Etude fonctionnelle du complexe CBP/TDG/RARα dans la régulation épigénétique de la transcription et la réparation de l'ADN

Couplage entre la réparation de l'ADN, la transcription et l'épigénétisme

Présentée par

Hélène LEGER

Le 24 mai 2012

En vue d'obtenir le grade de docteur en sciences de l'Université Lille 1 Spécialité : Biologie cellulaire et moléculaire.

Composition du jury :

François TRONCHE Brendan BELL Sylviane PIED Hélène COURVOISIER DR, CNRS UMR 7224, Université Pierre et Marie Curie, France PR, Université de Sherbrooke, Sherbrooke, Canada DR CNRS, Institut Pasteur de Lille, France MdC, CNPS-CNRS UMR 8195, Université Paris Sud, France

Rapporteur Rapporteur Membre du jury Membre du jury

Directeurs de Thèse : Arndt BENECKE

Sara MORLEY-FLETCHER MdC, CNRS UMR 8576, Université Lille1, France. CR, CNRS, IHES, Bures sur Yvette, France.

Remerciements

Je tiens à remercier vivement les Dr. Tronche et Bell qui ont accepté de juger mon travail de thèse et ont lu mon mémoire avec tant de rapidité. Un grand merci aussi aux membres du jury qui ont accepté d'être présents à la soutenance de ma thèse pour juger de mes travaux. Je remercie particulièrement le Dr Sylviane Pied qui a accepté de présider ce jury de thèse si éloigné de son domaine d'étude. J'ai une pensée spéciale pour Hélène Courvoisier dont les enseignements et les encouragements à continuer la recherche m'ont mené jusqu'ici. Merci d'avoir transmis, il y a bientôt 4 ans, mon CV à Arndt, me permettant ainsi de me lancer dans l'aventure TDG.

Je tiens à remercier également l'ensemble de nos collaborateurs, l'équipe de Sara Morley-Fletcher de Lille1, de Jean-Claude Sirard et Yvan de Launoit de l'IBL, de Maria-Christina Zennaro de l'HEGP et de Brendan Bell de l'Université de Sherbrooke avec qui j'ai travaillé pendant ces années. Je suis tout particulièrement reconnaissante envers Maria-Christina Zennaro et son équipe, Sheerazed, Edwige, Benoit, Fabio, Isabelle, Amanda ainsi que les autres membres de l'équipe du Dr. Xavier Jeunemaitre pour nous avoir si bien accueillis pendant plus d'un an. Je remercie également Sara Morley-Fletcher et Arndt Benecke, mes directeurs de thèse, qui m'ont accepté comme étudiant et guidé dans mes travaux.

Un grand merci à mes collègues lillois de l'IRI et du C9, Caroline Smet-Nocca, François-Xavier, Guillaume, Sebastian pour nos discussions scientifiques ou non, nos fous rires et ainsi que les soirées barbecues. Un merci particulier à Sebastian pour notre approvisionnement en bière allemande ainsi que pour son aide et son soutien au cours de nos diverses péripéties parisiennes. "Ich wollte mir bei Sebastian ihrer Hilfe und der Unterstützung während unserer Schicksalswenden vor allem bedanken, die Pariser ist und während meiner Doktorarbeit". Merci à mes collègues de l'IHES, Felipe, Bryce, Christophe et Nicolas pour leur contribution au développement d'ace.map dont j'ai été fière d'être la testeuse de bug naïve, pour nos discussions enflammées à l'IHES pendant tea times ainsi que pour les soirées cointreaupolitain (n'est-ce pas Brice !). Enfin merci aux membres de l'équipe de Yves Levy à Créteil pour leur chaleureux accueil et les dégustations de chocolat. Une pensée toute particulière à Clarisse et Julie pour leurs conseils, à Vedran, Mathieu, Aurélie et Ayrin pour leur gentillesse et les discussions enflammées dans le bureau et à Pascaline pour nos échanges cinématographiques.

Je tiens aussi à remercier mes amis de Lille qui avaient ensoleillé mon séjour làbas, les filles du rugby, Benoit, ainsi que Suzanne et Joanna pour nos footing, nos soirées entre filles et tous ces bons moments passés et futurs. Une pensée affectueuse pour vos moitié Régis et Dominique.

Sans oublier mes amis de Paris et d'ailleurs et mes loulous des *Gaillards Parisiens* pour tous ces entrainements de folie, ces matchs endiablés et surtout pour toutes nos belles soirées. Une pensée pour Sylvie Laurent pour son soutien quand j'ai décidé de faire de la biologie et non des sciences économiques.

Enfin un immense merci à ma famille, et surtout à mes parents et à mes deux sœurs adorées qui m'ont encouragée et soutenue depuis toujours, quelque soit les difficultés et les réussites. C'est un peu votre thèse aussi. Merci à la famille Prévot pour les bons moments et votre soutien surtout lors de la dernière ligne droite. Et enfin et surtout, merci à mon Régis. Comme sur le terrain, tu m'as encouragée, soutenue quand il le fallait et tu as même passé « l'éponge magique » parfois. Merci pour tout cela et pour le reste. Je suis impatiente de vivre d'autres aventures à tes côtés.

Etude fonctionnelle du complexe CBP/TDG/ RARα dans la régulation épigénétique de la transcription et la réparation de l'ADN

Résumé

Une perte locale ou globale de l'information épigénétique est souvent un facteur significatif dans les désordres génétiques notamment dans les cancers car elle mène à une dérégulation de l'expression des gènes.

Des études précédentes ont mis en évidence un lien, in vitro, entre la réparation de l'ADN des mésappariements G : U/T par la Thymine DNA Glycosylase (TDG) et les mécanismes de transcription via les récepteurs à l'acide rétinoïque (RAR) et le corégulateur transcriptionnelle Creb Binding Protein (CBP). Ces mésappariements G : U/T, produits lors de la déamination des paires G: meC (cytosine méthylée) dans les îlots CpG, contribuent à une déméthylation de l'ADN. TDG possède une fonction de régulation épigénétique de la transcription et joue donc un rôle essentiel dans les processus de différenciation et de développement. Mes travaux de thèse ont porté sur deux aspects de TDG : (i) l'étude de la régulation de la fonction de réparation de l'ADN de TDG via son interaction avec la protéine SUMO-1 et (ii) l'étude de la régulation de la transcription des gènes par TDG au sein des complexes CBP/TDG et CBP/TDG/RAR α dans les cellules.

L'interaction TDG/SUMO-1 a été caractérisée par RMN et à l'aide de différents mutants de TDG et à mis en évidence l'impact de TDG sur le turnover G/U/T de TDG. SUMO-1 permet donc à TDG de se détacher plus facilement de l'ADN et donc de compenser son faible turnover. Par ailleurs, j'ai confirmé l'interaction in vivo entre CBP et TDG et j'ai mis en évidence la présence d'un complexe tertiaire circulaire entre CBP, TDG et RARa. Par ailleurs, l'utilisation de mutants de TDG ont permis d'étudier le rôle transcriptionnel de TDG via son interaction avec CBP et RARa. Le complexe formé par CBP, TDG et RARa constitue donc le premier lien entre la réparation d'ADN par excision de base, la transcription, et l'épigenèse et est ainsi une nouvelle voie de régulation de l'intégrité et de l'expression génomique. En parallèle, nous avons étudiés les mécanismes épigénomiques *in vivo* par l'utilisation d'un modèle du stress prénatal chez le rat en collaboration avec le Dr. Morley-Fletcher. Ces données montrent que le stress prénatal déclenche une programmation de certains gènes impliqués dans le façonnage du phénotype vulnérable qui caractérise ce modèle animal et que cette programmation de ces gènes est transmise aux cours des générations. Par ailleurs, grâce à mon expertise en analyse transcriptomique, j'ai participé à cinq collaborations portant sur (i à iii) l'étude des mécanismes de défense innés de l'organisme face à une infection bactérienne à Streptococcus pneumoniae. Ces études ont démontré les infections bactériennes à Streptococcus pneumoniae induisent une réaction rapide de la signalisation TLR, de l'IL-17 ou de nombreux gènes stimulés par les interferons. Ces données permettent de mieux comprendre comment les infections à bactéries extracellulaires encapsulées sont contrôlées par les organismes, permettant ainsi de concevoir de nouvelles approches thérapeutiques. (iv) L'exploration des gènes régulés par HMGA1 et (v) par TAF6delta à l'échelle génomique. Ces données ont permis de montrer l'activité régulatrice de (iv) 7SK ARN L2, de HMGA1 et de (v) TAF 66 notamment sur la transcription des gènes régulés par HIV.

Mots clés : TDG, CBP, RARα, épigénétisme, transcription, réparation de l'ADN.

Functional studies of CBP/TDG/RAR complex reveal a close link between DNA repair, transcription, and epigenetic signaling

Abstract

The packaging of eukaryotic DNA into chromatin represents an essential organizational and an important regulatory feature. Chromatin structure can specifically contribute positively or negatively to the correct assembly of transcription factors and their activity. So a loss of epigenetic information is often a significant factor in genetic disorders including cancers because it leads to a deregulation of gene expression.

Previous studies showed an interaction between Thymine DNA Glycosylase (TDG) and the transcription factor retinoic acid receptor (RAR) and between TDG and the general coregulator of the transcription, the CREB Binding Protein (CBP). Furthermore, TDG has a potency to repair G: T mismatches, produced by deamination of G: meC (methylated cytosine) in CpG islands, to restore a G: C pair so TDG contributes to indirect demethylation of DNA. TDG thus playing an essential role in the process of differentiation and development. In addition, the malignant transformation associated with mutations of CBP could be due to dysfunction of major signaling pathways that require CBP as coactivators, such as retinoic acid pathway.

The aim of my work is to characterize this functional complex in transcriptional regulation and epigenetics and understand the genomic plasticity induced by the CBP/TDG complex. During my thesis, (i) I continue to characterized, in collaboration, the regulation of TDG by SUMO-1 and (ii) I studied the involvement of TDG in regulation of retinoic acid-dependent genes through its interaction with CBP and RAR in cells overexpressed for CBP and TDG in presence of retinoic acid.

We demonstrated that SUMO-1 increases the enzymatic turnover of TDG through a competitive DNA binding activity of SUMO-1 towards the regulatory domain of TDG. Moreover I showed that CBP, TDG and RAR formed a regulated tertiary complex in cellular extract which modified the transcriptional profile of our cellular models. CBP/TDG/RAR α complex is thus a new link between DNA repair by base excision and epigenetic regulation of the transcription. We also studied epigenetic mechanisms of prenatal restraint stress (PRS) in rat in collaboration with Dr. Morley-Fletcher. PRS induce

change of heritable expression profil of some gene involved in anxious/depressive phenotype.

Moreover, I used my expertise in microarrays experiment and transcriptome analyze to work on several projects about (i to iii) the study of innate defense against Streptococcus pneumoniae bacterial infection. Streptococcus pneumoniae bacterial infections induce an early reponse of TLR, IL-17 and interferons. These data provide a better understanding of control of extracellular encapsulated bacterial infections, allowing the design of new therapeutic approaches. (iv) The study of HMGA1 regulated genes and (v) TAF6delta regulated genes at wide genome level. These data have shown the regulatory activity of (iv) 7SK RNA L2, HMGA1 and (v) TAF 6δ including the regulation of genes regulated by HIV.

Key words: CBP, TDG, RARa, epigenetism, transcription, DNA repair.

Liste des publications

I_ Etude fonctionnelle du complexe CBP/TDG/RARa dans la régulation épigénétique de la transcription et la réparation de l'ADN

Smet-Nocca C, Wieruszeski J-M, <u>Léger H</u>, Eilebrecht S, Benecke A. SUMO-1 regulates the conformational dynamics of thymine-DNA Glycosylase regulatory domain and competes with its DNA binding activity. BMC Biochem. 2011 1; 12:4

<u>Léger H</u>, Eilebrecht S, Smet-Nocca C, Attmane A, Benecke A. Genome-wide activity of human Thymine-DNA-Glycosylase-containing transcription complexes in Retinoic Acid induced gene expression. Under soumission process

II_ Acquisition de profils transcriptomiques pour l'étude des mécanismes de réponse à une infection par Streptococcus pneumoniae.

Van Maele L, Carnoy C, Cayet D, Songhet P, Dumoutier L, Ferrero I, Janot L, Erard F, Bertout J, <u>Léger H</u>, Sebbane F, Benecke A, Renauld JC, Hardt WD, Ryffel B, Sirard JC. TLR5 signaling stimulates the innate production of IL-17 and IL-22 by CD3 (neg) CD127+ immune cells in spleen and mucosa. J Immunol 2010 15; 185 (2) : 1177-85

Marqués J. M, Rial A, Muñoz N, Pellay F-X, Van Maele L, <u>Léger H</u>, Camou T, Sirard J-C, Benecke A, Chabalgoity J. A. Protection against Streptococcus pneumoniae serotype 1 acute infection shows a signature of Th17- and IFN-γ- mediated immunity. Immunobiology 2011.

Ivanov S, Fontaine J, Paget C, Macho-Fernandez E, Van Maele L, Renesson J, Maillet I, Muñoz N, Wolf, Rial A, <u>Léger H</u>, Ryffel B, Chabalgoity J. A, Sirard J-C, Benecke A, Faveeuw C and Trottein F. Key role for respiratory dendritic cells, IFN- γ and IL-17 in protection against acute Streptococcus pneumoniae pneumonia in response to α galactosylceramide. Accepted pending minor revision. III _ Exploration des gènes régulés par HMGA1 (en collaboration avec le Dr. Eilebrecht) et par TAF6delta (en collaboration avec le professeur Bell de l'Université de Sherbrooke (Qc, Canada)) à l'échelle génomique.

Eilebrecht S, Bécavin C, <u>Léger H</u>, Benecke BJ, Benecke A. HMGA1-dependent and independent 7SK RNA gene regulatory activity. RNA Biol. 2011 1; 8 (1):143-57.

Sommaire

Remerciements	1
Résumé	3
Abstract	5
Liste des publications	7
I_ Etude fonctionnelle du complexe CBP/TDG/RARα dans la régrépigénétique de la transcription et la réparation de l'ADN	ulation 7
II_ Acquisition de profils transcriptomiques pour l'étude des mécanisn réponse à une infection par Streptococcus pneumoniae	nes de 7
III _ Exploration des gènes régulés par HMGA1 (en collaboration avec	le Dr.
Eilebrecht) et par TAF6delta (en collaboration avec le professeur Bell de l'Univer Sherbrooke (Qc, Canada)) à l'échelle génomique	sité de 8
Sommaire	9
Introduction	14
Chapitre I _ De la régulation de l'expression des gènes à la mise en plac organisme multicellulaire	e d'un 14
I.1. Régulation de l'expression génique	14
I.1.a Transcription de gènes et régulation de l'expression génique	14
I.1.b Structure du génome et régulation de la transcription des gènes	16
I.2. Mécanismes épigénétiques et différenciation cellulaire	20
I.2.a. La méthylation de l'ADN : marqueur épigénétique de la répr transcriptionnelle	ression 21
I.2.b Méthylation et empreinte parentale	25
I.2.c. Organisation structurale et fonctionnelle de la chromatine	27
I.3. Couplage entre régulation épigénétique, transcription des gènes et répa de l'ADN	aration 30
I.3.a Couplage entre épigénétique et oncogenèse	31

I.3.b Couplage entre la transcription et la réparation de l'ADN	
Chapitre II _ Les composants du complexe Creb Binding Protein/Thymi Glycosylase/Récepteur à l'Acide Rétinoïque (CBP/TDG/RAR)	ne DNA
II.1. La Thymine DNA Glycosylase (TDG)	40
II.1.a Structure et fonctions de TDG	
II.1.b TDG et la régulation de la transcription	45
II.2. La Creb Binding Protein (CBP ou CREBBP)	
II.2.a Structure et fonctions de CBP	
II.2.b Maladies liées à CBP	52
II.2.c Interaction CBP/TDG et ouverture vers un transcription/réparation de l'ADN	couplage
II.3. Le Récepteur à l'Acide Rétinoïque (RAR : RXR)	55
II.3.a Structure des récepteurs à l'acide rétinoïque RAR et RXR	55
II.3.b Implication dans la morphogenèse et l'oncogenèse	57
II.3.c Complexe CBP/TDG/RARa : cœur de mes travaux	60
Chapitre III _ L'analyse de données transcriptomiques par microarrays	
III.1. Présentation de la technologie des puces à ADN	
III.2. Les puces de la Technologie Applied Biosystems	65
III.2.a Caractéristiques des puces	65
III.2. b- Caractéristique du mode de lecture	68
III.3. Analyse transcriptomique par microarrays	69
Matériels et méthodes	71
I_Matériel	71
I.1. Description des lignées cellulaires utilisées	71
I.2. Description des plasmides utilisés	72
I.3. Description des anticorps utilisés	73
I.4. Description des traitements utilisés	74

II _ Méthodes
II. 1. Transformation, amplification de bactéries XL1 blues compétentes76
II. 2. Transfections
II.3. Extractions des ARN totaux78
II.4. Analyse micro-arrays
II.5. Analyse des données avec ace map
II.6. Extraction des protéines cytoplasmiques et nucléaires de cellules
transfectées
II.7. Dialyse des extraits protéiques totaux
II.8. Immunoprécipitation des protéines taggées
II. 9. Western blot
II. 10. Microthermophorèse
II. 11. Cinétique d'activité glycosylase de TDG
Résultats
Chapitre I _ TDG, un acteur de la réparation de l'ADN mais aussi un corégulateur
transcriptionnel
I.1. Mutagenèse dirigée envers la protéine TDG90
I.2. Les fonctions de TDG sont-elles régulées par sumoylation91
I.2.a Impacts de l'interaction entre SUMO-1 et TDG sur la conformation et la dynamique de TDG
I.2.b Etude des SUMO-binding domains (SBMs)
I.2.c Impacts des interactions covalentes ou non-covalentes sur l'activité glycosylase et le turnover de TDG
I.2.d Impacts des interactions covalentes ou non-covalentes avec SUMO-1 sur
l'interaction entre TDG et ses partenaires tels que CBP et RAR105
Chapitre II _ Le complexe tertiaire CBP/TDG/RARa peut-il expliquer, in vivo,
l'activité régulatrice de TDG sur la transcription et notamment sur la transcription
dependante de l'acide retinoïque ?

