Project (2018-): Theory of epigenome regulation: Physico-chemical models of chromatin-based control of genome regulation

7 janvier 2018

TABLE DES MATIÈRES

1 Objectifs **1**

2 Repliement 3D de l'épigénome : approche structurale **3**
2.1 **Etude théorique de l'organisation 3D : modèles de séparation en microphase des états chromatiniens** . . . 3

3 Régulation de l'épigénome : Relation structure/fonction **3**
3.1 **Régulation à grande echelle : modèle de la chromatine vivante** . . . 3
3.1.1 **Prolongements théoriques** . . . 4
3.1.2 **Applications expérimentales potentielles** . . . 4
3.2 **Régulation à petite échelle : modèles moléculaires des états chromatiniens** 4
3.2.1 **Théorie : du nucléosome à la fibre de chromatine et états chromatiniens** . . . 5
3.2.2 **Projets expérimentaux** . . . 6

4 Conclusion **8**

1 OBJECTIFS

The general objective of this research project is to build quantitative models of the epigenomic-based mechanisms of gene regulation involved in the short- and long-term cellular response. Using theoretical approaches from statistical and numerical physics and in close collaboration with experimentalists, we propose to study the "1D" assembly of the different eu- and heterochromatin states along the genome (nucleation, spreading, compartmentalization and mitotic/meiotic inheritance), and their "3D" folding and nuclear organization. At the nucleosomal and gene scales, we are currently developing molecular models accounting for the combined action of chromatin regulators and DNA binding proteins to understand the mechanisms by which chromatin states achieve their functions. At larger scale, we use coarse-grained models to derive the structural and dynamical properties of the epigenome : (i) the regulation of epigenomic domains (ii) their folding into spatial compartments and (iii) finally the coupling between 3D organization and 1D assembly. Our main goal is to provide a general framework to understand how the epigenome is regulated and how its affect gene expression : from its establishment during development and differentiation to its deregulation in diseases. Ultimately we propose to understand how the epigenome (via the chromatin regulators) "sense" environmental, notably metabolic changes as well as genomic changes and how in some cases, it can translate transient stimuli into stable, inheritable expression pattern changes while in other cases it can act as a "buffer" for these perturbations.

