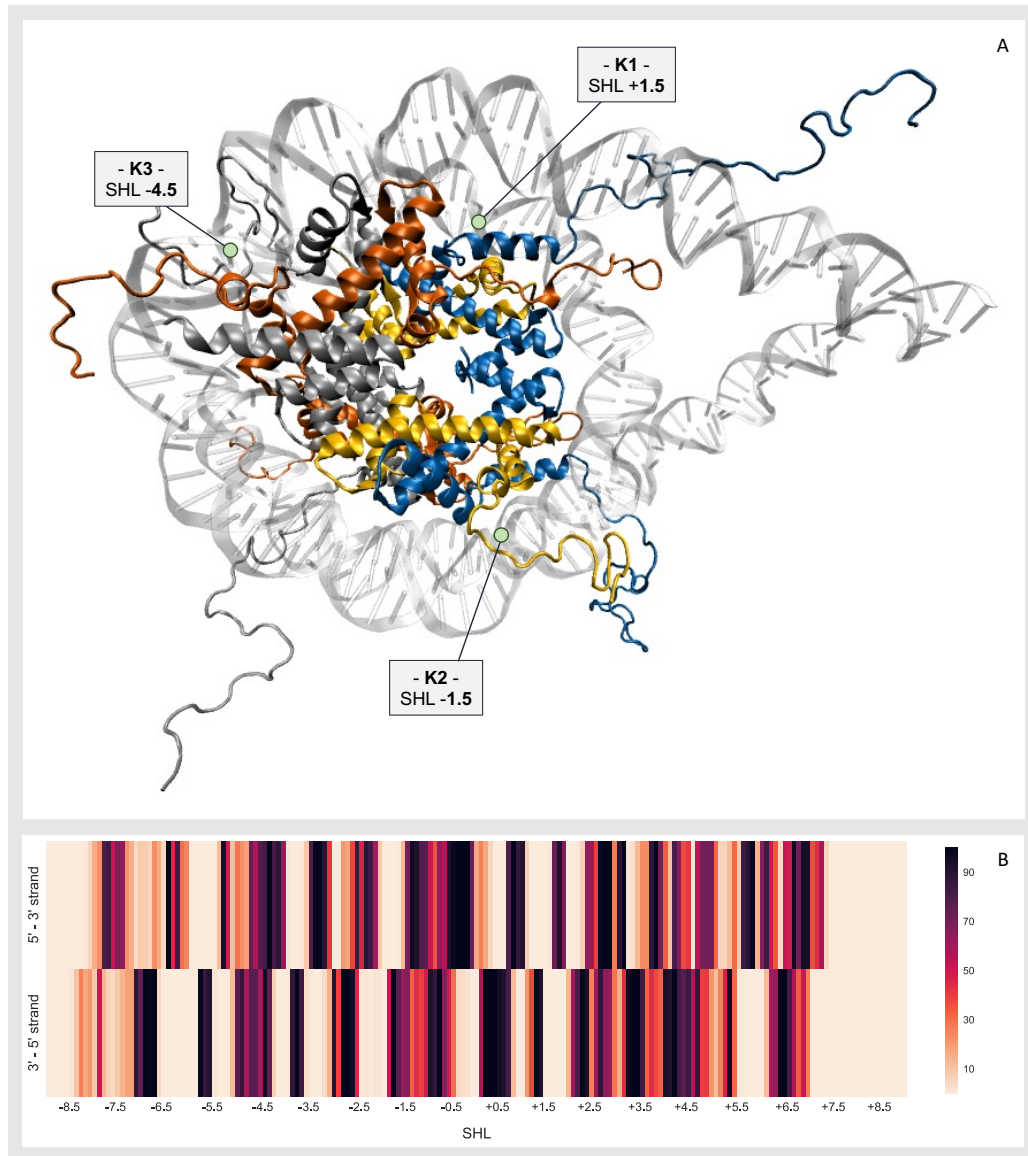


Figure 6.7: (A) Frame of the core histone octamer: histones H3/H3' (*blue*), H4/H4' (*yellow*), H2A/H2A' (*red*) and H2B/H2B' (*silver*). The local strain enforced by *blunt* double strand breaks of the DNA backbone has been assessed at three sites (K1, K2 and K3) of the nucleosome fold. These sites were chosen to account for the local robustness (B) of the nucleosome interface (section 5.2).

At no scenarios (K1 - K3) we ever observed either of the broken DNA moieties drift from the nucleosome fold. The π -stack contact between blunt DNA termini overall tolerates mild thermal fluctuations at 310 K, in spite of i) the

excess mechanical strain exerted by the wound DNA layout in the NCP, and ii) the twofold cleft of the DNA backbone.

At all three sites, however, the DSB loosens the torsional constraints of the helix fold, whereby the DNA termini resetttle about an overtwisted and broader π -stack interface (TABLE 6.2), associated with wider fluctuations of the contact nucleotides¹⁰ (FIG. 6.8).

	K1 site		K2 site		K3 site	
	<i>reference</i>	<i>fractured</i>	<i>reference</i>	<i>fractured</i>	<i>reference</i>	<i>fractured</i>
twist (°)	41.01 \pm 5.27	47.04 \pm 10.10	36.44 \pm 4.80	46.37 \pm 5.69	37.25 \pm 4.33	43.73 \pm 6.17
rise (Å)	3.57 \pm 0.40	3.65 \pm 0.39	3.70 \pm 0.38	3.71 \pm 0.35	3.55 \pm 0.32	3.49 \pm 0.35

Table 6.2: Helical *dofs* associated with the sites of *blunt* double strand breaks of the nucleosome fold.

This is in line with the inference that the forceful, coaxial detachment of π -stacked DNA moieties is aided via a twist/slide distortion of the blunt interface [135, 141]. Yet, we assessed a few subtleties underneath the dynamics of the three DSB sites/scenarios.

K1

The π -stack interface at K1 retains the local network of contacts between DNA and the helix fold of the reference nucleosome (FIG. 6.9), in spite of the twofold cleft of the DNA backbone. Moreover, broader fluctuations of the contact nucleotides are observed, whereas no excess mechanical strain is conveyed from the DSB site (FIG. 6.9).

K2

No major modifications of the local network of contacts (associated with the lesion of the DNA backbone) unfold at the nucleosome interface about **K2** either (FIG. 6.10). However, we noticed that a mild mechanical strain, exerted by the N-terminal tails of histones H3 and H4, is propagated from the lesion site, unlike K1.

In fact, both tails hover about the DNA between SHL -0.5 and SHL -1.5

¹⁰In fact, CYT110 dislocates from the DSB interface about 291.5 ns in the trajectory of K1. Likewise, both GUA79...CYT109 and GUA78...CYT110 fray reversibly, and a stacked layout is not achieved till 215 ns in the trajectory of K2.