II. 1. Analyse du complexe CBP/TDG 113
I.1.a Effets de TDG et de CBP sur le transcriptome des cellules 114
I.1.b Effets de CBP/TDG sur le transcriptome des cellules
II.2. Analyse du complexe tertiaire CBP/TDG/RARa 123
II.2.a Existence d'un complexe CBP/TDG/RARα123
II.2.b Le mutant TDG-P65A, un mutant de fonctionnalité 124
II.2.c Analyse transcriptomique du complexe CBP/TDG/RARa134
Discussion
I _ La régulation de la conformation, de la dynamique et des fonctions de TDG par SUMO-1
SUMO-1 peut réguler la conformation et la dynamique de TDG140
SUMO-1 peut réguler l'activité de réparation de l'ADN de TDG 143
SUMO-1 peut modifier la capacité de TDG à interagir avec ses différents
partenaires
II_ Etude in vivo des complexes CBP/TDG et CBP/TDG/RARα dans la régulation de la transcription
Etude in vivo du complexe CBP/TDG dans la régulation de la transcription 149
Etude in vivo du complexe CBP/TDG/RARa dans la régulation de la
transcription
Conclusion
Perspectives
Annexes
Projet N°1 _ Acquisition de profils transcriptomiques pour l'étude des mécanismes
de réponse à une infection par Streptococcus pneumoniae162
I.1. La signalisation via TLR5 stimule la production naturelle d'IL-17 et d'IL-22
par l'activation des cellules immunitaires CD3 (neg) CD127 + de la rate et de la muqueuse
102

Introduction

Chapitre I _ De la régulation de l'expression des gènes à la mise en place d'un organisme multicellulaire

La mise en place d'un organisme pluricellulaire, appelée morphogenèse, dépend de l'état des cellules de cet organisme au cours du temps c'est-à-dire de l'ensemble des ARN et des protéines présent au bon moment et dans les quantités nécessaires, en réponse à des stimuli extra et intracellulaires. La morphogenèse est donc coordonnée à l'aide d'un vaste programme génétique, lui-même fortement contrôlé et régulé au cours du temps et de l'espace.

I.1. Régulation de l'expression génique

Un organisme pluricellulaire possède de nombreux programmes génétiques, hérités de ses ancêtres au cours de l'évolution. Au cours de la morphogenèse, les cellules doivent donc choisir entre ces divers programmes, c'est-à-dire réguler l'expression de leurs gènes. Cette régulation de l'expression génétique commence à l'échelle du génome entier, suivant une signalisation complexe et coordonnée entre les différentes cellules qui constituent cet organisme et est fortement contrôlée à l'aide de signaux extra- et intercellulaires qui affectent de façon positive ou négative le niveau d'expression de leurs gènes-cibles. Des programmes spécifiques vont donc se mettre en place en fonction des besoins de la cellule, de son état ainsi que l'ensemble des signaux reçus. Une grande partie cette régulation génétique se situe au niveau de l'étape de transcription des gènes.

I.1.a Transcription de gènes et régulation de l'expression génique

La transcription d'un gène, autrement dit la production d'un ARN messager (ARNm) à partir d'une séquence d'ADN, est réalisée par une machinerie d'une grande complexité et très dynamique, comprenant notamment une ARN-polymérase ADNdépendante, de nombreux facteurs associés ainsi que des sous-complexes moléculaires. L'assemblage de cette machinerie est elle-même très finement régulée et représente donc le premier niveau de régulation transcriptionnelle.



Figure 1 : Structure d'un gène eucaryote. Anne Keriel (2002) Implications de la kinase de TFIIH dans les mécanismes de régulation de la transcription

On retrouve chez tous les Eucaryotes, une structure similaire des gènes composée d'un promoteur comprenant le promoteur minimal (site de liaison des corégulateurs transcriptionnels), les séquences régulatrices proximales (site de liaison des facteurs de transcriptions) ; d'un site d'initiation de la transcription ; du gène avec ses séquences codantes (exons) et ses séquences non codantes (introns) ; d'un site de terminaison de la transcription et enfin de séquences de régulations situées loin du gènes et appelées séquences régulatrices distales.

Le niveau prédominant de régulation de la transcription, emploie des protéines nommées facteurs de transcription qui reconnaissent des séquences de régulation spécifiques de chaque gène. L'ensemble des facteurs de transcription, de leurs séquences de liaison, de leurs cofacteurs et de leur capacité à répondre aux signaux forme le régulon génomique d'un organisme dont l'activité dépend en partie de la structure tridimensionnelle de l'ADN (Benecke et al. 2003).

I.1.b Structure du génome et régulation de la transcription des gènes

Chez les eucaryotes, l'ADN est associé à de nombreuses protéines et molécules qui constituent, tous ensemble la chromatine. La chromatine est constituée d'unités de base appelées nucléosomes qui représentent le premier niveau de compaction de l'ADN. Les nucléosomes sont donc un mécanisme de régulation de la transcription, de la réplication ou de la réparation de l'ADN. En effet, la compaction de ces nucléosomes contrôle l'accessibilité à l'ADN double-brin par les enzymes et facteurs de transcription. Un nucléosome est composé d'un cœur protéique constitué de huit protéines histones (deux exemplaires de chacune des histones H2A, H2B, H3 et H4) autour duquel s'enroule environ 146 paires de bases d'ADN double-brin. Un nucléosome a une masse de 205 kilo Dalton (kDa), histones et ADN contribuant chacun à la moitié de cette masse (Harp et al. 2000), Figure 2, p16). Les cœurs des nucléosomes étant espacés les uns des autres par des segments d'ADN, les segments de liaison, les 146 paires de bases enroulées autour de l'octamère d'histones sont six fois plus compactes qu'un fragment d'ADN nu de même longueur. Les nucléosomes représentent donc bien le premier niveau de compaction de l'ADN (

Figure 4, p17).



Figure 2 : La structure d'un nucléosome (Wheeler R, 2005). De gauche à droite : hétérodimère H3-H4 et hétérodimère H2A-H2B. 2 exemplaires de ces hétérodimères forment un octamère d'histones. En haut à droite : Octamère d'histone entouré d'environ 146 paires de bases d'ADN double-brin.

Le deuxième niveau de compaction de l'ADN est le chromatosome ((Segers et al. 1991), (Simpson et al. 1978)). Le chromatosome est formé par l'association d'une histone de liaison (il s'agit le plus souvent d'une histone H1) à une particule de cœur voisine. L'histone H1 lie l'ADN au niveau où celui-ci rentre et sort de la particule de cœur (

Figure 3, p17). H1 induit le rapprochement des ADN de liaison entrant et sortant sur une trentaine de paires de bases, limitant leurs mouvements et scellant ainsi le complexe nucléo-protéique. Le chromatosome modifie les propriétés du nucléosome ainsi que la capacité de la chromatine à former des structures d'ordre supérieur, comme les chromosomes condensés au moment de la mitose. En interphase, la chromatine est à l'état déconsensé, l'ADN formant des nucléofilaments de chromatine. Cette structure est composée de l'enchainement des nucléosomes en une fibre de chromatine de 11 nm où les particules de cœur y apparaissent séparées par des segments d'ADN de liaison, de longueur variable en fonction des espèces et des tissus ; cette structure est aussi appelée « collier de perles » (

Figure 3, Figure 4, p18).



Figure 3 : Schéma d'un chromatosome

(<u>http://AtlasGeneticsOncology.org/Educ/ChromatinEducEng.html</u>). A gauche, se trouve la représentation schématique de l'unité de base de la chromatine, le nucléosome, liés entre eux par une histone de liaison. A droite est représenté l'association de deux nucléosomes formant un chromatosome

Les nucléosomes peuvent par ailleurs être déplacés le long du brin d'ADN sous l'action de complexes de remodelage à activité ATPasique, comme par exemple la famille SWItch/Sucrose NonFermentable (SWI/SNF) chromatin remodeler.

La dynamique, globale et spécifique de certaines parties de la chromatine, représente une autre forme de régulation de l'expression génique, finement contrôlée. Cette régulation se fait via des modifications de ses constituants protéiques, les histones. Les modifications les plus importantes des histones sont l'acétylation ou la déacétylation ainsi la méthylation. On parle alors de code épigénomique, code sur lequel je reviendrais plus tard. Ces modifications au niveau du chromosome (Figure 3, p17), de la séquence des nucléosomes, des nucléosomes individuels et des histones (Figure 5, p18) ne sont pas complètement élucidées.



Figure 4 : Formes chromatiniennes du génome (Smet Nocca C., 2006)

À gauche sont représentées les formes majeures de la chromatine à différentes échelles, de l'ADN nu au chromosome condensé, via les fibres de 9 nm, 30 nm et 300 nm. Et à droite est représenté une visualisation schématique des différentes formes chromatiniennes du génome.

Les différentes modifications des histones sont effectuées par les corégulateurs transcriptionnels, appelés coactivateurs ou corépresseurs, qui agissent sur l'activité d'un facteur de transcription ou d'un ensemble de facteurs donné. Les corégulateurs n'interagissent pas spécifiquement avec l'ADN mais ils présentent toutefois une spécificité vis-à-vis des gènes dont ils corégulent l'expression. Les corégulateurs sont donc le lien entre les régulateurs spécifiques aux différents gènes et la signalisation cellulaire. Ils participent ainsi à la plasticité génomique ((Benecke et al. 2003), (Benecke et al. 2006)).

De nombreuses expériences suggèrent que la liaison des facteurs de transcription sur l'ADN n'est pas nécessairement la première étape de la régulation transcriptionnelle. Elles proposent un autre modèle pour expliquer comment la machinerie transcriptionnelle lit, interprète et exécute l'information concernant la régulation de l'expression des gènes. Ce modèle introduit le concept d'un code de chromatinien, avant pour support les modifications des histones et permettant aussi de réguler structurellement et temporellement la transcription génique ((Benecke et al. 2006), (Lavelle et al. 2006)). Selon la théorie actuelle, les corégulateurs, qui présentent chacun une affinité particulière pour une queue d'histone (Bartlett et al. 2006), vont ainsi reconnaître spécifiquement certaines «séquences chromatiniennes», ce qui induira l'augmentation de leur concentration au niveau de ces séquences entrainant ainsi un changement de la conformation locale d'ouverture et de fermeture alternée de la chromatine. Cet effet a été observé, par exemple, avec le complexe de remodelage ATP-dépendant SWI-SNF dont je vous parlerai plus tard. Les facteurs de transcription peuvent alors accéder jusqu'à l'ADN au niveau des sites où la chromatine est «pré-ouverte », et la stabilisent dans un état «ouvert », permettant ainsi l'engagement de la machinerie transcriptionnelle.

Les mécanismes épigénétiques qui contrôlent la structure et la dynamique de la chromatine sont donc aussi à la base de la transmission de l'information génétique.

I.2. Mécanismes épigénétiques et différenciation cellulaire

L'épigénétique concerne l'ensemble des modifications héréditaires pouvant être réversibles qui conduisent à la modulation de la fonction d'un gène et ne pouvant pas être expliquées par des modifications de la séquence nucléotidique du gène. D'un point de vue biologique, l'épigénétique se traduira par des variations phénotypiques, qui se perpétuent le long des lignées cellulaires mais ne proviennent pas de modifications du génotype de ces cellules. Les modifications épigénétiques participent donc à la régulation de l'expression des gènes dans le temps et dans l'espace. Elles incluent la méthylation des bases nucléotidiques de l'ADN, les modifications post-traductionnelles des histones (méthylation, acétylation par exemple), la position des nucléosomes le long de la séquence nucléotidique qui modifie la structure locale de l'ADN et enfin la structure globale de l'ADN comme par exemple son organisation en domaines et leur localisation dans le noyau.

L'acétylation et la déacétylation des histones ont été les phénomènes les plus étudiés jusqu'à présent mais il semble évident que la méthylation des histones ainsi que celle des bases cytosines de l'ADN est une autre voie de régulation transcriptionnelle importante ((Benecke et al. 2003), (Benecke et al. 2006)) (Figure 5, p21).

Je me suis ainsi particulièrement consacrée, ainsi que l'équipe de recherche dont je fais partie, au code épigénétique induit par la méthylation des bases cytosines de l'ADN et à son impact sur la régulation de la transcription des gènes.



Figure 5 : Quelques exemples de code épigénétique (Léger H).

En rouge sont représentées les protéines histones H2A, H2B, H3 et H4 de l'octamère d'histones, en jaune les protéines histones de liaison ou histone H1 et en bleu la molécules d'ADN. Sont représentés l'acétylation, la méthylation et la phosphorylation des protéines histones ainsi que la méthylation des bases cytosines de l'ADN ce qui met en avant les deux types de modifications épigénétiques : sur les protéines histones et sur les bases de l'ADN.

I.2.a. La méthylation de l'ADN : marqueur épigénétique de la répression transcriptionnelle

Toutes les cellules d'un organisme possèdent le même patrimoine génétique mais acquièrent au cours du développement des phénotypes différents dans le temps et dans l'espace ; c'est la différenciation. Ce processus se caractérise notamment par une méthylation différentielle des gènes selon leur fonction dans un tissu ou un organe donné.

La méthylation permet d'inhiber certains éléments régulateurs tels que des promoteurs, des amplificateurs, des isolateurs et/ou des inhibiteurs, et par conséquent, de contrôler l'expression des gènes. De plus, nous savons que les profils de méthylation sont transmis d'une cellule-mère à ses cellules-filles au cours de la mitose, également aux gamètes au cours de la méiose, puis des gamètes à l'embryon. La transmission de ces

profils au travers des divisions cellulaires permet aux cellules d'une même lignée de conserver leur phénotype. Cette information, tout comme la séquence nucléotidique de l'ADN, est donc héréditaire.

Il existe une forte corrélation entre le degré de méthylation chromatinienne d'un gène et la répression de l'expression de ce même gène ; une forte méthylation coïncide toujours avec les régions transcriptionnellement inactives de la chromatine, «l'hétérochromatine », tandis qu'un faible degré de méthylation coïncide avec les régions transcriptionnellement actives appelées « euchromatine » (Figure 5, p21, Figure 6, p22).



Figure 6 : Composition du noyau des cellules eucaryotes. (Comings, 1967).

Le noyau se compose de deux types de chromatines, «l'hétérochromatine» (en gris), régions transcriptionnellement inactives car présentant une forte méthylation et «l'euchromatine» (en blanc), régions transcriptionnellement actives présentant un faible degré de méthylation. On constate que le centrosome est constitué majoritairement d'euchromatine.

La méthylation des cytosines en position 5 (appelées 5-méthylcytosines ou 5meC) se produit essentiellement au sein des îlots CpG. Les îlots CpG sont des régions concentrées au niveau ou à proximité des régions régulatrices de la transcription et qui présentent un fort enrichissement en cytosines immédiatement suivies d'une guanine (C.G

ou C point G ou CpG) est élevée. Dans les cellules somatiques, environ 70% des paires CpG sont méthylées via des enzymes, les DNA-méthyltransférases (Dnmt). La méthylation des îlots CpG stabilise le nucléosome et est associée au recrutement de larges complexes capables de se lier spécifiquement à ces sites de méthylation, les methyl-CpG-binding proteins (MeCP1/2). Il existe aussi des îlots non CpG (CpA, CpT) pouvant aussi être méthylés mais ils sont minoritaires et leur fonction est peu connue.