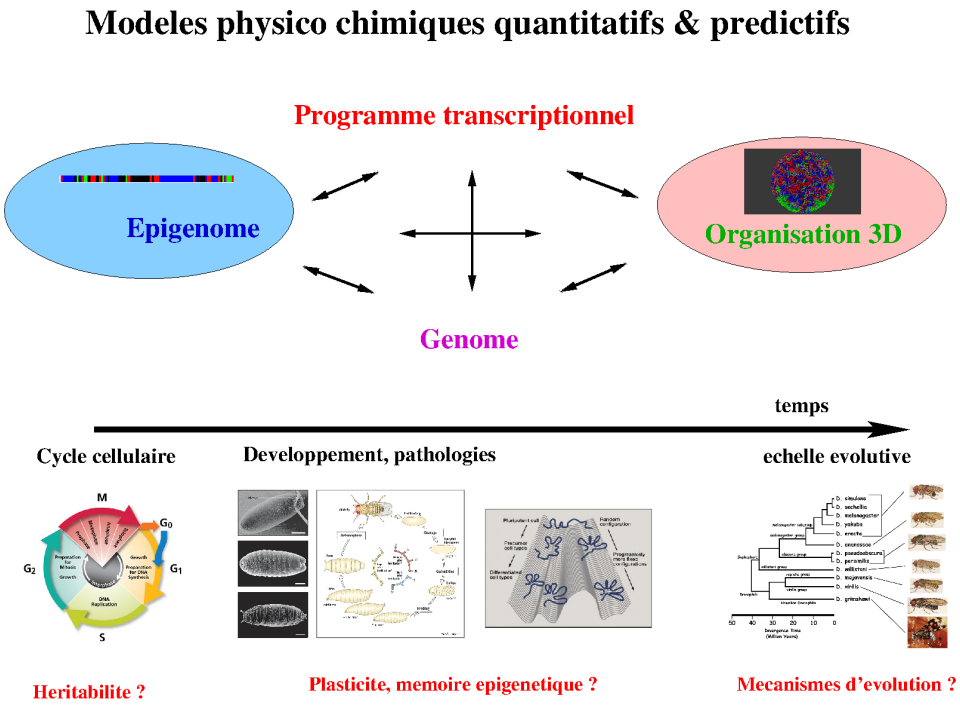

FIGURE 1 – Objectifs généraux

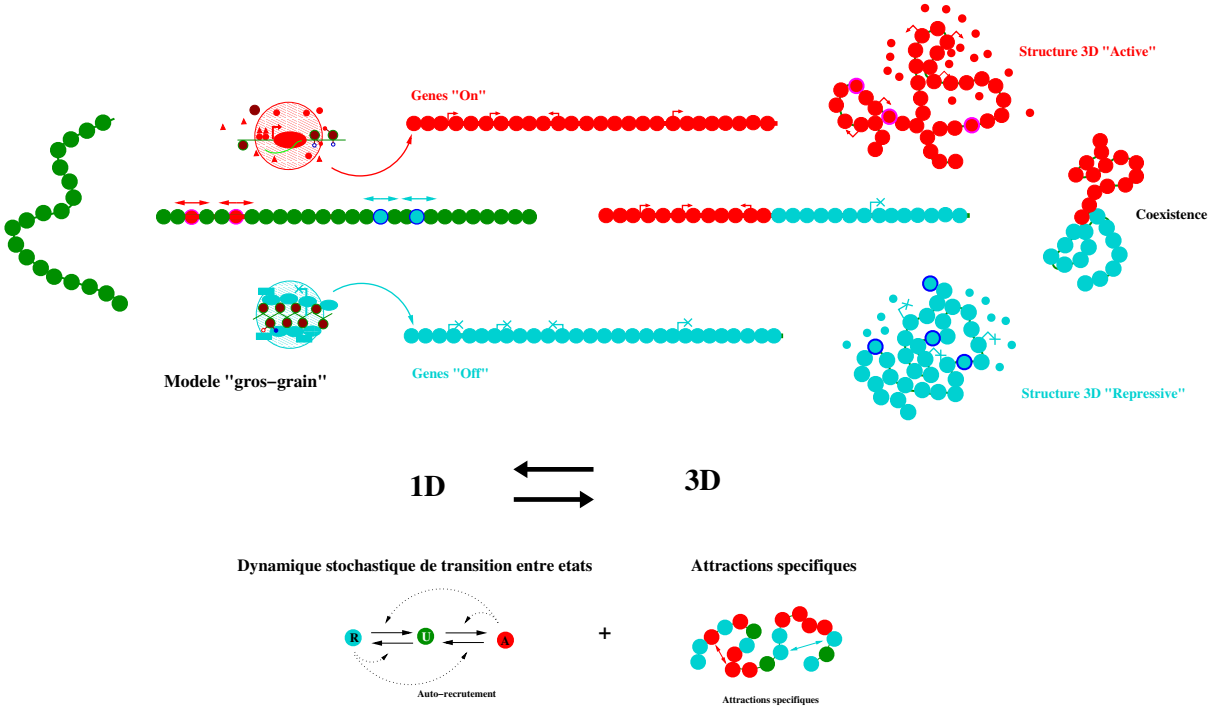

FIGURE 2 – Modèles effectifs de régulation 1D et 3D de l'épigénome

2 REPLIEMENT 3D DE L'ÉPIGÉNOME : APPROCHE STRUCTURALE

2.1 Etude théorique de l'organisation 3D : modèles de séparation en microphase des états chromatiniens Il s'agit ici de continuer nos travaux sur l'organisation spatiale des chromosomes notamment dans le cadre du modèle de copolymère par blocs (cf Rapport Activité) : à savoir poursuivre et enrichir notre interprétation théorique de l'organisation nucléaire comme résultant d'une séparation en microphase, "hors-équilibre" des différents types de chromatines. **A court terme, mon objectif est de soumettre notre papier en cours de rédaction sur la dynamique des chromosomes au cours de l'embryogenèse chez la Drosophile, compilant les résulats obtenus par P. Carrivain. (Projet ANR ("EpiDevoMath" 2015-2018) : Coord. G. Cavalli (IGH, Montpellier), Coll. : D. Jost, R. Everaers.)**. Après ce papier, tout en collaborant, je laisserai la pleine conduite de la modélisation à D. Jost et R. Everaers, plus compétents que moi dans ce domaine de la physique des polymères. L'objectif est d'enrichir le modèle de copolymère pour pouvoir notamment prendre en compte (i) les processus actifs d"'extrusion de boucles" à l'oeuvre notamemnt chez l'homme, (ii) les intéractions avce la membrane. Récemment, dans le cadre de la FRM nous avons montré, avec un modèle de copolymère avec, en plus, interaction avec la membrane nucléaire, que la formation des SAHF accompagnat la senescnce induite (OIS) est le fait d'une diminution des intéraction heterochromatine/heterochromatine et heterochromatine/membrane (Résultats non publiés). Un enjeu ici est de développer des algorithmes rapides pour pouvoir modéliser l'organisation 3D de grands génomes, tel le génome humain.

3 RÉGULATION DE L'ÉPIGÉNOME : RELATION STRUCTURE/FONCTION

3.1 Régulation à grande echelle : modèle de la chromatine vivante Comme précisé plus haut l'étude de la structure des chromosomes en elle-même n'est pas l'objet principal de mon projet. Par contre explorer comment la colocalisation spatiale induite par l'autoassociation des états chromatiniene permet de former des "nano-réacteurs" à savoir des compartiments fonctionnels en augmentant la concentration locale des régulateurs au voisinage de eurs cibles génomique est un projet que je compte continuer à mener, et ce en collaboration avce D. Jost. L'objectif est donc ici de poursuivre notre étude de la régulation des domaines épigénomiques (qui peuvent