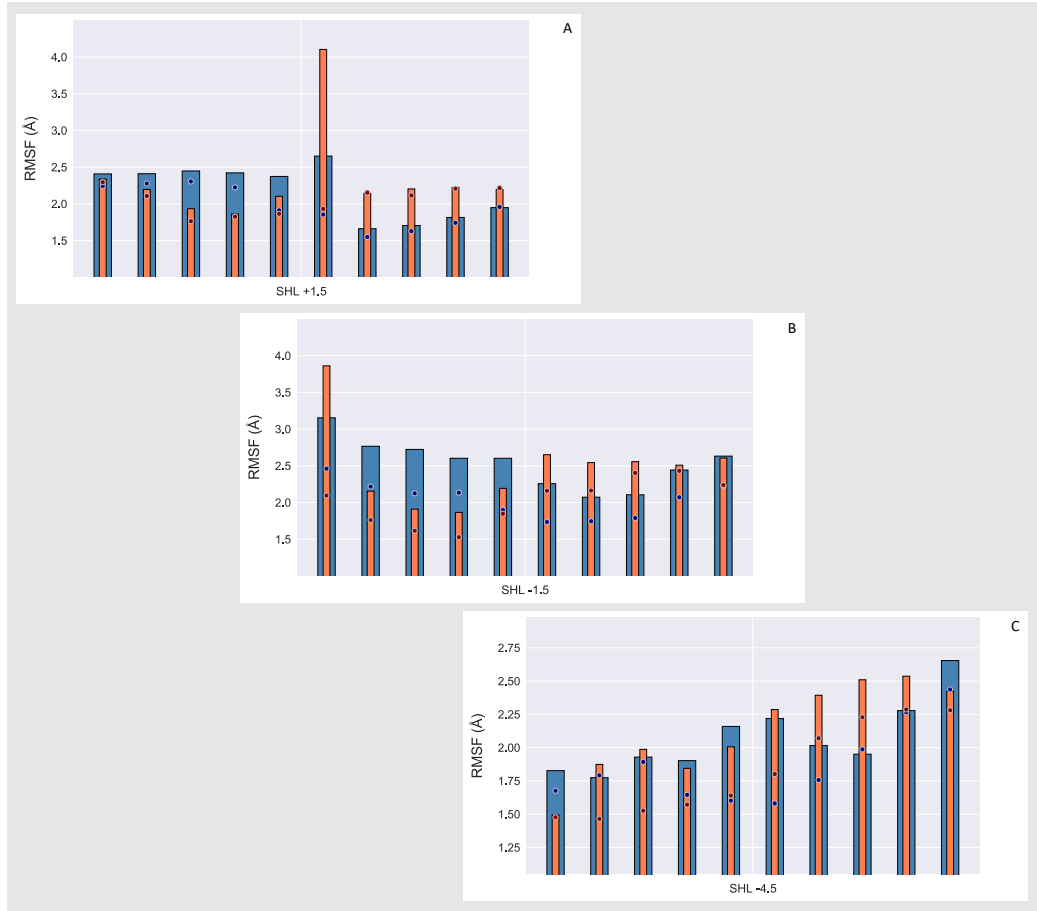


Figure 6.8: RMS fluctuations of ribose C1' atoms, at the 5' - 3' (blue) and 3' - 5' (red) DNA strands of the nucleosome fold, about the sites of the double strand breaks. Bars and circles are associated with the nucleotides of the fractured and reference nucleosome scenarios respectively. (A) The twofold cleft of the DNA backbone at **K1** modulates the excess fluctuations of the contact nucleotides (CYT110 frays from the DSB interface at 291.5 ns). (B), (C) The N-terminal tails of histones H3/H4 and H2A/H2B stir the DNA at SHL -1/-1.5 and SHL -4.5 respectively, whereby an excess micromechanical strain is propagated from the fracture sites at **K2** and at **K3**, and across the blunt π -stack interface.

(section 5.3.1), in the reference nucleosome and K2 scenarios alike, and lock the *innermost* DNA moiety at the lesion site.

About the π -stack contact interface, both the blunt DSB of the DNA backbone and the forceful motion of the N-terminal tails stir the DNA about the nucleosome fold. However, a broader strain is observed at the *outermost* DNA moiety, i.e. between SHL -1.5 and SHL -2 (FIG. 6.8), which establishes

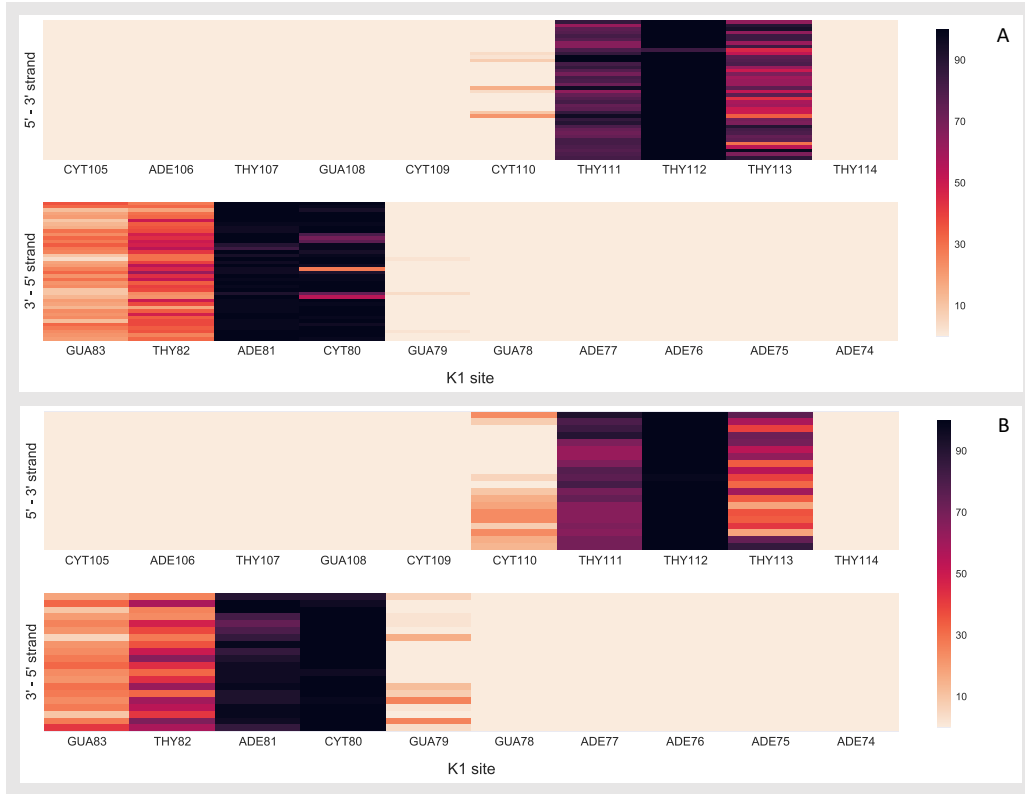


Figure 6.9: Marginal local robustness of the contact interface between DNA and the *helix fold* domain, about the **K1** site. Each layer/cell is associated with 25 nanoseconds in the trajectory of the reference (A) and fractured (B) nucleosome fold (time flows in the downward direction). The nucleosome interface overall retains the local network of contacts at K1, in spite of the twofold cleft of the DNA backbone.

no direct contact with either tail but via the blunt π -stack interface.

We reckon that the mechanical strain of DNA exerted by the H3 and H4 N-tails at K2 is conveyed across the π -stack contact and towards the outermost DNA moiety. Here, the excess distortion is not cushioned by the N-terminal tails (as with the innermost DNA moiety), which accounts for a mild RMSF imbalance between the broken DNA moieties.

K3

Lastly, the behavior of the π -stack interface at **K3** broadly resembles that observed at K2: The inward DNA minor groove at SHL -4.5 is braced by the N-terminal tails of histones H2A and H2B, whereby an excess mechanical

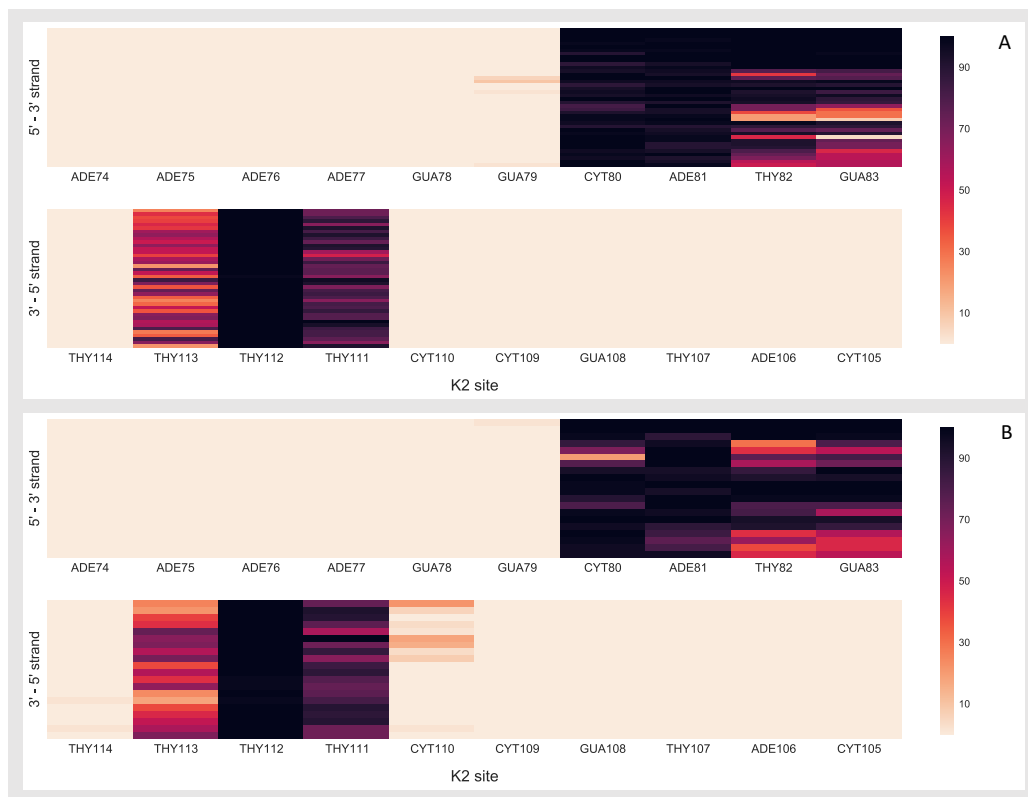


Figure 6.10: Marginal local robustness of the contact interface between DNA and the *helix fold* domain, about the **K2** site. Each layer/cell is associated with 25 nanoseconds in the trajectory of the reference (A) and fractured (B) nucleosome fold (time flows in the downward direction). As observed with K1, the nucleosome interface retains the local network of contacts at K2, in spite of the twofold cleft of the DNA backbone and the excess mechanical strain exerted via the N-terminal tails of histones H3 and H4.

strain resonates from the lesion site (FIG. 6.8).

Here, however, the cleft of the DNA backbone lies over a stretch of nucleotides directly involved with the H2A-R29 electrostatic hub¹¹. Thus, unlike K1 and K2, the local network of contacts at the interface between DNA and the core helix fold is affected majorly at K3 (FIG. 6.11).