Les MeCP1/2 comprennent des sous-unités capables de se lier à l'ADN méthylé de part et d'autre du site de méthylation, les methyl-binding domains (MBD). La liaison de ces complexes au niveau du promoteur ou de l'amplificateur a un effet répressif sur la transcription car la grande taille de ces complexes macromoléculaires masque la séquence nucléotidique aux facteurs de transcription. De plus les complexes MeCP1/2 recrutent des enzymes impliquées dans les modifications post-traductionnelles des histones telles que les histone-méthyltransférases (HMT) (Figure 7, pFigure 7 : Les enzymes responsables de la régulation de la méthylation (Bannister et al. 2005) ou les histone-déacétylases (HDAC). Ces modifications vont renforcer l'effet répressif sur la transcription (Esteller et al. 2006).

Histone and residue	Methyltransferase	Demethylase/deiminase	Function
H3R2	CARM1 (Mm, Hs)	-	
		PADI4 (Hs)	
H3K4	vSET1 (Sc)		Activator/euchromatin
	SET7/Set9 (Hs)	-	Activator
	MLL (Hs)	-	Activator
	Ash1 (Dm)	-	Activator
	Smyd3 (Hs)	-	Activator
		LSD1 (Hs)	Repressor
H3R8	PRMT5	-	Repressor
	-	PADI4 (Hs)	
H3K9	SUV39h1/SUV39H1 (Mm, Hs)	-	DNA methylation/repressor/heterochromati
	SUV39h2 (Hs)	-	DNA methylation/heterochromatin
	Clr4 (Sp)	-	Repressor/heterochromatin
	Dim5 (Nc)	-	DNA methylation
	Kryptonite (At)	-	DNA methylation
	G9a (Mm, Hs)	-	Imprinting/repressor
	Eu-HMTase1 (Hs)	-	Repressor
	ESET/SETDB1 (Mm, Hs)	-	Repressor/DNA methylation
	E(z)/EZH2 (Dm, Hs)	-	Repressor
	Ash1 (Dm)	-	Activator
	-	LSD1 (Hs)	Activator
H3R17	CARM1 (Mm, Hs)	-	Activator
		PADI4 (Hs)	
H3R26	CARM1 (Mm, Hs)	-	
		PADI4 (Hs)	
H3K27	E(z)/EZH2 (Dm, Hs)		Repressor
	Ezh2 (Mm)	-	X-chromosome inactivation/heterochromatin
H3K36	Set2 (Sc)	-	Activator
	NSD1 (Mm)	-	
H3K79	Dot1/DOT1L (Sc. Hs)	-	Repressor/DNA damage
H4R3	PRMT1	-	Activator
	-	PADI4 (Hs)	
H4K20	SET9 (Sp)		DNA damage
	Pr-SET7/Set8 (Hs. Dm)	-	Repressor
	SUV4-20 (Hs)	-	Heterochromatin
	Ash1 (Dm)	-	Activator
	NSD1 (Mm)	-	
H1K26	EZH2 (Hs)	-	

Figure 7 : Les enzymes responsables de la régulation de la méthylation (Bannister et al. 2005), leurs fonctions et leurs substrats.

La répression de la transcription à long terme, ou gene silencing, implique deux mécanismes à la fois : d'une part la modification de l'ADN et, de l'autre, des modifications spécifiques des protéines histones. En effet les régions réprimées de l'ADN présentent un taux de méthylation très important tandis que l'on observe une hypoacétylation des queues N-terminales d'histones et une méthylation de résidus lysine du corps des histones. Notre équipe est particulièrement intéressée par les mécanismes de méthylation car cette modification entraine un état stable et fermé du nucléosome, donc de l'ADN, contrairement à l'acétylation qui, elle, entraine une déstabilisation et un déroulement de l'ADN le rendant ainsi plus accessible à cet endroit. La méthylation de l'ADN est donc un facteur prépondérant dans les processus de différenciation, de développement et de morphogenèse (Esteller et al. 2006).

I.2.b Méthylation et empreinte parentale

Chez les organismes sexués, les gènes sont en général exprimés à partir des deux allèles parentaux, l'allèle paternel venant du matériel génétique du spermatozoïde et l'allèle maternel venant de l'ovule ; on parle alors d'expression bi-allélique. Il existe cependant certains gènes essentiels pour le développement, qui sont exprimés de manière monoallélique c'est-à-dire qu'un seul des deux allèles parentaux s'est exprimé ; cet allèle exprimé peut provenir soit du génome paternel, soit du génome maternel. Ces gènes sont soumis à un phénomène d'empreinte parentale ou genomic imprinting, dû à une information épigénétique différente. Le choix de l'allèle exprimé se fera essentiellement selon le niveau du profil de méthylation de l'ADN de chacun des chromosomes. La méthylation différentielle de ces gènes provient d'une différence de profil de méthylation des gamètes.

De façon intéressante, dans les gamètes, l'empreinte parentale est d'abord effacée par une déméthylation massive de l'ADN puis de nouvelles empreintes épigénétiques sont reconstruites au cours de la gamétogenèse en fonction du sexe de l'embryon (Esteller et al. 2006). Ce mécanisme permet aux gènes des chromosomes sexuels d'avoir une même expression monoallélique chez un individu XX que chez un individu XY. La méthylation joue aussi un rôle au cours du développement embryonnaire. En effet les premières phases de développement se caractérisent par une déméthylation complète du génome à l'exception des gènes soumis à l'empreinte parentale qui semblent protégés. Cette déméthylation a pour but d'effacer la régulation induite par le profil de méthylation des gènes parentaux. Au stade blastocyste, les gènes de l'embryon subissent une reméthylation massive qui conduit à l'établissement d'une nouvelle identité épigénétique propre à chaque individu. Cette reméthylation embryonnaire explique, en partie, pourquoi des vrais jumeaux ne sont pas des clones parfaits l'un de l'autre.

De plus, les gènes soumis à l'empreinte parentale apportent une information supplémentaire au niveau du génome et jouent un rôle primordial dans le développement. En effet, ces gènes modulent l'expression du génome dans l'espace (type cellulaire) et dans le temps (Esteller et al. 2006) de façon héréditaire et néanmoins réversible.

L'exemple le plus frappant d'empreinte parentale via la méthylation est l'inactivation du chromosome X chez les individus femelles (XX).

En effet le chromosome X, contrairement au chromosome Y, contient de nombreux gènes essentiels pour le fonctionnement des cellules. Les individus femelles qui en possèdent deux exemplaires, contrairement aux individus mâles, devraient donc présenter une expression au moins deux fois plus importante des produits de ces gènes que les individus mâles. Or ce n'est pas le cas. Il existe donc un phénomène de compensation du dosage de ces gènes chez les mammifères femelles. Cette compensation a lieu via l'inactivation, au hasard, de l'un des chromosomes X de chaque cellule de la blastula, après son implantation dans la paroi utérine.

Les chromosomes X inactivés chez les individus femelles sont constitués d'histones H3 fortement méthylées et d'histones H4 déacétylées au niveau de leurs gènes. Ces modifications des histones contribuent probablement à la formation de l'hétérochromatine précédant l'arrêt de la transcription (Wutz et al 2011). Cette hétérochromatine est localisée à la périphérie du noyau, et est appelée «corpuscule de Barr ». Par ailleurs, cette inactivation aléatoire de l'un des chromosomes X de chaque cellule de la blastula se transmet à l'ensemble de sa lignée cellulaire, induisant ainsi la formation d'un individu mosaïque dont toutes les cellules n'auront pas le même X inactivé (Figure 8, p27).



Figure 8 : Exemple d'inactivation du chromosome X (wikipedia).

A) la robe « écaille de tortue » et B) la robe « calico » chez les chats femelles. Chez ces animaux, le locus du gène « orange », situé sur le chromosome X, présente deux formes d'allèles ; l'allèle orange « O » qui est la forme dominante (XO) produit une fourrure rousse, tandis que l'allèle non-orange « O » est récessive (Xo). Le gène « noir » est lui sous l'influence du gène « orange ». En effet l'allèle « O » masque l'effet du gène noir, alors que l'allèle « o » ne masque pas le gène « noir ». Pour être écaille de tortue, une chatte doit exprimer simultanément les deux allèles, « O » et « o ». La robe tricolore (ou "calico") est une robe écaille de tortue à laquelle s'ajoute des taches blanches. Le gène « S », porté par un chromosome non sexuel, détermine l'apparition de ces tâches blanches.

I.2.c. Organisation structurale et fonctionnelle de la chromatine.

Les différents gènes de notre organisme sont exprimés à différents niveaux, dans des cellules spécifiques (régulation spatiale) et à un moment précis du développement (régulation temporelle). De nombreuses études ont mis en évidence l'existence de différents domaines chromatiniens de part et d'autre d'un locus (localisation précise d'un gène), domaines délimités par divers éléments génétiques ayant une fonction très spécifique. La régulation spatio-temporelle de l'expression des gènes semble donc contrôlée via l'organisation de la chromatine en différents domaines fonctionnels.

Ainsi en interphase, stade du cycle cellulaire où les chromosomes sont le plus décondensés, la chromatine est constituée de nucléosomes formant des fibres enroulées en boucles contenant entre 10 et 150 kb. Ces boucles auraient pour fonction de délimiter l'ADN en domaine, comprenant un groupe de gènes régulés par un même mécanisme. On constate cependant que des cellules en interphase contiennent 10% de leur génome sous

forme d'hétérochromatine, c'est-à-dire sous la forme condensée, située généralement à la périphérie du noyau ((Bartlett et al. 2006), (Gilbert et al. 2006)).

Certaines régions du génome sont ainsi constituées d'hétérochromatine facultative, c'est-à-dire de régions inactivées spécifiquement au cours de certains stades du développement ou de la différenciation ; le corpuscule de Barr (la forme inactive du chromosome X chez la femelle, comme présenté précédemment) est un bon exemple d'hétérochromatine facultative. D'autres régions de l'ADN se trouvent préférentiellement sous forme d'hétérochromatine et sont appelées hétérochromatine constitutive. Ce sont des régions qui semblent dépourvues de gènes codant des protéines et qui sont toujours transcriptionnellement silencieuses. Par exemple, les séquences fortement répétées ou ADN satellite, telles que les centromères et les télomères sont des exemples d'hétérochromatine constitutive.

L'organisation des chromosomes en domaines fonctionnels est requise pour divers processus biologiques. La localisation d'un gène dans le territoire chromosomique et dans le noyau modulera son accessibilité vis-à-vis de la machinerie transcriptionnelle et aura donc des effets marqués sur son expression. Parallèlement, l'état d'expression d'un gène peut également orienter sa localisation chromosomique et nucléaire ((Bartlett et al. 2006), (Gilbert et al. 2006)). Ainsi un gène actif sera généralement localisé en périphérie du territoire chromosomique et au centre du noyau, associé à des compartiments à forte activité transcriptionnelle, tandis qu'un gène inactif sera plutôt enfoui à l'intérieur du chromosome, parfois à l'interface avec des régions d'hétérochromatine constitutive comme les centromères ou les télomères et en périphérie du noyau ((Bartlett et al. 2006), (Gilbert et al. 2006)).

Je vais vous présenter certaines régions ayant des fonctions bien spécifiques.

a) Les locus control regions (LCR) et les amplificateurs sont des domaines ciblés par les facteurs de transcription ; ils interviennent dans la mise en place de la chromatine en une configuration spatiale appropriée permettant la transcription ; cette configuration est nommée active chromatin hub. Ils interviennent aussi dans le ciblage de la chromatine vers des sites nucléaires spécifiques, pro-transcription. Ils sont donc directement impliqués dans l'augmentation de l'expression d'un gène.

b) Les scaffold-matrix attachment regions (S/MAR) sont des domaines situés en bordure des différents domaines chromatiniens. Ils sont constitués de régions d'ADN non apparié qui coïncident avec les régions riches en AT, les origines de réplication, les sites de reconnaissance de la topoisomérase II et les 3'-UTR. Les S/MAR sont des points d'ancrage à la matrice nucléaire des différents facteurs agissants sur le génome. Par ailleurs, les S/MAR régulent l'activité transcriptionnelle en stabilisant la chromatine dans un état transcriptionnellement actif, en réduisant les contraintes mécaniques sur l'ADN crées au cours de la transcription et en localisant les domaines chromatiniens près de centres de transcription ce qui facilite la transcription des gènes.

c) Les isolateurs sont des éléments géniques qui bornent les frontières entre des domaines chromatiniens existant dans des états de condensation non identiques. Les isolateurs peuvent influencer la structure des domaines chromatiniens via le recrutement de facteurs de transcription pour contrôler l'expression des gènes.

Toutes ces régions jouent donc un rôle majeur dans le maintien de l'intégrité du génome et dans la transmission de l'information génétique. Je vais maintenant vous illustrer le rôle des télomères.

Les télomères sont des régions hautement répétitives (plusieurs centaines de fois sur chaque chromosomes humains), non codantes, d'ADN identiques chez tous les vertébrés et situées à l'extrémité des chromosomes. À chaque duplication d'un chromosome eucaryote, l'ADN polymérase s'avère incapable de copier les derniers nucléotides : la présence de télomères sans information génétique, à l'extrémité des chromosomes, préviendrait ainsi la perte de données (Zhou et al. 2012). De plus, les télomères protègent l'ADN des nucléases, évitent que le chromosome ne s'effiloche et que son extrémité ne soit considérée comme une rupture du double brin d'ADN ce qui pourrait conduire à la fusion des extrémités chromosomiques. Par ailleurs, les télomères favorisent l'interaction des chromosomes avec l'enveloppe nucléaire ((Bartlett et al. 2006), (Gilbert et al. 2006)). Malheureusement les télomères raccourcissent avec l'âge, l'inflammation et le stress. La plupart des cancers produisent de la télomérase, mais souvent à un stade tardif, la cancérogénèse débutant par une érosion importante des télomères (Zhou et al. 2012).

En conclusion, l'organisation des chromosomes en domaines fonctionnels est un niveau de la régulation dynamique de l'expression génique car l'état d'expression d'un gène peut également orienter sa localisation chromosomique et nucléaire (Benecke et al. 2003). Cette localisation spatiale change donc au cours du temps et de la différenciation cellulaire.

I.3. Couplage entre régulation épigénétique, transcription des gènes et réparation de l'ADN

Les domaines chromatiniens permettent de regrouper dans un même environnement divers composants régulés de la même manière, permettant la mise en place de complexes fonctionnels spécifiques à un moment et à une condition particulière.

Divers agents extérieurs, tels que les rayonnements ultraviolets ou certains agents chimiques, entrainent des lésions à l'ADN, dommages qui induisent des mécanismes de réparation. Par exemple, les lésions causées par les rayonnements ultraviolets induisent un mécanisme de réparation de l'ADN par excision de nucléotides ou nucleotide excision repair (NER), au cœur duquel on retrouve les protéines Xeroderma Pigmentosum (XP). Un défaut de ces protéines entraine des maladies génétiques présentant une augmentation du risque de cancer. La régulation de la dynamique des domaines chromatiniens est soumise à une régulation épigénétique (paragraphe I. 2. C, p27) qui sert à orchestrer tous les processus nucléaires liés à l'ADN tels que la régulation transcriptionnelle, la réparation de l'ADN, la réplication du génome et la condensation supérieure de la chromatine ((Benecke et al. 2003), (Smet-Nocca et al. 2006)).

Une perturbation locale ou globale de cette régulation épigénétique entraîne une dérégulation de l'expression des gènes, ce qui mène généralement à une modification du

fonctionnement cellulaire et l'apparition de désordres génétiques ou de cellules tumorales. En effet, on constate que dans presque tous les cancers on retrouve une dérégulation du code épigénétique (Lund et al. 2004).

I.3.a Couplage entre épigénétique et oncogenèse

Qu'est-ce que l'oncogenèse ?

Le terme oncogenèse défini l'ensemble des acteurs et des mécanismes entrainant la transformation de cellules saines en cellules tumorales (Figure 9, p31) qui peuvent proliférer de façon anarchique, via l'activation d'un certain nombre de gènes dits pro-oncogéniques. Ces pro-oncogènes sont inactifs en temps normal dans les cellules saines mais peuvent être réactivés par une modification de leur épigénome, c'est-à-dire de l'ensemble des informations portées par leur séquence d'ADN, par leur code épigénétique et par leur dynamique chromatinienne ((Benecke et al. 2003), (Kubicek et al. 2004), (Smet-Nocca et al. 2006)).



Figure 9 : Les six mécanismes fondamentaux induisant une transformation oncogénique ((Hanahan and Weinberg 2000).

Certaines mutations peuvent, par exemple, entrainer l'extinction ou la diminution de l'expression d'un gène ou au contraire induire l'augmentation de son expression génique ; l'acétylation de la chromatine est utilisée pour décondenser la chromatine et donc activer les gènes situés dans cet environnement tandis que la méthylation de l'ADN est, elle, utilisée pour « éteindre » des gènes (paragraphe I.2.a, p21).

Notre équipe s'est plus particulièrement intéressée à l'étude des profils de méthylation, à leur établissement et à leur maintien en conditions correctes permettant le développement et le fonctionnement normal des cellules et des organismes chez les mammifères.

Les cellules cancéreuses présentent de fortes modifications des profils de méthylation de l'ADN, avec généralement une forte hypo-méthylation de l'ensemble du génome mais aussi la présence d'une hyper-méthylation au niveau de régions très spécifiques de l'ADN ((Benecke et al. 2003), (Lund et al. 2004)).