etre chromosomiques, dans le cas de la compensation du dosage) dans le cadre notamment du modèle de la chromatine vivante (cf Rapport Activité) pour étudier l'influence de l'organisation spatiale à grande échelle. Dans ce modèle l'épigénome est caractérisé localement par différents états accessibles (ou des couleurs) caractérisant le type chromatinien en toute position génomique. Les taux de transitions entre états ainsi que les attractions entre monomères de même état sont des paramètres de controÎle effectifs libres. Ce type de modèles simples de l'épigénome nous permet d'accéder à des propriétes à grandes échelles, chromosomiques typiquement. Globalement il s'agira toujours d'établir des diagrammes de phases permettant, par comparaison avce des données expérimentales, de restreindre l'espace des paramètres cohérents ou si possible d'inférer ces paramètres. Idéalement la valeur de ces paramètres devraient pouvoir être extraite, tout au moins contraintes par le modèles moléculaires plus fins des différenst états chromatiniens.

3.1.1 **Prolongements théoriques** D'un point de vue théorique, tout en gardant le même modèle de copolymère par blocs, je propose plusieurs extensions à notre étude préliminaire :

1. *Extension à plusieurs couleurs.*
   Dans le modèle initial, nous n'avons considéér que deux éttats en "compétition" Actif vs Inactif (en plus d'un état intermédiaire) ; il s'agira d'étendre à 3 ou 4 états, en cohérence avce les observations expérimentales.
2. *Effet de titration :*
   Du fait d'un nombre limité de protéines architecturales (par ex les proteines hétérochromatiniennes SIR chez *S. cerevisiae*) ou d'un réduction de la concnetration en co-substrats des enzymes régulateurs (ATP, groupement Acetyl,Methyl etc..) suite à un désordre métabolique, le nombre de loci chromatiniens pouvant être associés à un certain état chromatinien peut se trouver être limité (titré). Comment cet effet de titration agit-il sur la stabilité des domaines ?
3. *Effet de la réplication : le modèle du "timing"*
   L'idée ici est de montrer que, du fait de cet effet de titration, les régions répliquées précocement auraient tendance à maintenir leur état chromatinien plus facilement selon la règle du "premier arrivé, premier servi". Des expériences indiquent en effet que des régions actives maiis dont on a retardé la réplication deviennet inactive, le "pool" d'histones acetylés requis pour leur maitien étant en quelque sorte "épuisé".