In the trajectory of the reference nucleosome, a conserved interface is established about the canonical contact site at SHL -4.5, where ADE48 and ADE49 are locked by an *arc-like* layout of the basic side chains of H2A-R29,

¹¹Unlike K1 and K2, where the blunt double strand break lies in-between of the manifold electrostatic interface of H3/H3'-R63 with the walls of the minor groove at SHL ± 1.5 (FIG. 6.2).

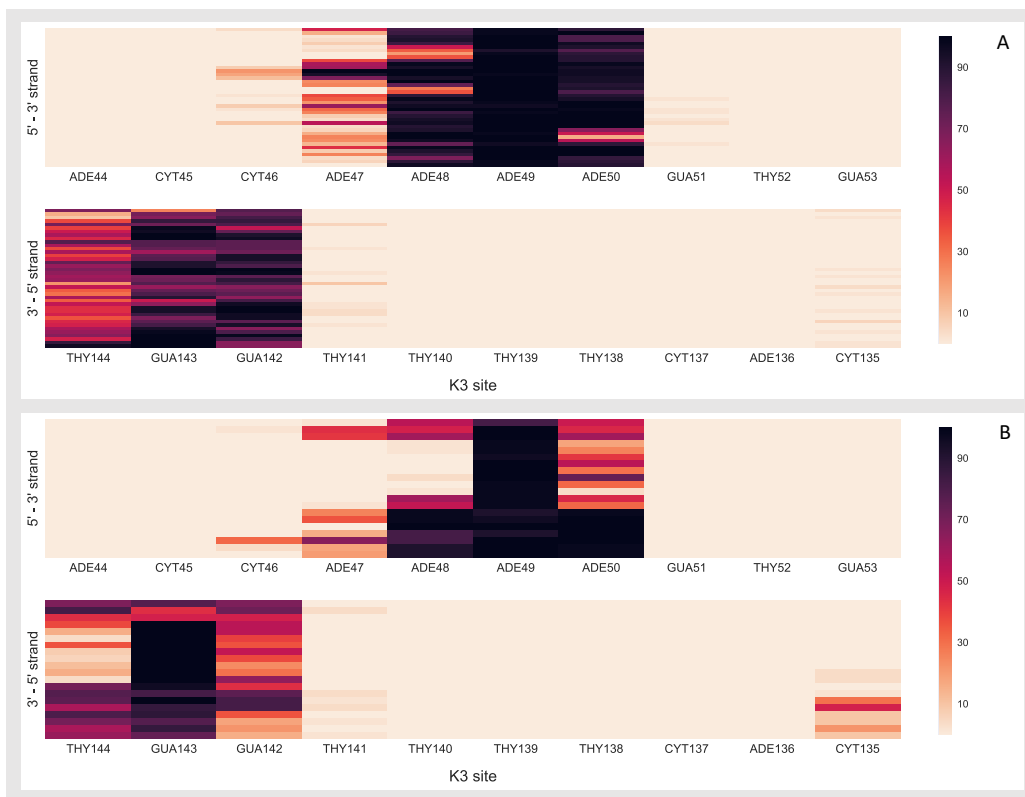
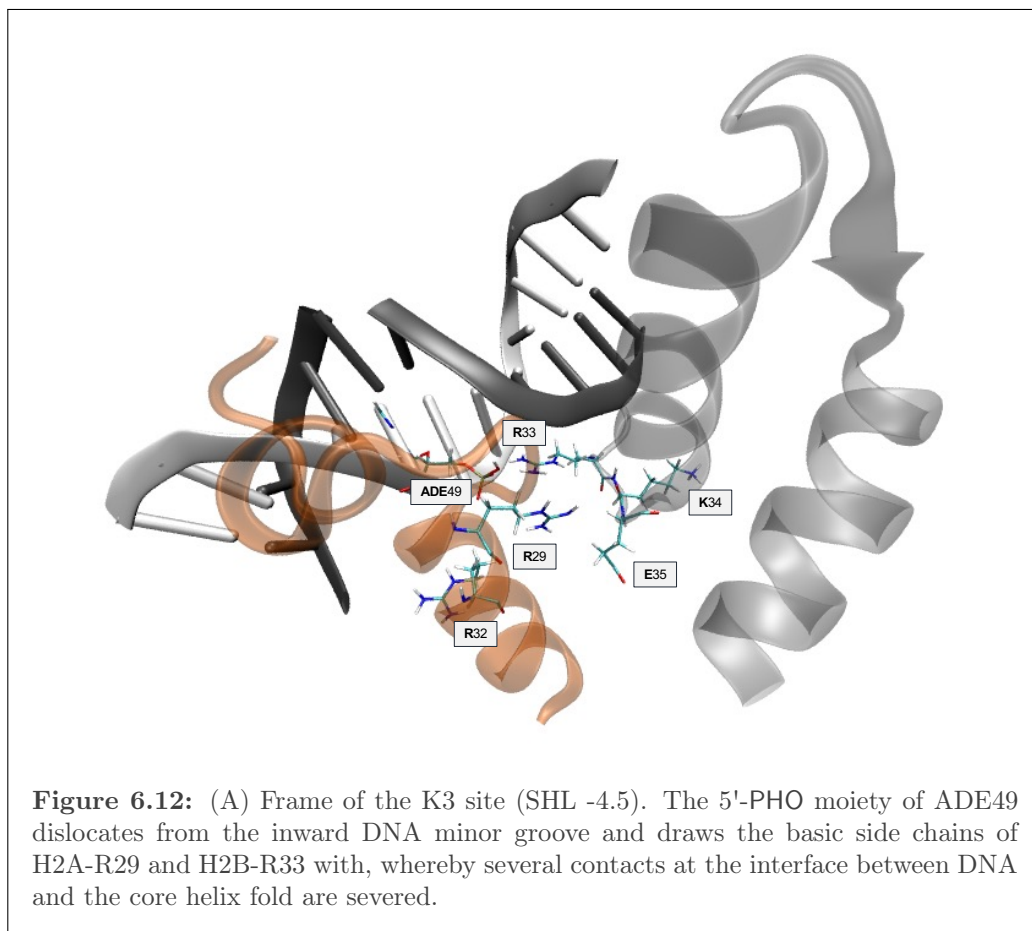


Figure 6.11: Marginal local robustness of the contact interface between DNA and the *helix fold* domain, about the **K3** site. Each layer/cell is associated with 25 nanoseconds in the trajectory of the reference (A) and fractured (B) nucleosome fold (time flows in the downward direction). The “arc-like” contact layout observed at SHL -4.5 in the reference nucleosome (section 5.3.3) is disrupted upon the dislocation of the 5'-PHO electrostatic hub of ADE49 from the lesion site, whereby major conserved contacts are severed. However, several metastable electrostatic contacts are restored as H2B-R33 relocates into the inward DNA minor groove (325 ns).

H2A-R32 and H2B-R33 (FIG. 5.7). This layout is disrupted upon the cleft of the DNA backbone and a streak of local dislocations unfolds at K3:

- at the core of the arc-like layout, the 5'-PHO moiety of ADE49 acts as an electrostatic hub of the three ARG residues. At K3, it is projected outwards of the π -stack interface and towards the core helix fold (FIG. 6.12);
- via the dislocation of ADE49, H2A-R29 is thus released from the minor groove, and establishes a twofold contact between the acidic H2B-E35 and ADE49 (FIG. 6.12), thereby severing two conserved electrostatic contacts with the walls of the DNA minor groove (H2A-R29/ADE50 and H2A-



R29/GUA142);

- likewise, H2B-R33 is drawn from the minor groove at SHL -4.5 by the 5'PHO moiety of ADE49. As it relocates into the groove (about 325 ns in the trajectory of K3), however, H2B-R33 drives ADE49 with, thereby distorting (slightly) the π -stack interface between ADE48...THY140 and ADE49...THY139;
- lastly, H2A-R32 is steered backwards and towards the inward DNA major groove at SHL -4, as it loses contact with ADE49.

Such a major downsize of the contact interface between DNA and the helix fold achieves but a mild distortion of the blunt π -stack interface at K3, as

observed with K1 and K2 (TABLE 6.2), and is counteracted effectively by the N-terminal tails of histones H2A and H2B.

6.5 Conclusive remarks

The dynamics of a *blunt* double strand break of the DNA backbone has been assessed over both a short DNA chain (to depict a DNA linker from a chromatin-like scenario) and the nucleosome fold, via classic, all-atom MD. In a DNA linker scenario, the strain of the disjointed DNA moieties is driven by the loss/reshuffle of the canonical layout of π -stack contacts and hydrogen bonds at the lesion interface, whereby we classify the outcome of a single instance of MD trajectory as i) **resilient** (i.e., the blunt π -stack interface is mildly distorted), ii) **cracked** (the fracture of the DNA helix), or iii) **intertwined** (an artificial layout of intercalated nucleic bases).