En effet, on constate cliniquement que les changements de méthylation des cellules impliquées dans un grand nombre de cancers sont des événements précoces dans la mise en place de l'oncogenèse. Par ailleurs, les études cliniques ont mis en évidence que la perte des profils normaux de méthylation est corrélée avec la sévérité de la maladie et le potentiel métastatique. Quelles en sont les raisons et comment ont lieu ces modifications ?

L'hypométhylation globale touchant les régions codantes, les introns et les séquences répétitives cela affecte la stabilité fonctionnelle des chromosomes et réactive les éléments transposables accumulés dans notre génome au cours de l'évolution ce qui entraine la mise en place d'un phénotype cancéreux. Les îlots CpG, situés principalement dans les régions régulatrices des gènes sont quant à eux fortement méthylés dans les cellules cancéreuses ; les gènes situés sous le contrôle de ces régions régulatrices sont alors éteints. Les gènes suppresseurs de tumeurs sont généralement fortement méthylés dans les cancers, perdant ainsi tout contrôle sur la survie et la prolifération des cellules cancéreuses (Fukushige et al. 2009).

Le niveau de méthylation de l'ADN est régulé par un type d'enzyme, les DNAméthyltransférases (Dnmt) ((Mund et al. 2004), (Walsh et al. 2006)) (Figure 10, p33). Ces enzymes permettent l'établissement et le maintien des profils de méthylation ; par ailleurs, elles recrutent les histone-déacétylases (HDAC) qui permet d'acétyler les promoteurs des gènes et de renforcer donc le silencing des gènes concernés (Zhou et al. 2008). On observe donc, dans les cellules cancéreuses, des profils très spécifiques pour la méthylation et l'acétylation des histones et de l'ADN (Esteller et al. 2006).



Figure 10 : Structures et expression (analyse en western blot) des différentes enzymes Dnmts (Mund et al. 2004) des cellules eucaryotes. Il existe trois grands types d'enzymes Dnmt, la famille des Dnmt1 qui possède un domaine catalytique proche de l'enzyme Dnmt2 et enfin les enzymes de la famille Dnmt3.

La détermination des profils de méthylation et/ou d'acétylation des cellules en condition normale comparée aux profils de méthylation et/ou d'acétylation des cellules en

condition tumorale permettrait d'avoir des marqueurs de détection des cellules cancéreuses et même de reconnaitre un état précancéreux des cellules et donc de soigner très précocement les patients ; cela augmenterait considérable les chances de succès des traitements (Morikawa et al. 2011). Les connaissances acquises sur les profils épigénétiques des cellules en condition normale ou en condition tumorale et sur les mécanismes de régulation des épigénotypes, permettraient de mettre en place de nouveaux agents thérapeutiques qui cibleraient l'épigénome des cellules cancéreuses les retransformant en cellules normales malades et donc accessibles à l'apoptose. Ces progrès thérapeutiques pourraient aussi toucher de nombreuses maladies liées entre autres à des désordres épigénétiques telles que les maladies auto-immunes ou certains cancers (Lund et al. 2004). Ces stratégies nécessitent une connaissance approfondies des mécanismes épigénétiques, de leurs régulations. Il est important aussi d'identifier précisément un groupe de gènes-cibles impliqués directement dans la pathologie (gènes-cibles primaires) et dérégulés en contexte pathologique.

I.3.b Couplage entre la transcription et la réparation de l'ADN

De nombreuses études précédentes ont montré qu'il existe un lien entre la réparation par excision de nucléotides ou NER (Nucleotide Excision Repair) et la transcription via le facteur de transcription TFIIH (décrit à l'origine comme un élément du complexe d'initiation de la transcription associé à l'ARN-polymérase II).

Le NER est un des mécanismes cellulaires qui permet de limiter les dommages créés par l'environnement et les erreurs de réplications du matériel génétique. Ce système permet notamment de réparer des dommages importants ou entrainants des déformations majeures de l'ADN. C'est aussi un des mécanismes qui lutte contre le vieillissement de l'organisme, limite les mutations délétères et le risque d'apparition de tumeurs et de cancers. Le NER entraine l'excision sur le brin endommagé d'un segment de plusieurs nucléotides encadrant la lésion. Après l'excision de la zone d'ADN endommagée, une ADN polymérase synthétise un nouveau brin d'ADN en utilisant le brin intact comme matrice et une ADN ligase soude les deux parties du squelette phosphodiester de l'ADN. Le NER est composé des deux mécanismes de réparation de l'ADN : le "global génome repair » (GGR) et le « transcription-coupled repair » (TCR) (Figure 11, p35). Ce dernier est spécialisé dans la réparation des lésions présentes sur le brin transcrit des gènes

transcriptionnellement actifs, où l'ADN polymérase II peut rencontrer une lésion ((Hanawalt et al. 2008), (Svejstrup et al. 2002)) tandis que le GGR est accompli par un complexe protéique DNA binding qui présente une très forte affinité pour les sites endommagés de l'ADN. Les protéines de la famille XP associées au facteur de transcription TFIIH sont impliquées dans ce processus. (Naegeli et al. 2011).

Une déficience dans le mécanisme NER chez l'homme est liée à des maladies sévères, comme Xeroderma Pigmentosum (XP) ou le syndrome de Cockayne (CS), résultant de la multiplication de lésions de l'ADN, affectant de nombreux gènes ou régions régulatrices. Ces maladies génétiques sont associées à une sensibilité accrue à la lumière solaire, à une augmentation du risque de développer des cancers de la peau et à des dysfonctionnements neurologiques (Riedl et al. 2003) (Figure 11, p35).



Figure 11 : Schéma des deux voies du nucleotide excision repair (Schumacher et al. 2009). Le NER est composé de la voie de réparation par excision des nucléotides mésappariés ou "global génome repair » (GGR) (A gauche) dont les défauts sont associés à la maladie Xeroderma pigmentusum (XP) et de la voie de réparation couplé à la transcription ou « transcription-coupled repair » (TCR) dont les défauts sont associés à la maladie Cockayne syndrom (CS).
Le Transcription Factor II H (TFIIH) est un facteur qui participe à la transcription des gènes. Il a été tout d'abord caractérisé comme étant un élément du complexe d'initiation de la transcription associé à l'ARN-polymérase II ; sans le système NER, la transcription est fortement perturbée. Le fonctionnement de TFIIH était l'exemple le plus connu de couplage entre transcription et réparation d'ADN ((Drapkin et al. 1994), (Hanawalt et al. 2008)).

Plus récemment, un couplage entre des processus similaires a été décrit au niveau de sites de régulation épigénétique, les cytosines méthylées et impliquant un couplage entre le mécanisme de réparation par excision de base ou base excision repair (BER) et la transcription via le complexe CBP/TDG, complexe que je vous présenterais en détail au paragraphe II.2.c (p54).

La réparation par excision de base (Base excision Repair ou BER) est un mécanisme de réparation des dommages situés au niveau individuel des bases de l'ADN. La reconnaissance de tels dommages causés à l'ADN est difficile à réaliser car les modifications des bases ne causent, le plus souvent, qu'un minimum de distorsions de la structure et de la thermodynamique de la double hélice d'ADN. Pour détecter ces lésions peu visibles mais néanmoins nocives, les cellules déploient un grand nombre de DNA glycosylases, enzymes qui reconnaissent pour chacune d'entre elles quelques types de bases erronées via une poche d'interaction spécifique ((Gallinari et al. 1996), (Hardeland et al. 2000)). Ces enzymes réparent ce type de dommage en coupant le pont N-glycosylique, ce qui élimine la mauvaise base, puis en resynthétisant la base correctement appariée (Huang et al. 2006).

Pourquoi y a-t-il couplage en BER et transcription au niveau des cytosines méthylées ?

La cytosine est la base la plus instable de l'ADN puisque la déamination spontanée d'une cytosine conduit à la formation d'une uracile et par conséquent, à un mésappariement G : U (Varghese et al. 1971). De plus, il existe une famille d'enzymes, les cytidine déaminases dont font partie les enzymes AID/APOBEC (Activation-induced (cytidine) deaminase/apolipoprotein B mRNA-editing, enzyme-catalytic), qui déaminent les cytosines en uraciles au sein des cellules (Conticello et al. 2008). Ces mésappariements sont aisément reconnus au sein de l'ADN et sont réparés par des enzymes appelées Uracile-DNA Glycosylases (UDG) (Figure 12, p37) (Vertessy et al. 2009). Ces enzymes se retrouvent tout le long de l'arbre de l'évolution depuis les procaryotes jusqu'à l'Homme et même chez certains virus. Elles permettent d'éliminer les mutations en éliminant les bases uraciles des molécules d'ADN.



Figure 12 : Structure en ruban de l'enzyme Uracil-DNA glycosylase (Vertessy et al. 2009). Ces protéines sont présentes de façon conservée tout le long de l'arbre phylogénique depuis les virus à ADN jusqu'aux êtres humains.

Les cytosines méthylées (Cme) sont, elles, susceptibles de former, par déamination, des thymines et donc des mésappariements G : T, spécifiquement réparés par la Thymine-DNA Glycosylase (TDG) ((Wiebauer et al. 1989), (Neddermann et al. 1993), (Neddermann et al. 1996), (Gallinari et al. 1996), (Cortázar et al. 2007)).



Figure 13 : Formation des mésappariements G : T et G : U dans les cellules (Benecke et al. 2012). Les bases cytosines, méthylées (en bas) ou non (en haut) peuvent se transformées par déamination spontanée ou sous l'influence des enzymes AID/APOBEC en bases thymines (en bas) ou uraciles (en haut). Chapitre II _ Les composants du complexe Creb Binding Protein/Thymine DNA Glycosylase/Récepteur à l'Acide Rétinoïque (CBP/TDG/RAR)

Comme je vous l'ai présenté au chapitre précédent, l'épigénétisme est un des mécanismes les plus importants de la régulation de la transcription. Le contrôle de la méthylation des cytosines joue un rôle particulièrement important dans la régulation des gènes notamment au cours du développement embryonnaire. Malheureusement les cytosines sont aussi les bases les plus instables de l'ADN, elles peuvent en effet se transformer en uracil par simple hydrolyse ; la forme méthylée des cytosines (Cme) est quant à elle susceptible de former des thymines (

Figure 13, p38). De plus il existe dans l'organisme des enzymes qui induisent ces transformations chimiques, ce sont les enzymes de la famille des AID/APOBEC. Ces enzymes sont des "cytidine désaminases"; ces enzymes possèdent une fonction peu commune, celle d'insérer des mutations dans l'ADN ou l'ARN en déaminant respectivement les cytosines et les cytosines méthylées en uracil ou thymine (Conticello et al. 2008). AID/APOBECs sont donc des acteurs majeurs de l'oncogenèse. Il semble que la famille des AID/APOBECs possède d'autres fonctions que son rôle mutagénique. En effet les AID/APOBECs jouent un rôle dans la diversification des anticorps ainsi que dans l'immunité innée contre les rétrovirus (Kohli et al. 2010).

Ces mésappariements G : U/T, n'entraînent pas seulement une modification de la séquence nucléotidique de l'ADN mais aussi la perte de l'information épigénétique portée par les cytosines méthylées. En effet, si ces mésappariements G : U/T ne sont pas réparés avant la réplication, la moitié du nouveau matériel génétique sera un binôme G : C et l'autre moitié sera un binôme A : T. Il est donc nécessaire de réparer ces mésappariements (

Figure 14, p40).



Figure 14 : Schéma de la transition G/A au cours de la réplication (Léger H.).

En haut : Si on ne répare pas les mésappariement G : U/T avant la réplication, on obtiendra 50% de cellules présentant un appariement G : C et donc une perte du patron de méthylation et 50% de cellules présentant un appariement A : T et donc une mutation du gène et une perte du patron de méthylation. Ces erreurs peuvent induire dérégulation et/ou modification de l'expression des gènes. En bas : Représentation en haut, de gauche à droite, des bases cytosines et thymines et en bas de gauche à droite, des bases cytosines méthylées et uraciles.

II.1. La Thymine DNA Glycosylase (TDG)

Très tôt au cours de l'évolution, sont apparues des enzymes permettant de prévenir ce type de mutation en éliminant les bases uraciles apparues dans l'ADN. Ces enzymes sont appelées des Uracil-DNA glycosylases ou UDG (Lindahl et al. 1977) (

Figure 15, p41).



Figure 15 : Arbre phylogénique des protéines Mistmatch Uracil DNA Glycosylase (MUG) (Cortazar et al. 2007).

Les protéines MUG (en gris) sont ubiquitaires et de structure très conservée tout le long de l'arbre phylogénétique, depuis les Uracil-DNA glycosylases des virus à ADN jusqu'aux Thymine DNA Glycosylases (entourées en rouge) des Eucaryotes, protéines ayant acquis la capacité à réparer les thymines issues de la déamination des cytosines méthylées.

Ces enzymes DNA glycosylases scannent l'ADN à la recherche de mésappariements, qu'elles reconnaissent grâce à leur site actif qui possède une spécificité de substrats (Hardeland et al. 2000). Certaines de ces enzymes possèdent un site actif très discriminant qui ne peut reconnaitre qu'un petit nombre de lésions. Par exemple, l'enzyme Uracil DNA Glycosylase (UDG) utilise une méthode d'exclusion stérique ainsi que des interactions électrostatiques spécifiques pour reconnaitre les bases uraciles présentes dans l'ADN et ne permet pas à d'autres bases de s'amarrer dans son site actif (Eftedal et al. 1993). Cependant, de nombreuses DNA glycosylases peuvent reconnaître un large spectre de bases erronées et même certaines bases normales via leur site actif. Pour discriminer les bases qu'elles peuvent exciser, ces enzymes utilisent certaines propriétés des bases

endommagées. Par exemple, les bases alkylées forment un pont N-glycosidique moins stable et plus facilement rompu que les bases non alkylées (Hardeland et al. 2000).

Les DNA glycosylases existent de façon très conservée tout le long de l'arbre phylogénétique, depuis les virus à ADN en passant par les procaryotes pour arriver jusqu'aux eucaryotes chez lesquels les Thymine DNA Glycosylases ont acquis la capacité à réparer les thymines issues de la déamination des cytosines méthylées ((Hardeland et al. 2003), (Cortázar et al. 2007)) (

Figure 15, p41).

II.1.a Structure et fonctions de TDG

La Thymine-DNA Glycosylase (TDG) a été découverte comme étant une enzyme qui excise les thymines lors des mésappariements guanine: thymine (G: T) dans les cellules HeLa (Neddermann et al. 1993). TDG, qui appartient à la famille des UDG, a donc acquis au cours de l'évolution la capacité à reconnaitre les mésappariements G : T.



Figure 16 : Présentation de différents substrats de TDG (Bennett et al. 2006)

De nombreuses études ont montré que TDG reconnait divers substrats dans les cellules en plus des bases thymine mal appariées. TDG peut ainsi exciser les bases 3-N4-éthenocytosine, 5-methylcytosine (5-meC), pyrimidines, 5-fluorouracile (FU), 5-hydroxymethyluracile (HmU), 5-bromouracile ((Hardeland et al. 2000), (Hardeland et al. 2003), (Bennett et al. 2006)) (

Figure 16, p42,

Figure 17, p43) ainsi que divers autres substrats situés dans l'ADN double brin avec une nette préférence pour les bases hybridées avec des guanines par rapport à celles hybridées avec des adénines. TDG possède une affinité variable en fonction des substrats qu'elle rencontre. En effet TDG présentera un turnover plus lent si elle rencontre un mésappariement G : T ou G : U que si elle rencontre un mésappariement G : FU (Kunz et al. 2009). La spécificité de TDG dépend de la stabilité du pont couplant la base à son sucre, stabilité qui détermine l'activité maximale (kmax) de TDG sur la base à exciser. TDG excise donc les bases 5-fluorouraciles 78 fois plus vite que les bases uraciles, ou 5-chlorouraciles 572 fois plus vite que les bases thymines. Ces différences de turnover peuvent être aussi expliquées par le fait que le site actif de TDG augmente cette spécificité via son interaction plus ou moins grande avec l'ADN (Kunz et al. 2009).