3.1.2 **Applications expérimentales potentielles** Il s'agit désormais de tester notre modèle de chromatine vivante dans des systèmes biologiques. L'idée qui émerge et que nous avons voulu mettre en avant dans notre papier est que la stabilité des domaines épigénomiques et de leur compartimentation dépend de l'activité enzymatique, de la force d'auto-association entre mêmes états épigénomiques, de la présence de sites de "forçage" (de sites de recrutement) et de la présence de sites génomiques insulateurs. Par l'utilisation de mutants, par ingénierie génomique, ou par des études génomique et épigénomqiue comparatives, il s'agira d'en préciser les différentes contributions.

1. *Compensation du dosage chez* C. elegans (cf Rapport d'activité) (Collab. P. Meister)
2. *Heterochromatine et clustering des télomères chez* S. cerevisiae (Collab. A. Taddei, I. Curie)
3. *Régulation des gènes Hox et plus globalement relation entre clustering et silencing des domaines PcG* (Collab. G. Cavalli & M. Nollman).
4. *Contrôle de la repression d'un chromosome par ingénierie génomique chez S. cerevisiae* (Collab, G. Fourel, LBMC, ENS Lyon)

3.2 Régulation à petite échelle : modèles moléculaires des états chromatiniens Mon projet principal sera de me concentrer de nouveau sur la modélisation de la structure et dynamique de la chromatine à l'échelle des nucléosomes, travail que j'avais laissé en suspend ces dernières années, me focalisant plus sur les grandes échelles en collaboration avec D. Jost. Il y a plusieurs raison à ce recentrage vers les petites échelle : (1) Mes travaux antérieurs font que, au sein de notre activité "Régulation de l'épigénomeé (avec D. Jost, R. Everaers et B. Audit) je suis plus compétent pour aborder ces échelles et de fait plus reconnu dans la communauté

scientifique comme ayant une réelle expertise et (2), le plus important : Qu'il s'agit selon moi de l'échelle pertinente pour mieux comprendre les mécanismes d'activation et de répressions de l'activité transcriptionnelle, mais pas seulement de la réplication et transposition. Il s'agit donc à présent, sur la base de mes travaux précédents, de proposer des modèles moléculaires de ces différents types d' états chomatiniens. L'idée est que ces différents états chromatiniens procurent des degrés d'accessibilités différents aux sites génomiques que ce soit aux polymérases et aux transposases et donc tout l'enjeu sera de pouvoir prédire l'accessibilité associés à ces états. Ces modèles nous renseignerons aussi sur les dynamiques de formation, de propagation et de transition ainsi que sur les modes d'auto-associations de ces états chromatiniens. On pourra, par des méthode de "moyennation" extraire les paramètre effectifs (au moins les ordres de grandeurs) pour les modèles de régulation à plus grande échelle.