Such manifold, rich dynamics is absent in the nucleosome fold, as blunt DSBs cannot elicit the fracture of DNA from the nucleosome interface - in fact, the disjointed DNA moieties resettle locally and achieve a mildly distorted π -stack layout. We thus collectively infer that:

- the dynamics of DNA termini from the additive CHARMM36 force field of nucleic acids is associated with *excess fraying* (i.e. beyond the outermost nucleotides) and cross-intercalation of the nucleic bases, which are both artificial distortions from the *in silico* depiction of DNA;
- *blunt* DSB motifs elicit the thermal fracture of DNA linkers at 310 K within a few (hundred) nanoseconds, via an abrupt “screw-like” dislocation or a step-wise decay of the blunt π -stack interface - the latter, however, may be mediated by the artificial fraying of the DNA termini;
- both of two DSB classes, which differ by the 5' termini (5'-OH v. 5'-PHO), adjust to a broad and over-twisted π -stack layout. However, the lack of the *sticky* phosphate moieties accounts for a further increase in *twist* and *rise* and fewer residual hydrogen bonds at the lesion site. Moreover, the 5'-PHO termini (and lack thereof) reframe the overall layout of hydrogen bond donors/acceptors at the π -stack interface;

- the nucleosome interface exerts a kinetic barrier towards the dislocation of (*blunt*) double strand breaks from the NCP, in line with earlier observations *in silico* [59], and contributes to the residual interface (hydrogen bonds, π -stack contacts) between the DNA moieties. Nevertheless, the lesion sites are mildly susceptible to the excess strain exerted by the histone tails onto the DNA, and conveyed across the contact interface.

Discussion and conclusive remarks

As a widely acknowledged paradigm [47], the earliest downstream cascade of events from the irradiation of cells (which is associated with a local loss of equilibrium) is described by three stages:

1. a *physical* stage of the effective collisions of the radiation carriers with the medium, whereby DNA and water molecules are locally excited or oxidized;
2. a *physicochemical* (or *radiochemical*) stage, which takes in the direct lesions of the DNA and the radiolytical decay of water molecules (10^{-15} - 10^{-12} seconds);
3. a *chemical* stage of the diffusion and reaction of radiolytical radicals and solvated electrons, which accounts for indirect DNA lesions (10^{-12} - 10^{-6} seconds).

These timescales are associated with both i) the mechanistic depiction of the radiation field *in silico* via state-of-the-art *Monte-Carlo track structure* (MCTS) toolkits, and ii) the all-atom *molecular dynamics* (MD) assessment of effective chromatin frameworks, such as (arrays of) nucleosomes [69].

In fact, a few efforts that unified the depiction of the stochastic decay of the radiation track and the dynamical evolution of the irradiated medium have been attempted. Abolfath and coworkers have employed the Geant4-DNA extension to distribute reactive oxygen species (ROS) about a DNA molecule, and evolved their subsequent diffusion/reaction via molecular dynamics [142, 143]. In the context of a multiscale framework, Surdutovich, Solov'yov and

coworkers have described *in silico* the propagation of *thermal shockwaves*¹² about the radiation track of heavy ions at elevated LET, from the dynamical evolution of the radial dose distribution (defined by an amorphous track structure formalism) [144].

Likewise, in this work, we have attempted to interlace the twofold framework of MCTS and MD, i.e. to forward the outcome of a MCTS assessment of a nucleosome fold (as an effective distribution of double strand breaks, enforced by a beam of hydrogen ions) and achieve a MD depiction of its subsequent behavior. However, we failed to isolate significant layouts of DNA lesions, nor we located *sensitive volumes* of a nucleosome fold, i.e. associated with a significant likelihood of lethal DNA lesions. Moreover, a “bijjective” association between MCTS and MD assessments is statistically flawed and/or unfeasible.

Thus, we redirected our efforts and assessed the elemental DNA lesions, i.e. the stochastic occurrence and strain of the DNA helix exerted by **blunt** double strand breaks. In this concern, we observed that such DSB *motifs* recur seldom by the direct effect of a beam of hydrogen ions, at clinically-relevant LET. Nevertheless, we wished to advance on the earlier work of Cleri and coworkers [48, 59], on the thermal and mechanical resilience of DSBs in the framework of classic and steered MD, who inferred that:

- short-distanced, *staggered* DSB alone elicit the fracture of the DNA helix;
- the electrostatic interface of the nucleosome fold holds the disjointed DNA moieties of a DSB lesion, which otherwise drift in a short DNA linker scenario.

That the twofold cleft of the DNA backbone elicits its fracture underlies the enzymatic activity of DDR sensors, such as the Ku heterodimer, which bind and involve the DNA termini with reversible supramolecular assemblies [145]. We corroborate the idea that nucleosomes exert a kinetic barrier towards the thermal dissociation of double strand breaks, thus delaying their mechanical

¹²Radiation-enforced shockwaves are elicited by local heat spikes, from the massive (vibrational) excitation of water molecules, which are not dissipated by molecular diffusion nor heat conduction.

decay, and overly stabilize short-distanced DSB motifs. Likewise, we reckon that the consensus threshold distance to define a DSB, i.e. within 10-BP of the twofold cleft of the DNA backbone, is too conservative a criterion for a lesion of the nucleosome fold¹³.

We remark, however, that (kinetic, thermodynamic) inferences about the resilience of DNA lesions shall not overlook the inherent weaknesses of its depiction *in silico*:

- it is debated whether actual all-atom force fields misestimate the π -stack forces between nucleic bases in DNA [119, 138, 146] and the cohesion of DNA termini;
- the vector of the DSB affects the layout - and subsequent behavior - of the DNA termini at the lesion interface, whether the cleft is enforced by radiation or nucleolytic enzymes¹⁴ [128].
- radiation-enforced shockwaves account for an excess loss of equilibrium and thermo-mechanical strain of the chromatin fold, a few nanometers about the radiation track of heavy ions; yet, their role has not been demonstrated in liquid water [144].

In conclusion, we reckon that further *in silico* contributions, within the context of multi-scale MD, will flank the microdosimetric assessment of the earliest events from the irradiation of cells, in view of the chronic lack of data *in vitro* at the micro- and nanoscale.

¹³That is, if atomistic MD assessments and molecular mechanics force fields are reliable.

¹⁴In which case the class of the excision enzyme exerts a further control of the DNA substrate [147].

Bibliography

- [1] S. Baldi, P. Korber, and P. B. Becker. “Beads on a string-nucleosome array arrangements and folding of the chromatin fiber”. *Nature Structural and Molecular Biology* 27 (2020), 109–118. DOI: <https://doi.org/10.1038/s41594-019-0368-x>.
- [2] C. L. Woodcock and S. Dimitrov. “Higher-order structure of chromatin and chromosomes”. *Current Opinion in Genetics and Development* 11 (2001), 130–135. DOI: [https://doi.org/10.1016/S0959-437X\(00\)00169-6](https://doi.org/10.1016/S0959-437X(00)00169-6).
- [3] A. R. Cutter and J. J. Hayes. “A brief review of nucleosome structure”. *FEBS Letters* 589 (2015), 2914–2922. DOI: <https://doi.org/10.1016/j.febslet.2015.05.016>.
- [4] G. E. Zentner and S. Henikoff. “Regulation of nucleosome dynamics by histone modifications”. *Nature Structural and Molecular Biology* 20 (2013), 259–266. DOI: <https://doi.org/10.1038/nsmb.2470>.
- [5] A. K. Sharma and M. J. Hendzel. “The relationship between histone posttranslational modification and DNA damage signaling and repair”. *International Journal of Radiation Biology* 95 (2019), 382–393. DOI: <https://doi.org/10.1080/09553002.2018.1516911>.
- [6] M. Ghoneim, H. A. Fuchs, and C. A. Musselman. “Histone tail conformations: A fuzzy affair with DNA”. *Trends in Biochemical Sciences* 46 (2021), 564–578. DOI: <https://doi.org/10.1016/j.tibs.2020.12.012>.

- [7] A. Ciccia and S. J. Elledge. “The DNA damage response: making it safe to play with knives”. *Molecular Cell* 40 (**2010**), 179–204. DOI: <https://doi.org/10.1016/j.molcel.2010.09.019>.
- [8] R. Scully, A. Panday, R. Elango, and N. A. Willis. “DNA double-strand break repair-pathway choice in somatic mammalian cells”. *Nature Reviews - Molecular Cell Biology* 20 (**2019**), 698–714. DOI: <https://doi.org/10.1038/s41580-019-0152-0>.
- [9] J. F. Ward. “DNA damage produced by ionizing radiation in mammalian cells: identities, mechanisms of formation, and reparability”. *Progress in Nucleic Acid Research and Molecular Biology* 35 (**1988**), 95–125. DOI: [https://doi.org/10.1016/S0079-6603\(08\)60611-X](https://doi.org/10.1016/S0079-6603(08)60611-X).
- [10] H. H. Y. Chang, N. R. Pannunzio, N. Adachi, and M. R. Lieber. “Non-homologous DNA end joining and alternative pathways to double-strand break repair”. *Nature Reviews - Molecular Cell Biology* 18 (**2017**), 495–506. DOI: <https://doi.org/10.1038/nrm.2017.48>.
- [11] S. J. McMahon. “The linear quadratic model: usage, interpretation and challenges”. *Physics in Medicine and Biology* 64 (**2018**), 01TR01. DOI: <https://doi.org/10.1088/1361-6560/aaf26a>.
- [12] L. Huang, A. R. Snyder, and W. F. Morgan. “Radiation-induced genomic instability and its implications for radiation carcinogenesis”. *Oncogene* 22 (**2003**), 5848–5854. DOI: <https://doi.org/10.1038/sj.onc.1206697>.
- [13] E. I. Azzam, S. M. de Toledo, and J. B. Little. “Oxidative metabolism, gap junctions and the ionizing radiation-induced bystander effect”. *Oncogene* 22 (**2003**), 7050–7057. DOI: <https://doi.org/10.1038/sj.onc.1206961>.
- [14] E. Rørvik et al. “Exploration and application of phenomenological RBE models for proton therapy”. *Physics in Medicine and Biology* 63 (**2018**), 185013. DOI: <https://doi.org/10.1088/1361-6560/aad9db>.