	N140A ♦ N140D ♦ A145S
hsTDG 123	TLPDILTFNLDIVIIGINPGUMAAYKGHHYPGPGNH 158
dmThd1 778	TIPDHLCDNLDIVIVGINPGLFAAYKGHHYAGPGNH 813
spThpl 141 (GVPDYICENPYAIIVGLNPGITSSLKGHAFASPSNR 176
ecMug 1 M	MVEDILAPGLRVVFCGINPGLSSAGTGFPFAHPANR 36
hsTDG 159	FWKCLFMSGLSEVQLNHMDDHTLPG-KYGIGFTNMV 193
dmThdl 814	FWKCLYLAGLTQEQMSADEDHKLIKQGIGFTNMV 847
spThpl 177	FWKMLNKSKLLEGNAEFTYLNDKDLPAHGLGITNLC 212
ecMug 37	FWKVIYQAGFTDRQLKPQEAQHLLDYRCGVTKLV 70
hsTDG 194	ERTTPGSKDLSSKEFREGGRILVOKLOKYOPRIAVF 229
dmThd1 848	ARATKGSADLTRKEIKEGSRILLEKLQRFRPKVAVF 883
spThpl 213	ARPSSSGADLRKEEMQDGARILYEKVKRYRPQVGLF 248
ecMug 71 l	DRPTVQANEVSKQELHAGGRKLIEKIEDYQPQALA I 106
hsTDG 230	N-GKCIYEI-ESKEVEGVKVKNLEEGLOPHKIPDTE 263
dmThd1 884	N-GKLIFEV-FSGKKEFHFGRQPDRVDGTD 911
spThpl 249	ISGKGIWEEMYKMLTGKKLPKTFVFGWQPEKFGD 282
ecMug 107	L-GKQAYEQGFSQRGAQWGKQTLTIGSTQ 134
hsTDG 264	TLCYVMPSSSARCAQFPRAQDKVHYYIKLKDLRD 297
dmThdl 912	TFIWVMPSSSARCAQLPRAADKVPFYAALKKFRD 945
spThpl 283	ANVFVGISSSGRAAG-YSDEKKONLWNLFAEEV 314
ecMug 135	IWV LPNPSGLSRIVSLEKLVEAYRELDQALVVRGR 168
150000	AHPSPLSV

Figure 17 : Evolution du site catalytique de TDG (Hardeland et al. 2000)

Séquence en acides aminés du site catalytique de la protéine humaine de TDG (hsTDG, EMBL accession number U51166) ainsi que de ses homologues (D. melanogaster (dmThd1, EMBL accession number AJ277789), S. pombe (spThp1, EMBL accession number AJ277958), and E. coli (ecMug; Swissprot accession number P43342)). Les résidus identiques à ces protéines sont grisés, les acides aminés essentiels au fonctionnement du site actif de TDG sont encadrés et les acides aminés en italique représentent les motifs structuraux identiques à ceux du gène de l'uracil DNA glycosylase du virus herpes simplex (hsvUDG; Swissprot accession number P10186). Enfin, les astérix représentent certains sites mutés connus chez la protéine hTDG (N140A, N140D, M269H, or A145S).

TDG scanne l'ADN à la recherche de bases mal appariées qu'elle reconnaît grâce à la spécificité de son site actif pour un certain nombre de bases ; elle se fixe donc à son substrat et coupe le pont Nglycosidique reliant la base à son sucre créant ainsi un site abasique (appelé aussi apurinic/apyrimidinic (AP)). TDG, contrairement à d'autres DNA glycosylases telles que l'enzyme Single-strand selective monofunctional uracil DNA glycosylase (SMUG1), ne peut pas se détacher du site abasique simplement par compétition avec l'endonucléase Apurinic/apyrimidinic (AP) endonuclease (APEX ou APE1) pour la liaison à l'ADN. TDG nécessite donc un changement conformationnel due à son interaction covalente avec SUMO-1 pour diminuer son affinité au site abasique. Ce mécanisme permet à TDG de rester accroché à l'ADN le temps de recruter APEX et ensuite de se détacher. APEX est une enzyme qui coupe les liaisons reliant le sucre formant le site AP au reste du brin d'ADN ; elle va recruter ensuite les autres enzymes du BER telle que la polymérase β par exemple (

Figure 18, p44).



Figure 18 : Schéma du Base excision repair ou réparation par excision de bases (Huang et al. 2006). Dans les cellules, 1) TDG scanne sans cesse l'ADN à la recherche de mésappariements G : U/T. 2) Quand il reconnait un mésappariement, ici G : T, il se fixe à l'ADN, 3) coupe le lien glycosylique reliant la base à son sucre créant ainsi un site abasique. 4-5) TDG doit ensuite être modifier par la conjugaison avec la protéine SUMO pour 6) se détacher de l'ADN, libérant ainsi la place pour les autres enzymes de la voie de réparation de l'ADN.

Je vais maintenant expliciter plus en détail le mécanisme de réparation des bases thymine, uracile et 5-fluorouracile (5-FU) pour un exemple de bases non canoniques de TDG (Kunz et al. 2009). Le 5-FU est un analogue de l'uracile qui possède une fluorine à la place du C5 ; c'est une molécule utilisée dans le traitement de nombreux cancers dont notamment les cancers colorectaux. Le 5-FU est converti dans les cellules en divers métabolismes dont la fluorodeoxyuridine monophosphate ou triphosphate (FdUMP et FdUTP respectivement), des molécules impliquées dans le métabolisme de l'ADN et de l'ARN.

TDG reconnait les mésappariements A.5-FU, G.5-FU et AP.5-FU et peut hydrolyser le pont N-glycosylique plus facilement c'est-à-dire qu'elle a besoin de moins d'énergie pour casser cette liaison et présente une plus faible affinité pour le site AP ainsi formé ce qui lui permet d'avoir un turnover fortement augmenté. En contrepartie, TDG présente à ce moment-là une plus grande toxicité qu'en présence de G.U, G.T ou G. AP car TDG sous l'action de SUMO-1 se détache de l'ADN avec une relative facilité, sans avoir forcément recruté l'endonuclésase APEX à chaque fois. Il y a donc multiplication des sites abasiques dans l'ADN des cellules traitées qui ne peuvent ainsi pas passer la transition G2/M (phase de préparation à la division cellulaire/mitose) ou bien restent bloquées en phase S (phase de synthèse de l'ADN) induisant l'activation de la voie apoptotique. On ne retrouve pas cette toxicité du 5-FU en présence de la DNA glycosylase SMUG1 car celleci est détachée de l'ADN par compétition de liaison entre le site AP formé et l'endonucléase APEX dont la capacité à lier SMUG1 prend le pas sur la capacité du site AP à lier SMUG1 ; il y donc pas, dans ce système BER simple, de dysfonctionnement dans le passage du bâton. Le 5-FU est donc utilisé en traitement anti cancéreux car il modifie la balance d'incorporation des uraciles et des thymines dans l'ADN en faveur des uraciles, 5-U et de ses métabolites. Cette surreprésentation de ces dérivés de l'uracile entraine donc la fragmentation de l'ADN après excision de ces bases via les TDG et donc apoptose ou bien arrêt en phase S (Kunz et al. 2009).

II.1.b TDG et la régulation de la transcription

Des études récentes ont mises en évidence que TDG a d'autres fonctions que celle de DNA glycosylase. TDG possède donc une activité transcriptionnelle qui diffère de son activité de glycosylation (Cortázar et al. 2011). TDG peut aussi interagir avec différents facteurs de transcription dont ceux de la famille des récepteurs nucléaires comme les récepteurs aux œstrogènes (Chen et al. 2003) ou bien les récepteurs à l'acide rétinoïque RAR : RXR (Um et al. 1998), récepteurs que je vous présenterais au paragraphe II.3 (p55). TDG peut aussi interagir avec certains corégulateurs transcriptionnels tels que SRC-1 (Lucey et al. 2005) ou bien CBP/p300 (Tini et al. 2002) que je vous présenterais au paragraphe II.2 (p48). Lors de ma thèse, je me suis particulièrement intéressée à la formation du complexe entre TDG et CBP/p300 ainsi qu'au complexe CBP/TDG/RARα.

Des études précédentes, l'une publiée dans le journal Nature (Cortázar et al. 2011) et l'autre dans le journal Cell (Cortellino et al. 2011), montre que des souris Tdg knockout (Tdg KO) présentent un arrêt du développement au jour embryonnaire E11,5. Cet arrêt est due à des hémorragies internes abdominales (au niveau du foie), des œdèmes péricardiques, une hypoplasie des arcs branchiaux (Cortázar et al. 2011) (

Figure 19, p46), un retard dans le développement des membres, des vésicules télencéphaliques trop proéminentes, ainsi que des lésions diffuses hémorragiques (Cortellino et al. 2011). Plus tard lors de la gestation, les embryons Tdg KO commencent à se résorber jusqu'à disparaitre complètement. Du point de vue microscopique, les embryons Tdg KO présentent des défauts de développement cardiaque, des défauts de l'aorte dorsale et des artères branchiales. Les embryons présentent aussi un défaut généralisé de l'angiogenèse.



Figure 19 : Comparaison entre le développement d'embryons Tdg ⁻/₋ et Tdg ⁺/₊ (Cortázar et al. 2011). On constate que les embryons Tdg -/- présentent de graves défauts embryonnaires au jour embryonnaire E10.5. Au jour E12.5 les embryons Tdg ⁻/₋ sont en voie de résorption.

Des fibroblastes issus des embryons de ces souris (Mouse Embryonic Fibroblastes ou MEFs) Tdg KO (MEFs Tdg /) ont été exposés à divers agents induisant des dommages à l'ADN, normalement réparés par TDG telles les radiations ionisantes ou bien l'eau oxygénée (H₂O₂). Ces MEFs Tdg ⁷/₋ ne présentent pas de mort cellulaire plus importante que les MEFs TDG ⁺/₊. De plus, la comparaison entre les profils d'expression issus de MEFs Tdg $^{-}/_{+}$ et de MEF Tdg $^{+}/_{+}$ a mis en évidence environ 400 gènes régulés par TDG (P≤0.05, FC ≥1.5) comprenant un grand nombre de facteurs de transcription. La majorité de ces gènes est impliquée dans le développement et l'embryogenèse sous le contrôle épigénétique via le système polycomb. Le rôle de réparation de l'ADN de TDG ne semble donc pas être la cause de la mort des embryons et est donc mineur par rapport au rôle de TDG sur la régulation de l'expression des gènes (Cortázar et al. 2011). Or, on sait que les îlots CpG situés au niveau des promoteurs des gènes contrôles par polycomb, sont généralement non méthylés en condition saine et présentent une méthylation de novo aberrante dans des cellules cancéreuses. Les auteurs ont constaté que les îlots CpG de ces promoteurs sont non méthylés dans les MEFs Tdg 7/.. Certains de ces facteurs de transcription ont été retrouvés parmi les facteurs liés à TDG lors d'expériences de chromatin immunoprecipitation (ChIP). TDG semble donc avoir un rôle épigénétique au niveau de certains gènes spécifiques. Il pourrait être un moyen de protéger ces gènes, importants lors du développement embryonnaire des vertébrés, contre l'apparition de profils de méthylation aberrants entrainant le « silencing » de ces gènes. Cette hypothèse est confortée par le fait que les cellules MEFs Tdg ⁺/₊ présentent beaucoup moins de liaison du cofacteur CBP/p300 aux promoteurs des gènes étudiés ((Cortázar et al. 2011) (Cortellino et al. 2011)).

TDG semble donc un acteur de la régulation épigénétique de la transcription de par son activité et via son interaction avec CBP/p300 ((Tini et al. 2002), (Cortázar et al. 2011), (Cortellino et al. 2011)).

Je vais donc vous présenter ce partenaire de TDG qui forme le complexe CBP/TDG, l'un des aspects de mes travaux de thèse.

II.2. La Creb Binding Protein (CBP ou CREBBP)

En effet, il a été précédemment montré que TDG peut s'associer à l'histone acétyltransférase CBP (Tini et al. 2002), qui est par ailleurs un co-activateur pour de nombreux facteurs de transcription de la famille des récepteurs nucléaires, dont les récepteurs à l'acide rétinoïque (RAR : RXR) (Um et al. 1998) et les récepteurs aux œstrogènes (Chen et al. 2003). Il a été aussi montré que TDG intervient lui aussi comme corégulateur de ces facteurs de transcription.

II.2.a Structure et fonctions de CBP

Le gène codant pour la CBP est exprimé de façon ubiquitaire chez les vertébrés et code pour une protéine histone acétyltransférase ou HAT (Arany et al. 1994) (Figure 20, p48).



Figure 20 : Structure en ruban de la Creb Binding Protein (Liu et al. 2008)

Les HATs sont des enzymes qui acétylent les protéines histones ainsi que des protéines non histones, tel le coactivateur NCOA3 ou TDG, pour réguler leur activité. Elles agissent sur le remodelage et la décondensation de la chromatine, ce qui permet à l'ARN polymérase II d'accéder à ses sites de liaison et aux protéines de régulation de la transcription de se lier à l'ADN. CBP et son proche homologue p300, sont donc des coactivateurs de la transcription (Kwok et al. 1994). CBP/p300 ont tout d'abord été isolées comme des protéines pouvant lier la protéine cAMP-response element binding (CREB) (Chrivia et al. 1993).

Le gène Creb, situé sur le chromosome 2 possède un épissage alternatif de son ARNm et produit deux transcrits différents et conduit à la formation de deux isoformes CREB-A et CREB-B tous deux reconnu par la CREBBP ou CBP (Wydner et al. 1995). La protéine CREB est un facteur de transcription ubiquitaire de 45 kDa appartenant à la famille des facteurs de transcription à domaine bZIP, comme CREM ou ATF-1 ; ces facteurs contiennent une région basique riche en leucine nommé aussi « leucine zipper » qui permet la dimérisation de la protéine et une reconnaissance spécifique à l'ADN aux niveau de ses éléments de réponse appelés séquences CRE ou cAMP response element ((Chrivia et al. 1993), (Kwok et al. 1994)). Les séquences sont composées d'un palindrome de 8 paires de bases «TGACGTCA» et permettent de répondre aux cascades moléculaires impliquant l'AMPc et menant à l'activation de CREB via sa phosphorylation. Un signal extracellulaire, tel une hormone, un neurotransmetteur arrive à la surface de la cellule, se fixe à son récepteur appartenant à la catégorie des récepteur à AMP cyclique (AMPc) et l'active. Ces récepteurs ont la propriété de générer un second messager, l'AMPc, qui va activer une protéine kinase A (PKA), qui est elle-même capable de phosphoryler la sérine 133 de CREB, sérine située dans son domaine KID (Kinase Inductible Domain) ce qui active CREB. CREB peut alors recruter les coactivateurs transcriptionnels CBP et P300 qui vont acétyler l'ADN et donc faciliter son accès aux facteurs de transcription présents dans l'environnement génomique (Chrivia et al. 1993).

CREB ainsi que sa cible CBP sont impliquées dans la régulation de nombreux facteurs de transcription via sa fonction de protéine de scaffold. Une protéine de scaffold permet la mise en place de complexes protéiques fonctionnels. En effet, CBP peut former des complexes contenant les protéines nuclear receptor co-activators NCOA2, NCOA3, les protéines IKKA, IKKB et IKBKG de la famille des serine/threonine protein kinases. NCOA2, NCOA3 sont des coactivateurs transcriptionnels des récepteurs stéroïdiens et des récepteurs nucléaires et IKKA, IKKB et IKBKG sont des protéines appartenant à la voie de signalisation du NF-kappaB. CBP/p300 peut très probablement former un complexe avec HIF1A qui est un régulateur transcriptionnel majeur de la réponse à un stress hypoxique. CBP peut aussi interagir avec divers autres protéines telles que le proto-oncogène MAF, le coactivateur CARM1 ou coactivator-associated arginine méthyltransferase 1, ou bien encore avec la protéine PCAF ou P300/CBP-associated factor qui est une lysine acétyltransférase 2B et une co-activateur transcriptionnel associé avec CBP/p300 mais aussi avec P53 ((Gu et al. 1997). CBP est entre autres associé avec les protéines promyelocytic leukemia ou PML (Doucas et al. 1999) qui sont des facteurs de transcription et des suppresseurs de tumeurs. La translocation de ce

gène avec le gène du récepteur à l'acide rétinoïque alpha est souvent associée avec les leucémies promyéloïdes aigues (Lin et al. 2001). De plus les protéines PML interagissent non seulement avec CBP mais aussi avec TDG (Lin et al. 2001) (

Figure 21, p51) et les récepteurs à l'acide rétinoïque (Zhong et al. 1999) que je vous présenterais au paragraphe II.3 (p55). Enfin on peut aussi signaler que CBP interagit avec la protéine codée par le gène Tat du virus HIV-1 via son domaine KIX ; HIV-1 TAT est une protéine transactivatrice de la transcription des gènes du virus ((Vendel et al. 2003), (Vendel et al. 2004)). L'activité de protéine de scaffold de CBP peut être régulée dans certains cas par sumoylation comme lors de son interaction avec DAXX (autre protéine de scaffold présentant de nombreuses fonctions cellulaires) (Huang et al. 2012). CBP subit de nombreuses modifications post-traductionnelles au niveau de ces différents domaines ; le KIX peut être méthylé par CARM1 ce qui inhibe l'interaction CBP/CREB et donc bloque la signalisation dépendante de CREB et inactive l'apoptose ; CBP peut aussi être phosphorylé ou sumoylée via DAXX (Huang et al. 2012) (

Figure 22, p51).



Figure 21 : TDG, un substrat de CBP (Léger H).

CBP est non seulement activer par TDG mais peut aussi acéyler TDG au niveau des lysines 59, 83, 84 et 87 situées dans le domaine de régulation de TDG.