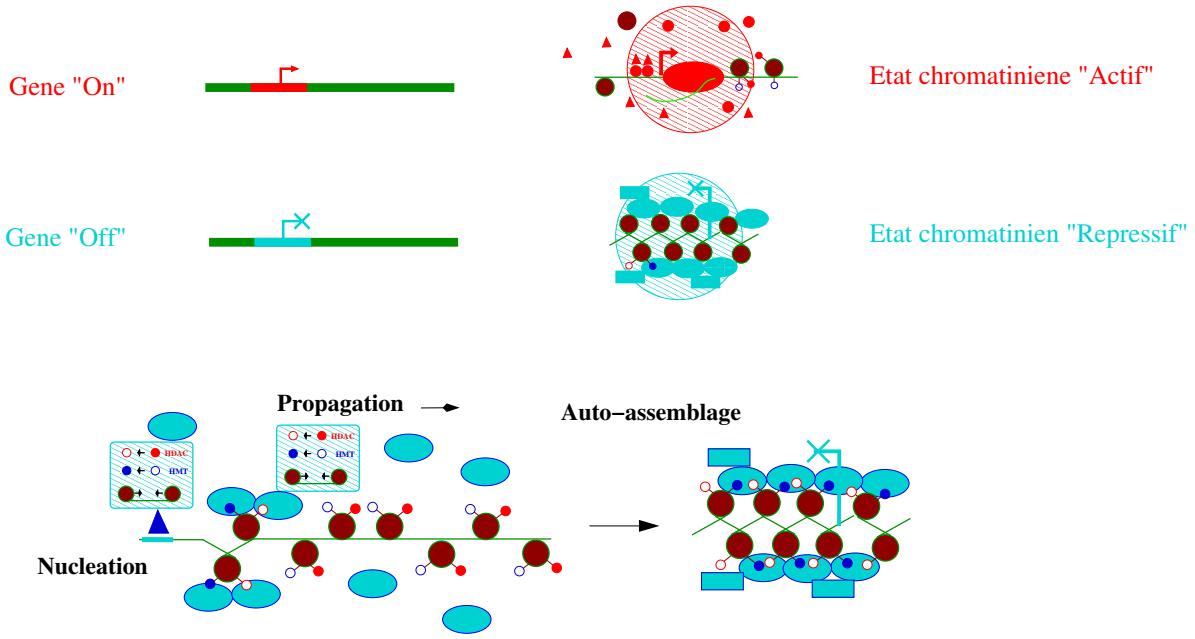

FIGURE 3 – Modèles moléculaires des états chromatiniens.

3.2.1 Théorie : du nucléosome à la fibre de chromatine et états chromatiniens

Collaboration : H. Schiessel (Inst. Lorentz)

— **Modèle 1D étendu du chapelet nucleosomal**
Il s'agit poursuivre notre effort de modélisation de la densité en nucléosemes le long des génomes en incorporant, à notre modèle de base (qui ne prend en compte que l'effet "intrinsèque" de la séquence génomque et l'interaction de volume exclu entre nucléosome, cf Rapport Activité), les ingrédients biologiques "extrinsèques" fondamentaux comme : (1) l'action des facteurs de remodellage (2) la competition avec des facteurs de transcriptions, insulateur ou autres protéines se liant à l'ADN (3) La respiration de l'ADN autour de l'octamère d'histone, (4) le contrôle de la densité globale en histone, (5) l'effet de l'insertion de variant (6) l'effet de des marques d'histones et de la méthylation de l'ADN et (7) l'effet perturbatif de la réplication. L'étude se fera par des méthodes numériques du type produit de Matrices de transferts ou méthodes Monte-Carlo pour extraire les propriétés à l'équilibre thermodynamique. Par ailleurs, un des apects nouveaux sera l'étude de la dynamique de relaxation du chapelet nucleosomal (de la densité locale en nucléosomes) par des simulation de Monte Carlo cinétique. Il sera à ce sujet intéressant d'étudier l'influence de la séquence génomique sur cette dynamique de relaxation (tout comme évidmment son influence sur la densité en nucléosome à l'équilibre). On calculera des profils de densité en nucléosomes et autres protéines (ou les profils d'accessibilté) qu'on pourra comparer à des profils experimentaux.

— **Modèle de fibre de chromatine : positionnement et conformation 3D des nucléosomes**
Jusqu'à présent nous n'avons donc considéré qu'un modèle 1D de positionnement des nucléosomes ; l'accessibilité est aussi controlée par la structure 3D de ces nucléosomes :

pour un même positionnement toutes les configurations 3D ne présentent pas une mêmem accessibilité à l'ADN inter-nucléosomes (les "linkers"). Par ailleurs, le 3D, par encombrement stériques en "volume" risque d'influencer le positionnement 1D des nucléosomes. Donc notre objectif est d'étendre le modèle 1D et de construire un modèle 3D de fibre (par des méthodes de Matrice de transfert et Monte-Carlo) et d'étudier les propriétés de la fibre (positionnement des nucléosomes, accessibilité, degré de compaction locale, propriétés élastiques/plastiques...) en fonction de la séquence génomique, de la densité globale en nucléosomes, de l'interaction histones-histones qui peut être modulée par l'acétylation des queues d'histone, de la présence de l'histone H1, d'autres protéines architecturales hétérochromatiniennes comme HP1 ou PRC1... Une première approche (plus simple !) sera d'étudier la structure de la fibre à positions de nucléosomes fixées : à partir des énergies libres associées on pourra construire des potentiels d'interaction effectifs nucléosomes/nucléosome qu'on pourra intégrer dans notre modélisation 1D. A terme il s'agira de relaxer la position des nucléosome dans notre modélisation 3D.
Ce travail est en cours et est mené par Pascal Carrivain, actuellement post-doctorant (Contrat FRM) (Cf Fig. 4).