- [15] H. Paganetti et al. “Relative biological effectiveness (RBE) values for proton beam therapy”. *International Journal of Radiation Oncology Biology Physics* 53 (**2002**), 407–421. DOI: [https://doi.org/10.1016/S0360-3016\(02\)02754-2](https://doi.org/10.1016/S0360-3016(02)02754-2).
- [16] H. Paganetti. “Relative biological effectiveness (RBE) values for proton beam therapy. Variations as a function of biological endpoint, dose, and linear energy transfer”. *Physics in Medicine and Biology* 59 (**2014**), R419–472. DOI: <https://doi.org/10.1088/0031-9155/59/22/R419>.
- [17] F. Tommasino and M. Durante. “Proton radiobiology”. *Cancers* 7 (**2015**), 353–381. DOI: <https://doi.org/10.3390/cancers7010353>.
- [18] M. Krämer, W. Kraft-Weyrather, and M. Scholz. “The increased biological effectiveness of heavy charged particles: from radiobiology to treatment planning”. *Technology in Cancer Research and Treatment* 2 (**2003**), 427–436. DOI: <https://doi.org/10.1177/153303460300200507>.
- [19] T. Inaniwa et al. “Treatment planning for a scanned carbon beam with a modified microdosimetric kinetic model”. *Physics in Medicine and Biology* 55 (**2010**), 6721–6737. DOI: <https://doi.org/10.1088/0031-9155/55/22/008>.
- [20] T. Inaniwa and N. Kanematsu. “Adaptation of stochastic microdosimetric kinetic model for charged-particle therapy treatment planning”. *Physics in Medicine and Biology* 63 (**2018**), 095011. DOI: <https://doi.org/10.1088/1361-6560/aabede>.
- [21] R. B. Hawkins. “A microdosimetric-kinetic model for the effect of non-Poisson distribution of lethal lesions on the variation of RBE with LET”. *Radiation Research* 160 (**2003**), 61–69. DOI: <https://doi.org/10.1667/RR3010>.
- [22] V. E. Bellinzona et al. “Linking microdosimetric measurements to biological effectiveness in ion beam therapy: A review of theoretical aspects of MKM and other models”. *Frontiers in Physics* 8 (**2021**), 578492. DOI: <https://doi.org/10.3389/fphy.2020.578492>.

- [23] A. M. Kellerer and H. H. Rossi. “A generalized formulation of dual radiation action”. *Radiation Research* 75 (1978), 471–488. DOI: <https://doi.org/10.2307/3574835>.
- [24] Y. Kase et al. “Microdosimetric measurements and estimation of human cell survival for heavy-ion beams”. *Radiation Research* 166 (2006), 629–638. DOI: <https://doi.org/10.1667/RR0536.1>.
- [25] T. Sato and Y. Furusawa. “Cell survival fraction estimation based on the probability densities of domain and cell nucleus specific energies using improved microdosimetric kinetic models”. *Radiation Research* 178 (2012), 341–356. DOI: <https://doi.org/10.1667/RR2842.1>.
- [26] M. Scholz, A. M. Kellerer, W. K.-Weyrather, and G. Kraft. “Computation of cell survival in heavy ion beams for therapy. The model and its approximation”. *Radiation and Environmental Biophysics* 36 (1997), 59–66. DOI: <https://doi.org/10.1007/s004110050055>.
- [27] B. Fertil, I. Reydellet, and P. J. Deschavanne. “A benchmark of cell survival models using survival curves for human cells after completion of repair of potentially lethal damage”. *Radiation Research* 138 (1994), 61–69. DOI: <https://doi.org/10.2307/3578847>.
- [28] T. Elsässer and M. Scholz. “Cluster effects within the local effect model”. *Radiation Research* 167 (2007), 319–329. DOI: <https://doi.org/10.1667/RR0467.1>.
- [29] T. Elsässer, M. Krämer, and M. Scholz. “Accuracy of the local effect model for the prediction of biologic effects of carbon ion beams in vitro and in vivo”. *International Journal of Radiation Oncology Biology Physics* 71 (2008), 866–872. DOI: <https://doi.org/10.1016/j.ijrobp.2008.02.037>.
- [30] T. Elsässer et al. “Quantification of the relative biological effectiveness for ion beam radiotherapy: Direct experimental comparison of proton and carbon ion beams and a novel approach for treatment planning”. *International Journal of Radiation Oncology Biology Physics* 78 (2010), 1177–1183. DOI: <https://doi.org/10.1016/j.ijrobp.2010.05.014>.

- [31] O. Mohamad et al. “Carbon ion radiotherapy: A review of clinical experiences and preclinical research, with an emphasis on DNA damage/repair”. *Cancers* 9 (2017), 66. DOI: <https://doi.org/10.3390/cancers9060066>.
- [32] B. Brzozowska, A. Tartas, and A. Wojcik. “Monte Carlo modeling of DNA lesions and chromosomal aberrations induced by mixed beams of alpha particles and x-rays”. *Frontiers in Physics* 8 (2020), 567864. DOI: <https://doi.org/10.3389/fphy.2020.567864>.
- [33] S. Agostinelli et al. “Geant4 - A simulation toolkit”. *Nuclear Instruments and Methods in Physics - Research Section A* 506 (2003), 250–303. DOI: [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).
- [34] J. Allison et al. “Geant4 developments and applications”. *IEEE Transactions on Nuclear Science* 53 (2006), 270–278. DOI: <https://doi.org/10.1109/TNS.2006.869826>.
- [35] J. Allison et al. “Recent developments in Geant4”. *Nuclear Instruments and Methods in Physics - Research Section A* 835 (2016), 186–225. DOI: <https://doi.org/10.1016/j.nima.2016.06.125>.
- [36] G. Battistoni et al. “The FLUKA code: An accurate simulation tool for particle therapy”. *Frontiers in Oncology* 6 (2016), 116. DOI: <https://doi.org/10.3389/fonc.2016.00116>.
- [37] S. J. McMahon and K. M. Prise. “Mechanistic modelling of radiation responses”. *Cancers* 11 (2019), 205. DOI: <https://doi.org/10.3390/cancers11020205>.
- [38] S. Incerti et al. “Simulating radial dose of ion tracks in liquid water simulated with Geant4-DNA: A comparative study”. *Nuclear Instruments and Methods in Physics Research Section B* 333 (2014), 92–98. DOI: <https://doi.org/10.1016/j.nimb.2014.04.025>.
- [39] S. Incerti et al. “Comparison of GEANT4 very low energy cross section models with experimental data in water”. *Medical Physics* 37 (2010), 4692–4708. DOI: <https://doi.org/10.1118/1.3476457>.

- [40] S. Incerti et al. “The Geant4-DNA project”. *International Journal of Modeling, Simulation, and Scientific Computing* 01 (**2010**), 157–178. DOI: <https://doi.org/10.1142/S1793962310000122>.
- [41] M. A. Bernal et al. “Track structure modeling in liquid water: A review of the Geant4-DNA very low energy extension of the Geant4 Monte Carlo simulation toolkit”. *Physica Medica* 31 (**2015**), 861–874. DOI: <https://doi.org/10.1016/j.ejmp.2015.10.087>.
- [42] S. Incerti et al. “Geant4-DNA example applications for track structure simulations in liquid water: A report from the Geant4-DNA Project”. *Medical Physics* 45 (**2018**), e722–e739. DOI: <https://doi.org/10.1002/mp.13048>.
- [43] S. Incerti, M. Douglass, S. Penfold, S. Guatelli, and E. Bezak. “Review of Geant4-DNA applications for micro and nanoscale simulations”. *Physica Medica* 32 (**2016**), 1187–1200. DOI: <https://doi.org/10.1016/j.ejmp.2016.09.007>.
- [44] H. Nikjoo, P. O’Neill, D. T. Goodhead, and M. Terrissol. “Computational modelling of low-energy electron-induced DNA damage by early physical and chemical events”. *International Journal of Radiation Biology* 71 (**1997**), 467–483. DOI: <https://doi.org/10.1080/095530097143798>.
- [45] W. Friedland, M. Dingfelder, P. Kunderát, and P. Jacob. “Track structures, DNA targets and radiation effects in the biophysical Monte Carlo simulation code PARTRAC”. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 711 (**2011**), 28–40. DOI: <https://doi.org/10.1016/j.mrfmmm.2011.01.003>.
- [46] N. Lampe et al. “Mechanistic DNA damage simulations in Geant4-DNA part 1: A parameter study in a simplified geometry”. *Physica Medica* 48 (**2018**), 135–145. DOI: <https://doi.org/10.1016/j.ejmp.2018.02.011>.
- [47] T. Friedrich. “Biophysical modeling of effects of ionizing radiation and associated uncertainties”. **2016**. URL: <http://tuprints.ulb.tu-darmstadt.de/6189/>.