Figure 22 : Schéma des domaines de CBP/p300 (adapté de (Smet-Nocca et al. 2008)).

CBP possède donc de multiples fonctions biologiques dont le contrôle de la prolifération cellulaire, l'homéostasie, le développement embryonnaire, la plasticité synaptique, la survie neuronale, le développement de l'apprentissage, la potentialisation à long terme c'est-à-dire la mémoire, l'hématopoïèse ou encore la spermatogénèse grâce à leur activité de remodelage de la chromatine qui est ainsi accessible pour les facteurs de transcription. Un exemple de l'interaction entre apprentissage et activité des récepteurs stéroïdiens est le rôle de CBP dans la transactivation des récepteurs aux androgènes et aux œstrogènes dans le cerveau des oiseaux lors de la mémorisation des chants sous l'influence des hormones sexuelles. CBP joue un rôle important pour la régulation de l'activité

transcriptionnelle des acteurs du système de chant sous l'influence des récepteurs stéroïdiens gonadiques (Auger et al. 2002).

II.2.b Maladies liées à CBP

Des mutations du gène Cbp et Ep300 ou des micro délétions de la région 16p13.3 du chromosome 16 sont la cause du syndrome Rubinstein-Taybi (SRT) (Coupry et al. 2002). Cette maladie génétique appelée aussi syndrome des pouces et des gros orteils trop larges ((Rubinstein and Taybi 1963), (Rubinstein and Taybi 1965)) est un syndrome dysmorphique combiné avec un retard mental, un retard de croissance, diverses malformations plus ou moins importantes selon la pénétrance de la maladie et une forte augmentation du risque de cancers, notamment du système nerveux (Iyer et al. 2004). C'est une maladie d'origine génétique présentant une incidence d'environ 1 naissance sur 125000 nouveau-nés vivants (Roelfsema et al. 2007).

Les malades présentent une voute du palais très ogivale, une orientation en bas et en dehors des fentes palpébrales, un ptosis, une orientation basse des oreilles, un nez prononcé, des pouces et des gros orteils plus larges que la moyenne (

Figure 23, p53), ainsi que diverses autres malformations dont la pénétrance est complète comme une microcéphalie, présente chez 50% des patients ou bien une cryptorchidie ou une cardiopathie dans 30% des cas (Roelfsema et al. 2007). Par ailleurs, de nombreux Rubinstein-Taybi présentent un retard de développement caractéristique du syndrome ; en effet l'âge moyen de la marche chez ces patients est de 30 mois, les premiers mots à 24 mois, tandis que la propreté est atteinte vers 5 ans (orphanet ; Syndrome Rubinstein-Taybi, Pr Lacombe, 2004).



Figure 23 : Patient RTS (Hennekam et al. 2006). A et B) Structure faciale caractéristique avec une orientation basse des oreilles, un nez prononcé ; C) voute du palais très ogivale ; D) pouces et des gros orteils plus larges que la moyenne.

L'ensemble des données du RTS indiquent que le niveau de CBP est limité dans la cellule, que CBP joue un rôle physiologique crucial dans l'embryogenèse, et que p300 ne peut pas compenser la fonction de CBP.

Par ailleurs, les mutations des gènes Cbp et Ep300 sont aussi impliquées dans l'apparition des leucémies myéloïdes. Ainsi lors des leucémies myéloïdes aigues ou acute myeloid leukemia (AML), il y a la présence d'une translocation chromosomique entre les gènes Moz (monocytic leukemia zinc finger) et Cbp aboutissant à l'expression d'une protéine de fusion MOZ-CBP comprenant le domaine acétyltransférase de MOZ fusionné à la quasi-totalité de CBP (Borrow et al. 1990). Il existe par ailleurs des protéines de fusion MLL-CBP, où la protéine Mixed Lineage Leukemia (MLL) (Lavau et al. 2000). MIl est un gène homologue aux gènes trithorax (trX) de chez la Drosophilia melanogaster ; ce gène est retrouvé dans plus de 80 % des leucémies de l'enfant et code pour une histone méthyltransférase impliquée dans le maintien de la régulation épigénétique de la transcription (Lavau et al. 2000).

II.2.c Interaction CBP/TDG et ouverture vers un couplage transcription/réparation de l'ADN

Nous savons que TDG interagit de façon fonctionnelle avec les corégulateurs de transcription CBP/p300 ((Tini et al. 2002)) et SRC-1 ((Lucey et al. 2005)) dans le cas qui nous intéresse. En effet, le domaine N-terminal de TDG est un domaine basique riche en lysines ciblé par les domaines –CH2 et – CH3 de CBP/p300 (Figure 21, p51 et

Figure 22, p51); le complexe CBP/TDG ainsi formé est impliqué dans la réparation de l'ADN par excision des bases thymines erronées dans les mésappariements G: T (Tini et al. 2002), et dans la régulation transcriptionnelle via l'interaction avec des facteurs de transcription tels que les récepteurs rétinoïques (Um et al. 1998). Il est donc le lien entre la transcription des gènes et la réparation de l'ADN via le BER (Tini et al. 2002) tout comme le fait l'enzyme TFIIH entre la réparation de l'ADN via le NER et la transcription des gènes (paragraphe I.3.b, p34). Le mécanisme de réparation de l'ADN initié par TDG fait de lui un facteur de signalisation épigénétique ; par conséquent, le complexe CBP/TDG donne pour la première fois l'aperçu d'un couplage direct entre signalisation épigénétique et régulation transcriptionnelle (Tini et al. 2002).

Des études précédentes ont mis en évidence que TDG interagit aussi fonctionnellement avec les facteurs de transcription de la famille des récepteurs nucléaires ligand-dépendants et plus précisément avec les récepteurs à l'acide rétinoïque RAR : RXR (Um et al. 1998) et avec les récepteurs aux œstrogènes ER (Chen et al. 2003). De plus CBP/p300 sont elles aussi impliquées dans l'expression des gènes liés aux récepteurs nucléaires ligand-dépendants RAR : RXR, ER, ainsi qu'aux récepteurs à la progestérone (PR), à l'hormone thyroïdienne (TR) ou aux glucocorticoïdes (GR). Ces récepteurs nucléaires opèrent en recrutant plusieurs coactivateurs pour exercer leur activité transcriptionnelle en réponse aux stimuli hormonaux dans diverses voies de différenciation cellulaire. Par exemple, les récepteurs ER et PR recrutent CBP qui lie, d'une part l'acétyltransférase, la SRC-1 (steroid receptor coactivator-1), via sa partie C-terminale et de l'autre part, les récepteurs nucléaires RAR : RXR via sa partie N-terminale. Ainsi, CBP et SRC-1 stimulent ces récepteurs à l'œstrogène et à la progestérone de façon coopérative. Un complexe fonctionnel de TDG avec CBP et RAR semblerait donc servir à l'intégration de la signalisation épigénétique via la méthylation de l'ADN dans la régulation transcriptionnelle de programmes génétiques majeurs.

II.3. Le Récepteur à l'Acide Rétinoïque (RAR : RXR)

II.3.a Structure des récepteurs à l'acide rétinoïque RAR et RXR

Comme je vous l'ai présenté ci-dessus, des études précédentes ont mis en évidence que TDG interagit aussi fonctionnellement avec les facteurs de transcription de la famille des récepteurs nucléaires ligand-dépendants et plus précisément avec les récepteurs à l'acide rétinoïque RAR : RXR. En effet, un criblage chez la levure des cofacteurs putatifs des récepteurs à l'acide rétinoïque RARs a montré que TDG, connu à ce moment-là, comme une enzyme de réparation de l'ADN, interagirait avec les RAR (Um et al. 1998). Les récepteurs à l'acide rétinoïque comprennent les retinoic acid receptors ou RARs et les retinoid X receptors ou RXRs. Les récepteurs nucléaires sont des récepteurs spécifiques des eucaryotes c'est-à-dire du règne animal. Ainsi les éponges Amphimedon queenslandica possèdent deux récepteurs nucléaires (Reitzel et al. 2011), les cnidaires Nematostella vectensis en possèdent dix-sept (Bridgham et al. 2010), le nematode C. elegans en possède 270 (Sluder et al. 2001) tandis que les humains, les souris et les rats en possèdent respectivement 48, 49, et 47 chacun (Zhang et al. 2004). Les 48 récepteurs nucléaires humains peuvent être classés en six familles selon leur homologie de séquence. Les récepteurs nucléaires RARs et RXRs appartiennent à la superfamille des récepteurs aux stéroïdes ou Thyroid Hormone Receptor-like ; ce sont des récepteurs nucléaires de type II, généralement activés via la liaison de leur ligand tel des hormones, la vitamine A ou la vitamine D sur leur site actif. Il existe cependant quelques récepteurs nucléaires orphelins qui sont activés par la liaison de produits de métabolisme tels les acides gras ou bien les stérols ; d'autres récepteurs nucléaires sont activés de façon ligand-indépendante. Les récepteurs RARs et RXRs existent tous les deux sous trois isotypes : alpha (α), beta (β) et gamma (γ) codés par des gènes situés sur le même locus (Germain et al. 2006). Ces différents isotypes diffèrent au niveau de leur structure primaire et au niveau de leur réponse aux rétinoïdes synthétiques. Tous les récepteurs RARs et RXRs possèdent la structure très conservée des récepteurs nucléaires c'est à dire qu'ils possèdent cinq (RXRs) ou six (RARs) régions nommées de A à E ou A à F, respectivement. Les récepteurs nucléaires présentent une région centrale C appelée domaine de liaison à l'ADN ou DNA binding domain (DBD), fortement conservée au cours de l'évolution, une région E contenant le domaine de liaison au ligand ou ligand binding domain (LBD), une ou deux régions N-terminales A/B et une région C-terminale F (Germain et al. 2006) (

Figure 24, p56). Comme tous les récepteurs nucléaires de type II ils possèdent deux interfaces d'homo RAR : RAR ou d'hétérodimérisation RAR : RXR, fonctionnant de manière autonome l'un de l'autre et localisées au niveau du DNA binding domain (DBD) et du ligand binding domain (LBD) (Allenby et al. 1993). Ces domaines permettent aussi d'interagir avec les cofacteurs tels que CBP.



Figure 24 : Schéma des récepteurs nucléaires (adapté de <u>http://nrresource.org</u>).

Les récepteurs nucléaires présentent donc plusieurs domaines appelés de A à F. Les domaines A/B sont les domaines d'activation indépendamment du ligand, la région centrale C est le domaine de liaison à l'ADN, la région C ou hinge domaine fait le lien entre la partie N et la partie C-terminales de la protéine et les régions E et F, situées en C-terminal de TDG, contiennent le domaine de liaison au ligand (Germain et al. 2006).

II.3.b Implication dans la morphogenèse et l'oncogenèse

Les récepteurs RAR et RXR diffèrent entre eux dans leur capacité à lier les formes isotopiques de l'acide rétinoïque ; en effet les RAR peuvent lier les deux isotopes de l'acide rétinoique, le all-trans-retinoic acid (ATRA) et le 9-cis-retinoic acid (9c-RA), tandis que les récepteurs RXR ne peuvent lier que le 9c-RA.

Les récepteurs RAR : RXR lient les différentes formes d'acide rétinoïque (RA) (Allenby et al. 1993). Le RA est un métaboliste de la vitamine A, une vitamine liposoluble qui existe dans l'organisme sous forme de rétinol, de rétinal, d'acide rétinoïque et de rétinyl phosphate. Elle est impliquée dans la croissance des os et la synthèse de pigments de l'œil. Un sévère déficit en vitamine A peut provoquer une cécité qui commence par une photophobie et une cécité crépusculaire auxquelles s'ajoute rapidement divers troubles de croissance, une dédifférenciation des épithéliums et une diminution de la résistance aux infections. Si elle est prolongée, la carence en vitamine A peut être mortelle, tout comme une surdose. Le RA agit au cours de l'embryogenèse comme un agent morphogène ; en effet lors de l'embryogenèse, certaines cellules embryonnaires comme celles du nœud de Hensen (gastrulation) ou celles de la zone d'action polarisante (ZAP) des bourgeons des membres, produisent de l'acide rétinoïque qui va diffuser dans l'embryon, grâce à son caractère lipophile, créant ainsi un gradient d'acide rétinoïque antéro-postérieur lors du développement précoce de l'embryon ou un gradient proximal-distal lors du développement des membres. Le RA va se fixer sur son récepteur qui active la transcription des gènes-cibles homéotiques Hox. La sensibilité des différents gènes Hox à l'acide rétinoïque dépend de leur position sur le chromosome, les gènes situés en position 3' étant les plus sensibles ; de plus, la sensibilité à l'acide rétinoïque est maximale dans la zone antérieure de l'embryon ou dans la zone des bourgeons des membres (http://homepage.mac.com/danielbalas/BDDMOL/embmol/embmolpdf/Lecture3.pdf) (Figure 25, p58).



Figure 25: Action de l'acide rétinoïque au cours du développement embryonnaire (Léger H) A) mise en place de l'axe antéro-postérieur ; B) mise en place de l'axe proximo-distal d'un membre via les gènes Hox-D (adapté de http://homepage.mac.com/danielbalas/BDDMOL/embmol/embmolpdf/Lecture3.pdf)

L'acide rétinoïque est nécessaire au développement homéotique des embryons des vertébrés, mais il peut aussi avoir une action tératogène ; en effet, un excès d'acide rétinoïque exogène entraine des transformations homéotiques en repoussant vers l'arrière la limite d'expression des gènes Hox les plus postérieurs. Il contrôle donc la différenciation et la prolifération cellulaire de façon spatio-temporelle (Choi et al. 2012). Le RA peut aussi réguler l'homéostasie des tissus ou bien contrôler l'apoptose (Todaro et al. 2012). Les récepteurs RAR et RXR sont avant tout des activateurs transcriptionnels mais ils peuvent aussi réprimer l'expression de certains gènes (Su et al. 2012).

Les récepteurs à l'acide rétinoïque possèdent aussi une activité de régulation épigénétique de la transcription via leur capacité à interagir avec des cofacteurs tels que CBP ou SRC-1 ou bien avec le facteur de transcription général TFIIH (Bour et al. 2007). La capacité de l'hétérodimère RAR : RXR à reconnaitre et à fixer spécifiquement les retinoic acid response elements (RARE) situées dans la région promotrice des gènes cibles de l'acide rétinoïque sont un autre mécanisme épigénétique sous la dépendance des RAR : RXR. Les RARE sont des séquences de l'ADN constituées d'un motif consensus semblable à -GGTAA, directement répété et espacé de cinq ou sept nucléotides. Les hétérodimères RAR: RXR peuvent se lier aux RAREs en absence ou en présence d'acide rétinoïque avec une conformation différente dans chaque cas. Lorsque RAR : RXR se fixe aux RAREs en absence d'acide rétinoïque, il est dans la conformation « OFF » c'est-à-dire qu'il découvre le CoR binding site qui permet de fixer le corépresseur des récepteurs à l'acide rétinoïque et des récepteurs à la thyroïde NCor (Glass et al. 2000) ou Smrt. Ce corépresseur contient un site de liaison pour le complexe histone déacétylase HDAC qui déacétyle l'environnement chromatinien et induit donc la condensation de la chromatine et ainsi la répression de la transcription. Lorsque l'hétérodimer RAR : RXR se fixe aux RARE en présence d'acide rétinoïque, le récepteur RAR est dans une conformation allostérique qui cache le domaine NCor et met en évidence un site de liaison pour le coactivateur CoA. Dans cette position "ON", le coactivator CoA possède un site histone acétyltransferase (HA) qui entraine la décondensation de la chromatine et l'interaction avec le complexe transcriptionnel ; il y a donc activation de la transcription des gènes situés dans cet environnement chromatinien (Linney et al. 2011) (

Figure 26, p60).



Figure 26 : Modèle de régulateur épigénétique de RAR : RXR (Linney et al. 2011) A gauche, complexe RAR : RXR / corépresseur de la transcription / HDAC qui inhibe la transcription par condensation de la chromatine. A droite, complexe RAR : RXR / coactivateur de la transcription / HAT qui active la transcription par acétylation de la chromatine. RARE : séquences régulatrices appelées retinoic acid response elements

Les récepteurs à l'acide rétinoïque jouent donc un rôle important de commutateur épigénétique de la transcription, de façon spatio-temporelle au cours du développement embryonnaire et plus particulièrement au cours du développement du système nerveux.

II.3.c Complexe CBP/TDG/RAR α : cœur de mes travaux

Comme je vous l'ai présenté au paragraphe précédent, les récepteurs à l'acide rétinoïque possèdent aussi une activité de régulation de la transcription via leur capacité à interagir avec des cofacteurs tels que CBP ou SRC-1 ou bien avec le facteur de transcription général TFIIH ou via leur activité de commutateur épigénétique. Par ailleurs, il a été récemment montré (Cortázar et al. 2011) que TDG possède une activité transcriptionnelle au cours du développement et plus précisément au cours de la mise en place du système nerveux sous l'influence de l'acide rétinoïque. L'activité de TDG semble par ailleurs régulée via son interaction avec CBP au sein d'un complexe CBP/TDG.