— **Hétérochromatine**
Une application théorique pourrait être la modélisation de la dynamique et de la structure de l'hétérochromatine de type HP1/H3K9me3 ou PRC1/H3K27me3 ; dans les deux cas, la fibre se caractérise par une répartition régulière des nucléosomes assurant certainement une plus forte compaction et peut-être une meilleur autoassociation (oligomérisation) des protéines HP1 et PRC1, renforÃğant la stabilité de la fibre et des nucléosomes, réduisant donc l'accessibilité aux sites génomiques. La régularité de position des nucléosomes est assurée par l'action de facteur de remodellage qui sont ciblée par coopérativité moléculaire avce HP1 (PRC1) eux mêmes ciblés aux nucléosomes marqués H3K9me3 via leur bromodomaine. Si on peut effectivement envisager une approche structurale de la fibre d'hétérochromatine, on pourra également aborder le problème avec un modèle 1D qui intègrerait de manière effective l'action des remodelleurs via un potentiel d'attraction entre nucléosomes marqués H3K9me3. On peut ainsi même envisager un modèle de propagation 1D d'états hétérochromatiniens tels ceux utilisés dans nos modèles effectifs de régulation de l'épigénome (cf Rapport d'Activité) mais désormais avec un chapelet nucléosomal "fluide" où la propagation d'une marque en cis dépendrait aussi de la distance génomique entre les deux nucléosomes : c'est toujours l'idée que nous avons formalisée dans notre modèle de chromatine vivante que la propagation d'un état dépend de la colocalisation spatiale ; ici à l'échelle du nucléosome c'est la propagation des marques qui dépend de la distance spatiale enter deux nucléosomes voisins et donc de manière effective de la distance génomique entre ces deux nuléosomes. Donc en partant d'une distribution quelconque de nucléosomes marqués et non marqués, on pourra (par simulation en Monte Carlo Cinétique) étudier à la fois la dynamique de relaxation de la position et des marques (H3K9me2/3...) des nucléosomes. On peut s'attendre à trouver une densité globale en nucléosomes critique en deçà de laquelle un domaine épigénomique ne soit plus stable.
L'objectif est de proposer un projet à l'appel d'offre ANR 2019 qui soit dédié à la modélisation de la chromatine à cette échelle. Le développemnt d'un tel modèle théorique aura un fort potentiel d'application comme l'indiquent déjà les différents projets expérimentaux dans lesquelles ce travail est impliqué (cf ci-après). Cela pourrait également avoir un intérêt en recherche appliquée par exemple dans le cas de l'édition de gènes par le système CRISPR-Cas9 (Isaac, R. S., F. Jiang, ..., R. Almeida. 2016. Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. eLife 5 :e13450).

3.2.2 Projets expérimentaux

- **Epigenome reprogramming during spermatogenesis** *Collaboration with Saadi Khochbin, Institut A. Bonniot,Genoble*

  En collaboration avec S. Khochbin qui étudie le processus de reprogrammation de l'épigénome lors de la spermatogenèse, et notamment le remplacement des histones par les protamines. L'objectif sera notamment ici de modéliser ce processus de remplacement et de comprendre pourquoi certaines régions génomiques semblent préserver leurs nucléo-