- [48] F. Landuzzi, P. L. Palla, and F. Cleri. “Stability of radiation-damaged DNA after multiple strand breaks”. *Physical Chemistry Chemical Physics* 19 (2017), 14641–14651. DOI: <https://doi.org/10.1039/c7cp02266b>.
- [49] D. I. D’souza and L. Harrison. “Repair of clustered uracil DNA damages in *Escherichia coli*”. *Nucleic Acids Research* 31 (2003), 4573–4581. DOI: <https://doi.org/10.1093/nar/gkg493>.
- [50] R. Hanai, M. Yazu, and K. Hieda. “On the experimental distinction between ssbs and dsbs in circular DNA”. *International Journal of Radiation Biology* 73 (1998), 475–479. DOI: <https://doi.org/10.1080/095530098142013>.
- [51] G. L. Dianov et al. “Repair of uracil residues closely spaced on the opposite strands of plasmid DNA results in double-strand break and deletion formation”. *Molecular and General Genetics* 225 (1991), 448–452. DOI: <https://doi.org/10.1007/BF00261686>.
- [52] L. Petrolli, F. Tommasino, E. Scifoni, and G. Lattanzi. “Can we assess early DNA damage at the molecular scale by radiation track structure simulations? A tetranucleosome scenario in Geant4-DNA”. *Frontiers in Physics* 8 (2020), 576284. DOI: <https://doi.org/10.3389/fphy.2020.576284>.
- [53] E. Delage et al. “PDB4DNA: Implementation of DNA geometry from the Protein Data Bank (PDB) description for Geant4-DNA Monte-Carlo simulations”. *Computer Physics Communications* 192 (2015), 282–288. DOI: <https://doi.org/10.1016/j.cpc.2015.02.026>.
- [54] T. Schalch, S. Duda, D. F. Sargent, and T. J. Richmond. “Structure of the 4_601_167 tetranucleosome”. 2005. DOI: <https://doi.org/10.2210/pdb1zbb/pdb>.
- [55] H. M. Berman et al. “The Protein Data Bank”. *Nucleic Acids Research* 28 (2000), 235–242. DOI: <https://doi.org/10.1093/nar/28.1.235>.

- [56] T. Schalch, S. Duda, D. F. Sargent, and T. J. Richmond. “X-ray structure of a tetranucleosome and its implications for the chromatin fibre”. *Nature* 436 (**2005**), 138–141. DOI: <https://doi.org/10.1038/nature03686>.
- [57] R. Brun and F. Rademakers. “ROOT - an object oriented data analysis framework”. *Nuclear Instruments and Methods in Physics - Research Section A* 389 (**1997**), 81–86. DOI: 10.1016/S0168-9002(97)00048-X.
- [58] J. Schuemann et al. “A new standard DNA damage (SDD) data format”. *Radiation Research* 191 (**2019**), 76–92. DOI: <https://doi.org/10.1667/RR15209.1>.
- [59] F. Cleri, F. Landuzzi, and R. Blossey. “Mechanical evolution of DNA double-strand breaks in the nucleosome”. *PLoS Computational Biology* 14 (**2018**), e1006224. DOI: <https://doi.org/10.1371/journal.pcbi.1006224>.
- [60] M. Souici et al. “Single- and Double-Strand Breaks of Dry DNA Exposed to Protons at Bragg-Peak Energies”. *The Journal of Physical Chemistry B* 121 (**2017**), 497–507. DOI: <https://doi.org/10.1021/acs.jpcb.6b11060>.
- [61] K. Vanommeslaeghe and A. D. Mackerell. “CHARMM additive and polarizable force fields for biophysics and computer-aided drug design”. *Biochimica et Biophysica Acta - General Subjects* 1850 (**2015**), 861–871. DOI: <https://doi.org/10.1016/j.bbagen.2014.08.004>.
- [62] J. Huang et al. “CHARMM36m: An improved force field for folded and intrinsically disordered proteins”. *Nature Methods* 14 (**2017**), 71–73. DOI: <https://doi.org/10.1038/nmeth.4067>.
- [63] A. D. Mackerell, M. Feig, and C. L. Brooks III. “Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations”. 25 (**2004**), 1400–1415. DOI: <https://doi.org/10.1002/jcc.20065>.

- [64] M. Tiberti et al. “PyInteraph: A framework for the analysis of interaction networks in structural ensembles of proteins”. *Journal of Chemical Information and Modeling* 54 (**2014**), 1537–1551. DOI: <https://doi.org/10.1021/ci400639r>.
- [65] N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein. “MDAnalysis: A toolkit for the analysis of molecular dynamics simulations”. *Journal of Computational Chemistry* 32 (**2011**), 2319–2327. DOI: <https://doi.org/10.1002/jcc.21787>.
- [66] R. J. Gowers et al. “MDAnalysis: A Python package for the rapid analysis of molecular dynamics simulations”. *Proceedings of the 15th Python in Science Conference - SciPy, Austin, TX* (**2016**), 98–105. DOI: <https://doi.org/10.25080/Majora-629e541a-00e>.
- [67] M. Biswas, J. Langowski, and T. C. Bishop. “Atomistic simulations of nucleosomes”. *WIREs Computational Molecular Science* 3 (**2013**), 378–392. DOI: <https://doi.org/10.1002/wcms.1139>.
- [68] K. Zhou, G. Gaullier, and K. Luger. “Nucleosome structure and dynamics are coming of age”. *Nature Structural and Molecular Biology* 26 (**2019**), 3–13. DOI: [10.1038/s41594-018-0166-x](https://doi.org/10.1038/s41594-018-0166-x).
- [69] J. Huertas and V. Cojocaru. “Breaths, twists, and turns of atomistic nucleosomes”. *Journal of Molecular Biology* 433 (**2021**), 166744. DOI: <https://doi.org/10.1016/j.jmb.2020.166744>.
- [70] C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, and T. J. Richmond. “X-ray structure of the nucleosome core particle, NCP147, at 1.9 Å resolution”. **2002**. DOI: <https://doi.org/10.2210/pdb1KX5/pdb>.
- [71] A. Šali and T. L. Blundell. “Comparative protein modelling by satisfaction of spatial restraints”. *Journal of Molecular Biology* 234 (**1993**), 779–815. DOI: <https://doi.org/10.1006/jmbi.1993.1626>.
- [72] M. D. Hanwell et al. “Avogadro: an advanced semantic chemical editor, visualization, and analysis platform”. *Journal of Cheminformatics* 4 (**2012**). DOI: <https://doi.org/10.1186/1758-2946-4-17>.

- [73] W. Humphrey, A. Dalke, and K. Schulten. “VMD: Visual molecular dynamics”. *Journal of Molecular Graphics* 14 (**1996**), 33–38. DOI: [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- [74] A. K. Shaytan et al. “Coupling between histone conformations and DNA geometry in nucleosomes on a microsecond timescale: Atomistic insights into nucleosome functions”. *Journal of Molecular Biology* 428 (**2016**), 221–237. DOI: <https://doi.org/10.1016/j.jmb.2015.12.004>.
- [75] J. C. Phillips et al. “Scalable molecular dynamics on CPU and GPU architectures with NAMD”. *The Journal of Chemical Physics* 153 (**2020**), 044130. DOI: <https://doi.org/10.1063/5.0014475>.
- [76] K. Hart N. Foloppe, C. M. Baker, E. J. Denning, L. Nilsson, and A. D. MacKerell Jr. “Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium”. *Journal of Chemical Theory and Computation* 8 (**2012**), 348–362. DOI: <https://doi.org/10.1021/ct200723y>.
- [77] J. Yoo and A. Aksimentiev. “New tricks for old dogs: Improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions”. *Physical Chemistry Chemical Physics* 20 (**2018**), 8432–8449. DOI: <https://doi.org/10.1039/C7CP08185E>.
- [78] F. Pedregosa et al. “Scikit-learn: Machine learning in Python”. *Journal of Machine Learning Research* 12 (**2011**), 2825–2830.
- [79] P. Virtanen et al. “SciPy 1.0: Fundamental algorithms for scientific computing in Python”. *Nature Methods* 17 (**2020**), 261–272. DOI: <https://doi.org/10.1038/s41592-019-0686-2>.
- [80] P. Liu, D. K. Agrafiotis, and D. L. Theobald. “Fast determination of the optimal rotational matrix for macromolecular superpositions”. *Journal of Computational Chemistry* 31 (**2010**), 1561–1563. DOI: <https://doi.org/10.1002/jcc.21439>.