Lors de mon projet de thèse, j'ai commencé à caractériser les paramètres fonctionnels de l'interaction CBP/TDG ainsi que l'impact de ce complexe sur la transcription des gènes dépendant des récepteurs à l'acide rétinoïque RAR : RXR. La caractérisation de l'interface CBP/TDG, l'étude de la réparation de l'ADN et de l'activité

transcriptionnelle de TDG ainsi que l'obtention de profils épigénomiques fonctionnels du complexe CBP/TDG via des études à haute précision du transcriptome permettront à terme de mieux comprendre le rôle de ces protéines dans la mise en place et le maintien d'un organisme sain. Ils permettront aussi de caractériser de nouvelles cibles thérapeutiques dans les leucémies et le syndrome de Rubinstein-Taybi.



Figure 27 : Le complexe CBP/TDG, un lien entre la réparation de l'ADN et la régulation épigénétique de la transcription via son interaction avec les récepteurs nucléaires (adapté de Smet Nocca)

Chapitre III L'analyse de données transcriptomiques par microarrays

Depuis plusieurs décennies, les équipes de recherche s'intéressent à mieux connaitre et comprendre les mécanismes qui contrôlent l'expression de nos gènes. Pour cela il existe depuis plus de 15 ans la technique des puces à ADN qui permettent d'accéder à une partie ou à la totalité de différents génomes. Dans ce chapitre, je vous présenterais la technologie des puces à ADN de façon globale en décrivant rapidement les caractéristiques des différents types de puces à ADN ; je vous présenterais ensuite plus en détail la/les technologies utilisées au cours de ma thèse et je vous expliquerais comment j'ai réalisé mes analyses transcriptomiques.

III.1. Présentation de la technologie des puces à ADN

L'étude de l'ADN a débuté par la mise au point et l'utilisation de la technique du Northern-Blot inversé qui consiste à faire migrer par électrophorèse, sur gel d'agarose ou de polyacrylamide, des molécules d'ADN qui vont aller s'hybrider à des ARNm marqués à l'aide de nucléotides radioactifs ou fluorescents. Les progrès réalisés en synthèse d'oligonucléotides in situ (Maskos et al. 1992), l'apparition de librairies d'ADNc de plus en plus complètes (Lennon et al. 1991) et les progrès réalisés en chimie de surface ont permis l'apparition, au cours des années 90, de la technologie des puces à ADN (Ekins et al. 1999); (Schena et al. 1998)). Des travaux portant sur la mesure de l'expression différentielle simultanée de 45 gènes d'Arabidopsis Thaliana grâce à utilisation des puces à ADN ont étaient publiés pour la première fois en 1995 par M. Schena et ses collaborateurs dans la revue Sciences (Schena et al 1995). La même équipe a publié l'année suivante un article relatant l'utilisation de puces pouvant observer plus de 1000 gènes humains (Schena et al. 1996). Depuis ces articles, le nombre de ces publications et les progrès techniques réalisés ont augmenté de façon continue et corrélée, notamment pour l'étude du génome humain et de certaines espèces telles la souris ou le rat et permettent actuellement l'étude du génome entier pour la souris, le rat ou l'humain.

De nombreuses plateformes de données transcriptomiques ont vu le jour et il en existe actuellement onze dont les données sont regroupées au niveau du site internet Gene Expression Omnibus (GEO) ((Edgar et al. 2001) (http://www.ncbi.nlm.nih.gov/geo/)). Ces plateformes proposent des puces à ADN de différents types car il existe différents moyens de marquer les oligonucléotides hybridés sur les puces à ADN ainsi que différentes façon de lire les puces et d'analyser les résultats par la suite.

On a deux grands types de puces, les puces à ADNc, les premières existantes historiquement (Schena et al. 1995), et les puces à oligonucléotides, elles-mêmes pouvant être diversifiées en puces à simple ou double marquage, utilisées pour mesurer l'expression d'un seul échantillon ou plusieurs simultanément.

Les puces à ADNc sont composées de fragments d'ADNc obtenus par réactions par polymérisation en chaine (PCR) et liés par liaison covalente sur la puce, sonde par sonde dans chaque puits. Les liaisons sont réalisés à l'aide de diverses techniques telles que les micropipettes (Schena et al. 1995), la photolithographie ((Pease et al. 1994), (Stengele et al. 2005), (Albert et al. 2003)) ou l'électrochimie. L'ARNm est rétro-transcrit et marqué dans une même étape pour donner l'ADNc marqué qui s'hybridera aux sondes fixées à la membrane de la puce. Il existe actuellement des versions commerciales de ce type de puce mais la plupart du temps les puces à ADNc sont utilisées lorsque l'on souhaite réaliser ses propres microarrays ou puces à façon, car elles possèdent l'avantage de pouvoir assez facilement être produites et donc personnalisées, tant pour le choix des sondes que pour leur localisation sur la puce. De plus elles présentent une grande sensibilité car les sondes utilisées sont longues (entre 100 et 200 nucléotides de long). Malheureusement leur flexibilité d'utilisation (puces à façon) entraine de plus grandes lacunes au niveau des contrôles et de la quantification des ARN mesurés (Bammler et al. 2005).

Les puces à oligonucléotides possèdent des sondes sous forme de courtes molécules d'ADN monobrin (de 25 nucléotides pour les puces Affymetrix à 60 nucléotides pour les puces des technologies Applied Biosystems et Agilent) mais elles proviennent de séquences génomiques ou de séquences label exprimées (EST : Expressed sequence Tag) issues de fragments d'éléments transcrits. La principale différence avec les puces à ADNc réside dans le fait qu'à la place des séquences complètes des gènes sont utilisées des fragments de quelques nucléotides de long, spécifiques de zones ciblées du gène, placées ou synthétisées *in situ* sur la puce par photolithographie (Fodor et al. 1991). La différence

de longueur des sondes de chaque plateforme implique une différence de spécificité et de sensibilité ((Kane et al. 2000), (He et al. 2005)).

Les puces Affymetrix, dont les sondes de 25 nucléotides ou 25-mers sont plus faciles à produire et autorisent une plus grande densité de sondes. Elles présentent une plus faible spécificité de probe compensée par un plus grand nombre de sondes par gène que les technologies à utilisant des sondes de 60-mers telles que Applied Biosystems et Agilent. Chaque technologie possède en outre ses propres controles d'hybridation. Ainsi les puces Affymetrix, la spécificité de l'hybridation est évaluée par la présence pour chaque probe d'une probe dont la séquence est identique à l'exception du nucléotide central, et qui ne doit donc théoriquement pas hybrider le transcrit. Ces sondes servent de témoin négatif.

Les puces à simple marquage permettent d'analyser une condition par puce et le marquage se fait généralement par des fluorochromes ; certaines plates-formes utilisent n marquage radioactif par incorporation de P32 dans les acides nucléiques telle BD Biosciences Clontech tandis que d'autres utilisent la chimiluminescence. C'est le cas de la technologie Applied Biosystems que j'ai utilisée pendant la majeure partie de mes travaux de thèse. Les ADNc et ARNc sont alors marqués par l'incorporation de nucléotides DIG-UTP et DIG-dUTP (en fonction du type d'amplification du matériel génétique : simple rétro-transcription ou rétro-transcription suivit d'une transcription in vitro). Les puces à double marquage permettent quant à elles de marquer deux échantillons, le plus souvent avec deux fluorophores, Cy3 et Cy5, que l'on hybride ensemble sur une même puce. Leur fluorescence est mesurée subséquemment de façon séparée et les images produites sont analysées en termes de ratio de l'émission des deux fluorophores.

Les plateformes commerciales offrent des puces présentant un nombre de sondes pouvant aller de 24267 pour les puces TORAY à 400000 pour les puces de chez NimbleGen, en fonction de leur besoin en spécificité. En effet, posséder plusieurs sondes par gènes, correspondant à différentes régions de ces derniers, permet de limiter au maximum les problèmes dus à l'hybridation non-spécifique et aux imprécisions d'annotations, et ainsi que d'être certains de reconnaitre tous les isoformes possibles d'un gène en possédant une sonde par exon (Pellay 2008). Ces plateformes commerciales proposent en plus des puces pangénomiques, des puces spécifiques à un nombre plus restreint de gènes ; ces puces sont à un prix plus abordable pour les utilisateurs qui ne cherchent pas analyser le génome entier. De nombreuses plateformes proposent également la production de puces à façon afin de cibler l'analyse aux gènes d'intérêt de l'utilisateur.

Par ailleurs les progrès mathématiques et informatiques (méthodes de normalisation, des analyses statistiques ou des méta-analyses des données) ont permis d'améliorer l'analyse des données issues de ces puces à ADN. En effet, des milliers de publications ont été consacrées ces dernières années à l'amélioration des analyses microarrays. L'équipe de recherche à laquelle j'appartiens travaille évidemment aussi à l'amélioration des méthodes d'analyse des données issues de notre plateforme transcriptomique en plus de notre intérêt dans les mécanismes de régulations épigénétiques de la transcription.

III.2. Les puces de la Technologie Applied Biosystems

III.2.a Caractéristiques des puces

Lors de ma thèse, j'ai étudié le rôle du complexe CBP/TDG sur la régulation de l'expression des gènes et notamment de ceux sous la dépendance de l'acide rétinoïque. Pour cela j'ai réalisé des expériences de puces à ADN à l'aide de la technologie des puces Applied Biosystems.

Les puces pangénomiques Applied Biosystems existent pour trois organismes modèles, les humains (Homo sapiens), les souris (Mus musculus), et les rats (Rattus norvegicus). Ce sont des puces pangénomiques à oligonucléotides simples marquages.

Ces puces sont formées d'une plaque de verre recouverte d'une membrane sur laquelle sont fixées par liaison covalente 34000 sondes de 60-mers de long couvrant environ 29000 gènes pour les puces humaines, 32000 pour les puces souris et 27000 pour les puces rats (

Figure 28, p66). Les gènes identifiés viennent de la compilation des données issues de quatre bases de données différentes, la « Celera Genomics Human Genome Database », la « LocusLink public database », la « Mammalian Gene Collection » et la « GenBank »

complétés par certains gènes identifiés par la compagnie Applied Biosystems elle-même (Applied Biosystems 2004). Les sondes dessinées et synthétisées par Applied Biosystems sont plus longues que les sondes utilisées par Affimetrix pour ses puces, ce qui apporte une meilleure sensibilité mais augmente aussi le risque de mésappariements. Pour compenser cela, les sondes ont été choisies majoritairement dans la partie 3' UTR non-codante des gènes car c'est la région la plus spécifique des gènes mais aussi celle partagée par la majorité des variants d'épissage des gènes. Les polymorphismes génétiques, les séquences répétitives de l'ADN et divers autres mécanismes entrainent l'apparition de 15% de gènes qui sont reconnus par deux à douze sondes tandis que les 85% des sondes Applied Biosystems restantes sont spécifiques d'un seul gène.



Figure 28 : Puces pangénomiques Applied Biosystems A) humaine, B) souris, C) rat.

La société Applied Biosystems a validé expérimentalement la qualité de ses sondes en testant ses puces humaines sur 52 tissus humains ce qui a permis d'éliminer les sondes ayant un comportement aberrant et ne conserver que celles ayant le meilleur signal, la meilleure concordance avec la PCR quantitative, et la meilleure spécificité dans le cas de gènes ayant plusieurs sondes possibles. Les annotations écrivant les gènes détectés par les puces Applied Biosystems ont été obtenues en combinant les informations issues de différentes bases de données libres d'accès telles que GenBank, Swiss-Prot, Medline, Unigene ou encore RefSeq. Les bases de données fonctionnelles Gene Ontology et de PANTHER (Thomas et al. 2003) apportent une information fonctionnelle pour chaque gène dans le fichier d'annotation lié aux puces de chaque espèce.

Les puces Applied Biosystems sont enfermées dans une cassette de plastique lors des étapes de pré-hybridation et d'hybridation qui ont lieu à 55°C sous agitation et n'en sont sorties que pour les étapes de lavage et lecture grâce à l'analyser Applied Biosystems 1700 (



Figure 29, p67).

Figure 29 : Caméra et analyseur de la technologie Applied Biosystems 1700 (Applied Biosystems).

III.2. b- Caractéristique du mode de lecture

La plupart des compagnies de puces à ADN utilise la fluorescence pour détecter les transcrits tandis que la société Applied Biosystems a fait le choix d'utiliser la chimiluminescence comme méthode de détection. En effet, les ARNc et ADNc obtenus respectivement par simple rétro-transcription ou rétro-transcription suivie d'une transcription *in vitro* sont marqués par l'incorporation de nucléotides digoxygenin-UTP (DIG-UTP) ou digoxygenin –dUTP (DIG-dUTP). Une fois hybridés sur la puce, ils sont reconnus par un anticorps anti- digoxygenin couplé à une alcaline phosphatase qui coupera le dérivé de 1,2-Dioxetan présent dans le substrat de chemilumiscence déposé sur la puce juste avant sa lecture par l'analyser Applied Biosystems 1700 et émettra de la chemilumiscence (Figure 30, p68).



Figure 30 : Hybridisation et détection par chimiluminescence des acides nucléiques marqués avec de la digoxygenine sur des puces pangénomiques de la technologie Applied Biosystems (Applied Biosystems).

Ce système permet d'amplifier fortement le signal émis par chaque transcrit hybridé et donc d'obtenir une grande sensibilité de détection en réduisant la limite de détection à 2.10⁻²¹ mol (Figure 31, p69). Cette technologie de puce permet donc de détecter des gènes même faiblement exprimés. En contrepartie, il y a risque accru de détecter des hybidations non-spécifiques. Pour séparer ces faux-positifs des vrais résultats, l'émission de chaque probe est effectuée plusieurs fois et comparée à celle de sondes de contrôle.

Marqueur	Méthode	Limite de détection (mol)
Fluorescéine	Fluorescence	2x10 ⁻¹⁴
Rhodamine	Fluorescence	5x10 ⁻¹⁵
Péroxidase	Colorimétrique	1x10 ⁻¹⁶
Phosphatase alcaline	Colorimétrique	2x10 ⁻¹⁶
Phosphatase alcaline	Chimiluminescence	2x10 ⁻²¹

Figure 31 : Tableau de comparaison de la sensibilité des différents types de marquages des acides nucléiques (Pellay 2008)

III.3. Analyse transcriptomique par microarrays

Les données brutes exportées depuis la plate-forme ont été analysées avec le logiciel ace.map. Les analyses nécessitent l'entrée des informations sur les données et la création de fichiers ma0, utilisés pour le reste de l'analyse. Chaque puce est représentée par un fichier ma0 et l'ensemble des puces d'un même réplicat, technique ou biologique, peut être fusionné en un seul fichier ma1. L'homogénéité entre les différents réplicats biologiques a été vérifiée en utilisant les moyennes de coefficient de corrélation de Pearson sur l'ensemble des sondes de la puce.

Les fichiers ma0 et ma1 peuvent ensuite être utilisés pour mesurer des différences d'expression entre différentes conditions biologiques par l'intermédiaire de soustractions qui donnent les fichiers ma2. Les calculs de soustraction des profils transcriptomiques des différentes conditions ont été réalisés en utilisant la procédure standard de type "everyone-against- everyone".

Le calcul de la différence d'expression pour chaque probe a été déterminé sous la forme d'un logarithme à base 2 (log-Q) et correspond à la moyenne, pondérée par les variances, des différentes valeurs de log-Q. Ce calcul utilise la méthode NeONORM (Noth

et al. 2006) pour la normalisation des valeurs de signal. Les valeurs de p-value ont été obtenues en utilisant une méthode ANOVA (Bécavin et al. 2011), (Brysbaert et al. 2010), (Brysbaert et al. 2007)

Les fichiers ma0, ma1 et ma2 sont ensuite filtrés selon différents critères et les listes de sondes filtrées sont contenues dans des fichiers ma3. Les fichiers ma4 sont construits à partir de fichier ma0 et permettent l'analyse d'études cinétiques.

Les analyses par ontologies des données transcriptomiques ont été réalisées en utilisant les annotations fonctionnelles de chaque gène décrites dans Gene Ontology et PANTHER. Les sondes considérées comme significativement différentiellement exprimées entre deux conditions sont analysées en terme de voie de régulation, de processus biologique et de fonction moléculaire. La surreprésentation d'une ontologie dans un des groupes de sondes analysées est mesurée en terme de p-value, calculée par une loi binomiale et corrigée pour tenir compte de la dépendance entre les ontologies.

Matériels et méthodes

I_Matériel

I.1. Description des lignées cellulaires utilisées

L'étude de l'interaction CBP/TDG/RAR et de son impact sur le transcriptome a été réalisée *in vivo* dans des cellules COS-7 (African Green Monkey SV40-transfd kidney fibroblasts) ainsi que dans des cellules HEK293 (Human Embryonnic Kidney 293).