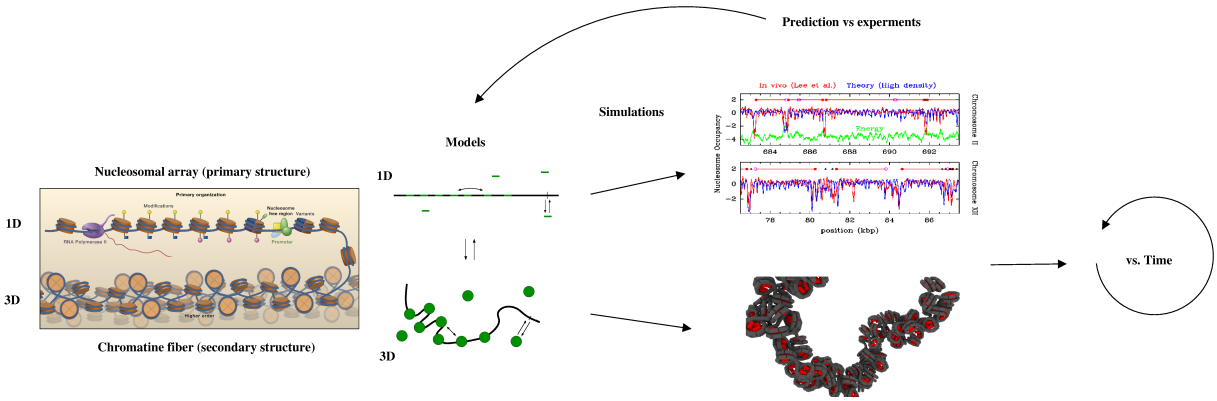

FIGURE 4 – Approches 1D et 3D : Les modèles 1D et 3D permettront d'avoir accès aux densités en nucléosomes (haut) et à la la structure spatiale (bas), et ce, à l'équilibre ou hors équilibre, au cours de la phase de relaxation après une perturbation (par ex. réplication) ou du fait d'un forçage extérieur, par exemple au cours du cycle circadien.

somes.

- **Circadian Regulation of Chromatin** *Collaboration with K. Padmanabhan, Institut de Genomique Fonctionnelle*

  C'est une collaboration débutée il y a peu avec le groupe de K. Padmanabhan, qui consiste à modéliser la dynamique de la chromatine au cours du cycle circadiens notamment au niveau des gènes qui sont contrôlés par le complexe protéique BMAL1 :CLOCK :PER :CRY qui s'assemble dynamiquement au cours du cycle. Ce projet a fait l'objet d'un pré-proposition à l'ANR 2018.

- **Chromatin structure as a constraint to genome evolution in vertebrates** *Collaboration with J.N Volff, Institut de Genomique Fonctionnelle, B. Audit Lab. Phys. ENSL*

  La chromatine peut contraindre l'évolution et la sélection naturelle. Cependant, la structure de la chromatine est bien plus difficile à comparer entre espèces que les régions codantes. Basé sur le calcul du coût d'énergie libre de la courbure d'un fragment d'ADN, un modèle physique simple d'assemblage de nucléosomes rend compte des données d'occupation de nucléosomes chez l'homme. Avec ce modèle, nous proposons de caractériser chez les Vertébrés les mécanismes évolutifs de l'agencement intrinsèque des nucléosomes, et la conservation fonctionnelle et la dynamique évolutive des barrières énergétiques inhibant la formation des nucléosomes ("Nucleosome Inhibiting Energy Barriers",NIEBs). Les bords des NIEBs étant associés à des séquences Alu chez l'homme, nous analyserons le lien entre organisation intrinsèque de la chromatine et éléments transposables. Ce projet apportera une nouvelle lumière sur l'évolution de l'architecture de la chromatine chez les vertébrés, avec un rôle potentiellement "utile" pour "l'ADN parasite".

- **Comprendre les bases physicochimiques de la réorganisation de la chromatine lors de la sénescence cellulaire**

  (Projet "Fondation pour la Recherche Médicale", Coord. CAVALLI Giacomo, (2016-2018).) (Cf Rapport d'Activité)