- [81] X.-J. Lu and W. K. Olson. “3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures”. *Nucleic Acids Research* 31 (2003), 5108–5121. DOI: <https://doi.org/10.1093/nar/gkg680>.
- [82] R. Kumar and H. Grubmüller. “do_x3dna: a tool to analyze structural fluctuations of dsDNA or dsRNA from molecular dynamics simulations”. *Bioinformatics* 31 (2015), 2583–2585. DOI: <https://doi.org/10.1093/bioinformatics/btv190>.
- [83] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond. “Crystal structure of the nucleosome core particle at 2.8 Å resolution”. *Nature* 389 (1997), 251–260. DOI: <https://doi.org/10.1038/38444>.
- [84] C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, and T. J. Richmond. “Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution”. *Journal of Molecular Biology* 319 (2002), 1097–1113. DOI: [https://doi.org/10.1016/S0022-2836\(02\)00386-8](https://doi.org/10.1016/S0022-2836(02)00386-8).
- [85] R. Ettig, N. Kepper, R. Stehr, G. Wedemann, and K. Rippe. “Dissecting DNA-histone interactions in the nucleosome by molecular dynamics simulations of DNA unwrapping”. *Biophysical Journal* 101 (2011), 1999–2008. DOI: <https://doi.org/10.1016/j.bpj.2011.07.057>.
- [86] A. Elbahnsi, R. Retureau, M. Baaden, B. Hartmann, and C. Oguey. “Holding the nucleosome together: A quantitative description of the DNA-histone interface in solution”. *Journal of Chemical Theory and Computation* 14 (2018), 1045–1058. DOI: <https://doi.org/10.1021/acs.jctc.7b00936>.
- [87] M. A. Hall et al. “High-resolution dynamic mapping of histone-DNA interactions in a nucleosome”. *Nature Structural and Molecular Biology* 16 (2009), 124–129. DOI: <https://doi.org/10.1038/nsmb.1526>.

- [88] D. Winogradoff and A. Aksimentiev. “Molecular mechanism of spontaneous nucleosome unraveling”. *Journal of Molecular Biology* 431 (2019), 323–335. DOI: <https://doi.org/10.1016/j.jmb.2018.11.013>.
- [89] G. A. Armeev, A. S. Kniazeva, G. A. Komarova, M. P. Kirpichnikov, and A. K. Shaytan. “Histone dynamics mediate DNA unwrapping and sliding in nucleosomes”. *Nature Communications* 12 (2021). DOI: <https://doi.org/10.1038/s41467-021-22636-9>.
- [90] K. Chakraborty and S. M. Loverde. “Asymmetric breathing motions of nucleosomal DNA and the role of histone tails”. *The Journal of Chemical Physics* 147 (2017). DOI: <https://doi.org/10.1063/1.4997573>.
- [91] Y. Chen et al. “Asymmetric unwrapping of nucleosomal DNA propagates asymmetric opening and dissociation of the histone core”. *Proceedings of the National Academy of Sciences of the United States of America* 114 (2017), 334–339. DOI: <https://doi.org/10.1073/pnas.1611118114>.
- [92] Z. Li and H. Kono. “Distinct roles of histone H3 and H2A tails in nucleosome stability”. *Scientific Reports* 6 (2016). DOI: <https://doi.org/10.1038/srep31437>.
- [93] K. Andresen, I. Jimenez-Useche, S. C. Howell, C. Yuan, and X. Qiu. “Solution scattering and FRET studies on nucleosomes reveal DNA unwrapping effects of H3 and H4 tail removal”. *PLoS ONE* 8 (2013). DOI: <https://doi.org/10.1371/journal.pone.0078587>.
- [94] N. P. Nurse, I. Jimenez-Useche, I. T. Smith, and C. Yuan. “Clipping of flexible tails of histones H3 and H4 affects the structure and dynamics of the nucleosome”. *Biophysical Journal* 104 (2013), 1081–1088. DOI: <https://doi.org/10.1016/j.bpj.2013.01.019>.
- [95] K. Luger and T. J. Richmond. “DNA binding within the nucleosome core”. *Current Opinion in Structural Biology* 8 (1998), 33–40. DOI: [https://doi.org/10.1016/S0959-440X\(98\)80007-9](https://doi.org/10.1016/S0959-440X(98)80007-9).

- [96] S. Pepenella, K. J. Murphy, and J. J. Hayes. “Intra- and inter-nucleosome interactions of the core histone tail domains in higher-order chromatin structure”. *Chromosoma* 123 (2014), 3–13. DOI: <https://doi.org/10.1007/s00412-013-0435-8>.
- [97] S. Pepenella, K. J. Murphy, and J. J. Hayes. “A distinct switch in interactions of the histone H4 tail domain upon salt-dependent folding of nucleosome arrays”. *Journal of Biological Chemistry* 289 (2014), 27342–27351. DOI: <https://doi.org/10.1074/jbc.M114.595140>.
- [98] R. Collepardo-Guevara et al. “Chromatin unfolding by epigenetic modifications explained by dramatic impairment of internucleosome interactions: A multiscale computational study”. *Journal of the American Chemical Society* 137 (2015), 10205–10215. DOI: <https://doi.org/10.1021/jacs.5b04086>.
- [99] S. Bilokapic, M. Strauss, and M. Halic. “Cryo-EM of nucleosome core particle interactions in trans”. *Scientific Reports* 8 (2018). DOI: <https://doi.org/10.1038/s41598-018-25429-1>.
- [100] B. Brower-Toland et al. “Specific contributions of histone tails and their acetylation to the mechanical stability of nucleosomes”. *Journal of Molecular Biology* 346 (2005), 135–146. DOI: <https://doi.org/10.1016/j.jmb.2004.11.056>.
- [101] M. S. Shukla et al. “The docking domain of histone H2A is required for H1 binding and RSC-mediated nucleosome remodeling”. *Nucleic Acids Research* 39 (2011), 2559–2570. DOI: <https://doi.org/10.1093/nar/gkq1174>.
- [102] G. Nahum et al. “A ratchet mechanism of transcription elongation and its control”. *Cell* 120 (2005), 183–193. DOI: <https://doi.org/10.1016/j.cell.2004.11.045>.
- [103] C. Hodges, L. Bintu, L. Lubkowska, M. Kashlev, and C. Bustamante. “Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II”. *Science* 325 (2009), 626–628. DOI: <https://doi.org/10.1126/science.1172926>.

- [104] J. Jin et al. “Synergistic action of RNA polymerases in overcoming the nucleosomal barrier”. *Nature Structural and Molecular Biology* 17 (2010), 745–752. DOI: <https://doi.org/10.1038/nsmb.1798>.
- [105] K. Brogaard, L. Xi, J.-P. Wang, and J. Widom. “A map of nucleosome positions in yeast at base-pair resolution”. *Nature* 486 (2012), 496–501. DOI: <https://doi.org/10.1038/nature11142>.
- [106] S. S. Teves, C. M. Weber, and S. Henikoff. “Transcribing through the nucleosome”. *Trends in Biochemical Sciences* 39 (2014), 577–586. DOI: <https://doi.org/10.1016/j.tibs.2014.10.004>.
- [107] Y. Arimura, H. Tachiwana, T. Oda, M. Sato, and H. Kurumizaka. “Structural analysis of the hexasome, lacking one histone H2A/H2B dimer from the conventional nucleosome”. *Biochemistry* 51 (2012), 3302–3309. DOI: <https://doi.org/10.1021/bi300129b>.
- [108] G. N. Rychkov et al. “Partially assembled nucleosome structures at atomic detail”. *Biophysical Journal* 112 (2017), 460–472. DOI: <https://doi.org/10.1016/j.bpj.2016.10.041>.
- [109] S. Bilokapic, M. Strauss, and M. Halic. “Histone octamer rearranges to adapt to DNA unwrapping”. *Nature Structural and Molecular Biology* 25 (2018), 101–108. DOI: <https://doi.org/10.1038/s41594-017-0005-5>.
- [110] T. Kameda A. Awazu and Y. Togashi. “Histone tail dynamics in partially disassembled nucleosomes during chromatin remodeling”. *Frontiers in Molecular Biosciences* 6 (2019). DOI: <https://doi.org/10.3389/fmolb.2019.00133>.
- [111] S. Bowerman and J. Wereszczynski. “Effects of macroH2A and H2A.Z on nucleosome dynamics as elucidated by molecular dynamics simulations”. *Biophysical Journal* 110 (2016), 327–337. DOI: <https://doi.org/10.1016/j.bpj.2015.12.015>.
- [112] T. T. M. Ngo, Q. Zhang, R. Zhou, J. G. Yodh, and T. Ha. “Asymmetric unwrapping of nucleosomes under tension directed by DNA local flexibility”. *Cell* 160 (2015), 1135–1144. DOI: <https://doi.org/10.1016/j.cell.2015.02.001>.

- [113] K. K. Sinha, J. D. Gross, and G. J. Narlikar. “Distortion of histone octamer core promotes nucleosome mobilization by a chromatin remodeler”. *Science* 355 (**2017**), eaaa3761. DOI: <https://doi.org/10.1126/science.aaa3761>.
- [114] S. Bilokapic, M. Strauss, and M. Halic. “Structural rearrangements of the histone octamer translocate DNA”. *Nature Communications* 9 (**2018**). DOI: <https://doi.org/10.1038/s41467-018-03677-z>.
- [115] J. P. Armache et al. “Cryo-EM structures of remodeler-nucleosome intermediates suggest allosteric control through the nucleosome”. *eLife* 8 (**2019**). DOI: <https://doi.org/10.7554/eLife.46057>.
- [116] A. Hada et al. “Histone octamer structure is altered early in ISW2 ATP-dependent nucleosome remodeling”. *Cell Reports* 28 (**2019**), 282–294. DOI: <https://doi.org/10.1016/j.celrep.2019.05.106>.
- [117] B.-R. Zhou et al. “Histone H4 K16Q mutation, an acetylation mimic, causes structural disorder of its N-terminal basic patch in the nucleosome”. *Journal of Molecular Biology* 421 (**2012**), 30–37. DOI: <https://doi.org/10.1016/j.jmb.2012.04.032>.
- [118] S. O. Rabdano et al. “Histone H4 tails in nucleosomes: A fuzzy interaction with DNA”. *Angewandte Chemie* 60 (**2021**), 6480–6487. DOI: <https://doi.org/10.1002/anie.202012046>.
- [119] P. D. Dans, J. Walther, H. Gómez, and M. Orozco. “Multiscale simulation of DNA”. *Current Opinion in Structural Biology* 37 (**2016**), 29–45. DOI: <https://doi.org/10.1016/j.sbi.2015.11.011>.
- [120] H. C. Kung and P. H. Bolton. “Structure of a duplex DNA containing a thymine glycol residue in solution”. *Journal of Biological Chemistry* 272 (**1997**), 9227–9236. DOI: <https://doi.org/10.1074/jbc.272.14.9227>.
- [121] K. Miaskiewicz, J. Miller, and R. Osman. “Ab initio theoretical study of the structures of thymine glycol and dihydrothymine”. *International Journal of Radiation Biology* 63 (**1993**), 677–686. DOI: <https://doi.org/10.1080/09553009314552071>.

- [122] S. D. Wang, L. A. Eriksson, and R. B. Zhang. “Dynamics of 5R-Tg base flipping in DNA duplexes based on simulations - Agreement with experiments and beyond”. *Journal of Chemical Information and Modeling* 62 (2022), 386–398. DOI: <https://doi.org/10.1021/acs.jcim.1c01169>.
- [123] T. Dršata, M. Kara, M. Zacharias, and F. Lankaš. “Effect of 8-oxoguanine on DNA structure and deformability”. *Journal of Physical Chemistry B* 117 (2013), 11617–11622. DOI: <https://doi.org/10.1021/jp407562t>.
- [124] U. K. Shigdel et al. “The trajectory of intrahelical lesion recognition and extrusion by the human 8-oxoguanine DNA glycosylase”. *Nature Communications* 11 (2020), 4437. DOI: <https://doi.org/10.1038/s41467-020-18290-2>.
- [125] X. Cheng et al. “Dynamic behavior of DNA base pairs containing 8-oxoguanine”. *Journal of the American Chemical Society* 127 (2005), 13906–13918. DOI: <https://doi.org/10.1021/ja052542s>.
- [126] C. von Sonntag. “Free-radical-induced DNA damage and its repair”. 2006. DOI: <https://doi.org/10.1007/3-540-30592-0>.
- [127] M. Dizdaroglu and P. Jaruga. “Mechanisms of free radical-induced damage to DNA”. *Free Radical Research* 46 (2012), 382–419. DOI: 10.3109/10715762.2011.653969.
- [128] A. Schipler and G. Iliakis. “DNA double-strand-break complexity levels and their possible contributions to the probability for error-prone processing and repair pathway choice”. *Nucleic Acids Research* 41 (2013), 7589–7605. DOI: <https://doi.org/10.1093/nar/gkt556>.
- [129] J. SantaLucia Jr. and D. Hicks. “The thermodynamics of DNA structural motifs”. *Annual Review of Biophysics and Biomolecular Structure* 33 (2004), 415–440. DOI: <https://doi.org/10.1146/annurev.biophys.32.110601.141800>.
- [130] M. Nakata et al. “End-to-end stacking and liquid crystal condensation of 6- to 20-base pair DNA duplexes”. *Science* 318 (2007), 1276–1279. DOI: <https://doi.org/10.1126/science.1143826>.

- [131] G. P. Smith et al. “Backbone-free duplex-stacked monomer nucleic acids exhibiting Watson–Crick selectivity”. *Proceedings of the National Academy of Sciences of the United States of America* 115 (2018), E7658–E7664. DOI: <https://doi.org/10.1073/pnas.1721369115>.
- [132] M. Todisco, G. P. Smith, and T. P. Fraccia. “Liquid crystal ordering of DNA Dickerson dodecamer duplexes with different 5’-phosphate terminations”. *Molecular Crystals and Liquid Crystals* 683 (2019), 69–80. DOI: <https://doi.org/10.1080/15421406.2019.1581706>.
- [133] C. Maffeo, B. Luan, and A. Aksimentiev. “End-to-end attraction of duplex DNA”. *Nucleic Acids Research* 40 (2012), 3812–3821. DOI: <https://doi.org/10.1093/nar/gkr1220>.
- [134] P. Banáš et al. “Can we accurately describe the structure of adenine tracts in B- DNA? Reference quantum-chemical computations reveal overstabilization of stacking by molecular mechanics”. *Journal of Chemical Theory and Computation* 8 (2012), 2448–2460. DOI: <https://doi.org/10.1021/ct3001238>.
- [135] F. Häse and M. Zacharias. “Free energy analysis and mechanism of base pair stacking in nicked DNA”. *Nucleic Acids Research* 44 (2016), 7100–7108. DOI: <https://doi.org/10.1093/nar/gkw607>.
- [136] C. H. Mak. “Unraveling base stacking driving forces in DNA”. *The Journal of Physical Chemistry B* 120 (2016), 6010–6020. DOI: <https://doi.org/10.1021/acs.jpcc.6b01934>.
- [137] P. D. Dans et al. “How accurate are accurate force-fields for B-DNA?” *Nucleic Acids Research* 45 (2017), 4217–4230. DOI: <https://doi.org/10.1093/nar/gkw1355>.
- [138] J. Yoo, D. Winogradoff, and A. Aksimentiev. “Molecular dynamics simulations of DNA–DNA and DNA–protein interactions”. *Current Opinion in Structural Biology* 64 (2020), 88–96. DOI: <https://doi.org/10.1016/j.sbi.2020.06.007>.
- [139] A. B. Imeddourene et al. “The intrinsic mechanics of B-DNA in solution characterized by NMR”. *Nucleic Acids Research* 44 (2016), 3432–3447. DOI: <https://doi.org/10.1093/nar/gkw084>.

- [140] E. Protozanova, P. Yakovchuk, and M. D. F.-Kamenetskii. “Stacked-unstacked equilibrium at the nick site of DNA”. *Journal of Molecular Biology* 342 (2004), 775–785. DOI: <https://doi.org/10.1016/j.jmb.2004.07.075>.
- [141] F. Kilchherr et al. “Single-molecule dissection of stacking forces in DNA”. *Science* 353 (2016), aaf5508. DOI: <https://doi.org/10.1126/science.aaf5508>.
- [142] R. M. Abolfath, D. J. Carlson, Z. J. Chen, and R. Nath. “A molecular dynamics simulation of DNA damage induction by ionizing radiation”. *Physics in Medicine & Biology* 58 (2013), 7143–7157. DOI: <https://doi.org/10.1088/0031-9155/58/20/7143>.
- [143] R. M. Abolfath, D. Grosshans, and R. Mohan. “Oxygen depletion in FLASH ultra-high-dose-rate radiotherapy: A molecular dynamics simulation”. *Medical Physics* 47 (2020), 6551–6561. DOI: <https://doi.org/10.1002/mp.14548>.
- [144] P. de Vera, E. Surdutovich, and A. V. Solov’yov. “The role of shock waves on the biodamage induced by ion beam radiation”. *Cancer Nanotechnology* 10 (2019), 1–31. DOI: <https://doi.org/10.1186/s12645-019-0050-3>.
- [145] M. R. Lieber. “The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway”. *Annual Review of Biochemistry* 79 (2010), 181–211. DOI: <https://doi.org/10.1146/annurev.biochem.052308.093131>.
- [146] P. S. Nerenberg and T. Head-Gordon. “New developments in force fields for biomolecular simulations”. *Current Opinion in Structural Biology* 49 (2018), 129–138. DOI: <https://doi.org/10.1016/j.sbi.2018.02.002>.
- [147] S. S. Wallace. “Enzymatic processing of radiation-induced free radical damage in DNA”. *Radiation Research* 150 (1998), S60–S79. DOI: <https://doi.org/10.2307/3579809>.