  Durant la senescence la production d'histone est stoppée et il y a surexpression des protéines HMGA. Or HMGA est un compétiteur de l'histone H1 qui est enrichi dans les domaines hétérochromatiniens. Donc l'objectif est étudier l'effet d'un réduction de la densité en nucléosomes aisni que la diminution du niveau de H1 sur les propriétes structurales de la fibre hétérochromatiene. Des mesures par MNase-seq et Chip-seq de l'occupation en Nucleosomes et HMGA dans les cellules proliferatives vs. cellules senescentes sont en prévision. La portée de cette étufe dépasse largment le cadre de la senescence cellulaires induite *in vitro*. Le vieillissement normal des organismes est accompagné d'une réduction de la densité en histones induisant une destabilisation de l'épigénome ; des expériences ont montré qu'augmenter le niveau d'expression des histones accroit la durée de vie (chez la levure...). Pour le cancer, il a été montré récemment (C. Morales Torres *et al.*, Science 353,

aaf1644 (2016)) que l'hétérogénéĺité phénotypique des cellules cancéreuses dans le cas du cancer du sein, notamment l'existence de cellules prolifératives pouvait être imputée à une sous-expression d'une histone H1 induisant une destabilisation de la chromatine dans les régions riches en AT (là où les nucléosomes sont les moins stables) conduisant à l'expression de gènes de prolifération. Tout facteur, que ce soit des chaperones telles CAF-I ou des protéines architecturales telles H1 ..., contrôlant la déposition ou stabilité des histones, et donc contrôlant la densité en nucléosome serait susceptible de jouer un rôle crucial dans la régulation des états chromatiniens et donc de la plasticité cellulaire ainsi que de l'intégrité génomique (via la destabilisation de la chromatine au niveau des éléments transposables). Cette régulation pourrait également jouer un rôle dans le maintien de la pluripotence des cellules ES ainsi que dans le reprogrammation cellulaire.

A travers ce projet on pourra également étudier dans quelle mesure la séquence génomique participe à la régulation des l'épigénome via la spécificité de séquence des nucléosomes, mais aussi via les nombreuses protéines (à "AT hook" telles HMGA1,2, ou à "CXXC domains" telles MLL1,2) qui reconnaissent des petits motifs (AT runs pour les premières et CpG non méthylés pour les secondes) et qui, selon leur niveau d'expression, pourraient "modeler" l'épigénome de façon différentielle (T. Quante an A. Bird, Do short, frequent DNA sequence motifs mould the epigenome ? Mol Cell Biol 17 257-262 (2016)).

C'est donc une étude très fondamentale et certainement très prometteuse. J'envisage de déposer uen ANR sur cet aspect particulier, à savoir comment la densité globale en nucléosome permet de contrôler la plasticité transcriptionnelle via le contrôle de l'épigénome. Et comment la susceptibilité "locale" à ces perturbations "globales" (perturbation de la densité globale en nucleosomes) peut effectivement dépendre du contexte génomique (% AT/GC etc...).

- **Discovery and modeling of epigenetically regulated genomic domains in lung cancer**
  (Projet INSERM, Coord. D. Jost, Coll. : S. Khochbin and E. Brambillat (Institute Albert Bonniot, Grenoble) (2015-2018))

  Les premier résultats de ce projet indiquent une surexpression du gène ATAD2. Ce facteur protéique serait impliqué dans le controĺe du dosage en histones. Or il a été montré par des expériences pilotes menées par le groupe de S. Khochbin dans les cellules ES que ATAD2 controĺe la dynamique de la fibre. Il s'agit donc ici d'étudier comment le controĺe de la densité en histones modifie la dynamique de la fibre ainsi que la stabilité et l'auto-association des états chromatiniens comme dans le projet sur la senescence.

4 CONCLUSION

Ce projet fixe les grandes lignes et orientations de mes activités de recherche future. Ces activités peuvent être pleinement réalisées au Laboratoire de Physique de l'ENS et notamment au sein de l'Equipe 3, qui est une équipe de traitement de données où travaille également Benjamin Audit avec qui je compte collaborer (de nouveau) pour la partie analyse de données génomiques et epigénomiques que je n'ai pas présentée ici. L'inférence de paramètres des modèles évoqués à partir des jeux de données pourra bénéficier de l'expertise de Nelly Pustelnik dans la théorie des problèmes inverses et d'estimation de paramètres. Mon implantation au Centre Blaise Pascal est par ailleurs un atout important pour la mise en place de simulations numériques performantes (parallélisme, technologie GPU...). Je compte sur la réussite des projets ANR déposés et où je suis partenaire pour recruter au moins un doctorant à la rentrée 2018.