Le nom, COS est un acronyme significant CV-1 (singe) in Origin, and carrying the SV40 genetic material. Les lignées cellulaires COS sont donc des lignées obtenues après immortalisation de cellules CV-1, dérivées de fibroblastes de singes verts africains *Cercopithecus aethiops* ou African green monkey, à l'aide d'un virus SV40 contenant le large antigen T. Ces cellules sont donc un modèle de fibroblastes « humains », poussant en monocouche ; lorsqu'elles sont transfectées, elles permettent de produire des protéines recombinantes.

Les cellules HEK293 sont issues du 293ème clone de cellules de tissus de reins embryonnaires transformées à l'aide d'ADN de l'adénovirus 5 par Alex Van der Eb dans les années soixante-dix. Ces cellules sont relativement faciles à cultiver, ont une croissance cellulaire rapide et se transfectent très bien. Ces cellules ont été pendant longtemps considérées comme des dérivés de fibroblastes soit endothéliales soit épithéliales qui sont tous très abondant dans les reins. Plus récemment, Graham et ses collègues ont mis en évidence que les cellules HEK293 ainsi que d'autres cellules rénales embryonnaires humaines transformées ont de nombreuses propriétés de neurones immatures. Ces caractéristiques, fort taux de croissance cellulaire, facilité de transfection ont orienté notre choix vers ce type de cellules lors d'une seconde approche de l'impact du complexe CBP/TDG/RAR sur le transcriptome.
I.2. Description des plasmides utilisés

Les constructions utilisées lors de ma thèse ont été réalisées par Arndt Benecke et Amel Attmane (Systems Epigenomics Group) précédemment à ma thèse. Ces constuctions ont été réalisées par coupure enzymatique des plasmides codants pour les protéines sauvages de CBP, TDG et RARα suivi de leur insertion dans divers plasmides. De la mutagenèse dirigée a ensuite été réalisée par le Dr. Caroline Smet-Nocca (Unité de glycobiologie structurelle et fonctionnelle, CNRS UMR8576, anciennement postdoctorante au sein de l'équipe) pour obtenir les mutants P65A, K330A, K59A et K59/83/84/87A de TDG. Cette mutagenèse a été réalisée grâce au kit Quick Change II Site-Directed Mutagenesis Kit (Stratagene).

Lors de l'étude en FRET/FLIM de l'interaction CBP/TDG, réalisée par Corentin Spriet (IRI, Lille), les protéines TDG-wt et TDG mutants ont été clonées dans un plasmide pEYFP-C1 (Clontech, Cat N. 6005-1) au niveau des sites de clonage BglII/SalI tandis que la protéine CBP wt a été cloné dans un plasmide mCFP-C1 (Clontech, Mountain View, C) au niveau des sites de clonage BglII. Ces plasmides codent pour des protéines fluorescentes respectivement jeune-vert (EYFP) et cyan (CFP), dérivées de la Green fluorescente protein (GFP) produite naturellement par la méduse *Aequorea victoria* (Figure 32, p73). Par ailleurs CBP wt a été lui aussi cloné dans un plasmide pEYFP-C1 au niveau des sites de clonage BglII.

J'ai utilisé ces plasmides (EYFP –TDG-wt, EYFP-TDG mutants et CFP-CBP) pour l'étude de l'impact du complexe CBP/TDG puis du complexe CBP/TDG/RARa sur le transcriptome humain lors de transfections dans des cellules Cos-7. De plus, j'ai utilisé le plasmides EYFP-TDG-wt et EYFP-TDG-P65A ainsi que le plasmide EYFP-CBP pour étudier l'impact du complexe CBP/TDG/RAR dans les cellules HEK293. Le plasmide EYFP-CBP a permit de réaliser des expériences de thermophorèse pour étudier la dynamique d'interaction entre les protéines CBP/TDG en présence ou non de RARα et/ou d'acide rétinoïque (RA).



Figure 32 : Structure et spectres de la de la Green Fluorescente Protein (GFP) et de ses dérivés. A) Structure Green Fluorescente Protein (GFP) ; B et C) Spectres d'excitation et d'émission de la GFP et de ses dérivés.

Les protéines TDG-wt, TDG mutantes et CBP ont aussi été clonées dans le plasmide pSG5 (Agilent Cat No. 216201) au niveau des sites de clonage BglII/SacI dans le but d'obtenir des protéines recombinantes HH25-TDG possédant un tag HA. Enfin la protéine RARa a été obtenue grâce au plasmide Flag (pSG5-A4, Benecke A.) au niveau des sites de clonage dans le but d'obtenir des protéines recombinantes Flag- RARa, c'està-dire possédant un tag Flag.

I.3. Description des anticorps utilisés

Lors de ma thèse, j'ai utilisé divers anticorps primaires et divers anticorps secondaires couplés avec une peroxydase de raifort ou horseradish peroxidase pour l'étude en western blot de l'expression des protéines CBP, TDG et RARa. J'ai donc utilisé les anticorps primaires rabbit polyclonal anti-CBP (Cat. N° sc-1211) et le goat polyclonal anti-

hTDG (Cat. N° sc-22847) de chez Santa Cruz Biotechnology. Ce dernier n'étant pas assez sensible, j'ai essayé l'anticorps anti-TDG rabbit anti-TDG (Cat. N° SAB4502939) de chez Sigma-Aldricht. J'ai par ailleurs utilisé l'anticorps mouse anti-living color JL8 (Cat. N° 632380) de chez Clontech, l'anticorps mouse anti-HA (Cat No. 11 666 606 001) de chez Roche et l'anticorps mouse anti-Flag pour détecter les protéines taggées YFP, HA ou Flag respectivement. Pour purifier les protéines couplées à un tag HA ou un tag Flag, j'ai utilisé les anticorps mouse anti-HA agarose-conjugate, clone HA-7 (Cat. N° A 2095) ainsi que le kit d'immunoprécipitation associé (Cat. N° IP0010), et le kit d'immunoprécipitation anti Flag (Cat. N° FLAGIPT1) contenant l'anticorps anti-Flag-agarose-conjugate (Cat. N° A2220) tous trois fourni par Sigma-Aldricht.

Les anticorps secondaires avec une peroxydase de raifort utilisés ont été l'anti-goat (Sigma, Cat. N° A 5 420-1ML), l'anti-mouse (Sigma-Aldricht, Cat. N° A 9 044) et l'anti-rabbit (Ozyme, Cat. N°7 074).

I.4. Description des traitements utilisés

Lors de l'analyse transcriptomique des cellules COS-7, j'ai réalisé une série de données concernant la réponse de ces cellules à divers traitements en présence ou en absence des complexes CBP/TDG.

Les cellules ont été traitées avec une concentration de 5.10⁻⁷ M de 5-Aza-2'-Deoxycytidine (Sigma-Aldricht, Cat N. A3656) (Li et al. 2001). Le 5-Aza-2'-Deoxycytidine est un agent qui entraine une déméthylation complète ou une hémiméthylation de l'ADN ((Winkley and Robins 1970), (Chiurazzi et al. 1999), (Slack et al. 1999), (R. Z. Chen et al. 1998)). Pour les traitements j'ai préparé une solution stock de 5-Aza-2'-Deoxycytidine (5-Aza-dC) (Sigma-Aldricht, Cat. N°) à 0.5.10⁻⁴ M dans du DMSO (MW du 5-Aza-dC : 228.21 g.mol-1) que l'on conserve à l'abri de la lumière et à -80°C quelques semaines.

Les cellules ont aussi été traitées avec une concentration de 5.10⁻⁷ M de 5-Fluorouracile (Sigma-Aldricht, Cat N° F 6627). Le 5-Fluorouracile ou 5-FU est un agent chimique utilisé dans le traitement de certaines tumeurs solides. En effet, il s'intègre au niveau de l'ADN entrainant l'apparition de mésappariements 5-FU : G. Le 5-Fluorouracile est dissous dans du DMSO a une concentration de 5.10^{-3} M, cette solution stock est conservée à l'abri de la lumière et à 4°C, de façon stable au moins quatre mois.

Les cellules ont aussi été traitées avec une concentration de 5.10⁻⁷ M d'acide rétinoïque all trans (ATRA) (Sigma-aldricht, Cat. N° R 2 625). L'acide rétinoïque est dissous dans de l'éthanol absolu pour former une solution stock à 5.10⁻⁴ M, conservée à - 20°C et à l'abri de la lumière. A partir de cette solution stock, j'ai préparé une solution d'acide rétinoïque à 5.10⁻⁷ M dans une solution de MgCl² à 20 mM utilisée pour traiter les extraits protéiques lors des expériences de microthermophorèse.

II_Méthodes

II. 1. Transformation, amplification de bactéries XL1 blues compétentes

Lors de ma thèse, j'ai amplifié des bactéries transformées par le plasmide EYFP-TDG-wt, TDG mutant ou CBP induisant une résistance à la kanamycine à une concentration de 100 mg / l, par le plasmide pSG5-TDG-wt, TDG mutant ou CBP et par le plasmide RAR α induisant une résistance à l'ampiciline à une concentration de 100 mg / l.

Transformation de bactérie XL1 blue compétentes

50 uL de bactéries compétentes XL1 blue (Agilent, Cat. N° 200249) sont décongelées dans la glace et mises en présence de 1 μ g d'ADN plasmidique pendant 30 min. la transformation des bactéries à lieu par choc thermique pendant 45 s à 42°C puis à 4°C pendant 2 min. Pour une meilleure transformation des bactéries, les bactéries sont amplifiées une première fois dans 3 ml de LB contenant le bon antibiotique pendant 3h à 37°C sous agitation. 250 μ l du mélange bactérien est ensuite étalé sur une boite LB-agar contenant le bon antibiotique et sont incubées over-night (O/N) à 37°C.

Amplification des bactéries transformées

Préparation du LB medium : dissoudre 20 g de LB dans 11 d'eau stérilisée ; mettre 500 ml / bouteille et faire stériliser le LB. Le laisser refroidir jusqu'à la température ambiante puis ajouter, en milieu stérile, l'antibiotique correspondant, à sa concentration appropriée.

Pour une meilleure amplification des bactéries, les bactéries issues d'un stock glycérol ou d'une colonie de bactérie transformées précédemment, sont pré-amplifiées dans 3 ml de LB contenant le bon antibiotique pendant 3h à 37°C sous agitation. Puis on prend 200 ul de cette préparation et on l'ajoute à 500 ml de LB contenant le bon antibiotique à la concentration adéquate et on l'incube O/N à 37°C sous agitation.

La purification de l'ADN plasmidique se fait à l'aide du kit Nucleobond de chez Macherey-Nagel (Cat. N° 740414.5).

Stock de bactéries transformées

Pour réaliser des stocks glycérol de bactéries transformées, il faut mélanger 200 μ l des bactéries transformées à 800 μ l de glycérol dans un cryotube sous environnement stérile. Les bactéries sont stockées à -80°C.

II. 2. Transfections

Les transfections ont été réalisées avec du Fugen HD (Promega, Cat. N° E 2 312). Pour cela, les cellules sont posées dans les boites la veille de jour de transfection à une concentration suffisante pour être à 70% de confluence lors de la transfection.

Le jour de la transfection, le Fugen HD ainsi que les plasmides sont placés à température ambiante pendant une heure. La quantité d'ADN plasmidique utilisé est fonction de la quantité de cellules à transfecter (

Figure 33, p77). Dans un tube eppendorf, on mélange l'ADN plasmidique à du DMEM ne contenant ni sérum ni antibiotiques, on vortexe puis on centrifuge brièvement le mélange. On ajoute le Fugen HD, on vortexe puis on centrifuge brièvement le mélange que l'on laisse reposer 1h à température ambiante, le temps pour les micelles de Fugen-ADN plasmidique se forment. Le mix sera ensuite ajouté au milieu de culture des cellules pendant 6h et les cellules sont mises à proliférer à 37° C et à 5% de CO₂.

contenant	Quantité	Quantité	ug d'ADN	Ratio 8 :2
	totale de	milieu de		Fugene : DNA
	milieu	transfection		
24-well plate	0.5 mL	25 uL	5	2 uL
12-well plate	1.0 mL	50 uL	1	4 uL
6-well plate	2.0 mL	100 uL	2	8 uL
60-mm dish	5.0 mL	250 uL	5	20 uL
100-mm dish	8.0 mL	400 uL	8	32 uL

Figure 33 : Protocole de transfection au Fugen HD.

Au bout de 6 heures, on changera le milieu de culture par du milieu frais supplémenté en sérum et en antibiotiques et les cellules sont mises à proliférer à 37°C et à

5% de CO_2 pendant 18h supplémentaires si elles sont traitées 24h ou bien à 42h supplémentaires si elles ne sont pas traitées.

II.3. Extractions des ARN totaux

Les cellules sont lavées au Dulbecco's Phosphate Buffered Saline ou DPBS (Sigma-Aldricht, Cat N° D8537) puis lysées à l'aide de buffer RLT fourni dans le kit RNesay mini (Quiagen, Cat N° 74106) supplémenté en β -Mercaptoethanol. Le β -Mercaptoethanol est utilisé pour éviter la dégradation des ARN par d'éventuelles RNases. Le lysat cellulaire est ensuite collecté à l'aide d'un grattoir à cellules. L'ARN est complété par 1 volume d'éthanol à 70% pour obtenir un mélange bien homogène et passé sur colonne RNeasy spin. Les colonnes sont tout d'abord lavées avec le buffer RW1 puis 2 fois avec le buffer RPE avant l'élution des ARN dans 30 à 50µL d'eau RNase-free.

La concentration de la solution d'ARN sera évaluée sur nanodrop. On peut aussi analyser la qualité des ARN à l'aide du bio-analyseur Agilent 2100 Bioanalyzer. La qualité des ARN extraits est évaluée grâce au rapport d'absorbance 260/280 nm qui doit être compris entre 1.8 et 2.1. Si ce rapport est inférieur à 1.8, l'échantillon peut avoir une contamination partielle par des protéines. S'il est supérieur à 2.1, l'ARN peut être en partie dégradé ou il peut contenir un excès de nucléotide libre.

Les ARN sont stockés -20°C pour une durée maximale de 2 mois et à -80°C pour plus longtemps.

II.4. Analyse micro-arrays

Les ARN totaux extraits vont ensuite subir une étape de marquage qui peut se présenter sous 2 formes : le marquage avec retro-transcription seule (RT), avec le kit Applied Biosystems Chemiluminescent RT Labeling Kit (4340415), ou suivi d'une étape de transcription in vitro (RT-IVT), avec le kit NanoAmp[™] RT-IVT Labeling Kit (4365715). Ces deux types de marquages donnent au final deux produits différents : de

l'ADN complémentaire ou de l'ARNm complémentaire analysables par la biopuce. Le marquage proprement dit a consisté à remplacer une partie des uridines de l'ARN messager d'origine par des nucléosides possédant un groupement digoxygenin, digoxygenin-11-UTP (Roche, 11209256910) pour le marquage en RT-IVT ou digoxygenin-11-dUTP (11093088910) pour le marquage RT.

Lors de mon projet, j'ai utilisé la technique RT-IVT pour amplifier plus fortement les ARN et obtenir des signaux détectables par la caméra Applied Biosystems 1700 Chemiluminescent Microarray Analyzer.

Ensuite, j'ai utilisé la technique RT-IVT pour amplifier plus fortement les 2 ug d'ARN en ARNc marqués à la digoxygenin (ou en ADNc en cas de simple rétrotranscription) et obtenir des signaux détectables par la caméra Applied Biosystems 1700 Chemiluminescent Microarray Analyzer.

Dans mes expériences, les ARNc marqués ont été hybridés sur les puces Applied Biosystems Human Genome Survey Microarray (Cat N° PN 4359029). J'ai par ailleurs utilisé des puces Applied Biosystems Mouse Genome Survey Microarray (Cat N° PN 4382672) et Applied Biosystems Rat Genome Survey Microarray (Cat N° PN 4357832) pour des projets collaboratifs.

Ces puces sont composées de plus de 32 000 sondes, constituées de polymères de 60 nucléotides (60mers) complémentaires de régions spécifiques aux gènes recherchés. L'hybridation et la détection des puces ont été réalisées avec le kit Applied Biosystems Chemiluminescence Detection Kit (Cat N° PN 4342142) en suivant le protocole fourni (Cat N° 4338727) que je vais vous présenter rapidement ci-après.

Les micropuces sont hybridées avec 10 µg d'ARNc-DIG-UTP pendant 16 h sous agitation (100 rpm) et à 55°C, avant d'être lavées de nombreuses fois selon le protocole du kit Applied Biosystems Chemiluminescence Detection Kit (Cat N° 338727). Les micropuces sont ensuite incubées pendant 20 minutes avec la solution contenant l'anticorps anti-digoxigenin-AP puis de nouveau lavées de nombreuses fois. On procède ensuite à la réaction de chemiluminescence, puis au dernier lavage qui sert aussi à

empêcher les micropuces de sécher en attendant la lecture à l'aide de la caméra Applied Biosystems 1700 Chemiluminescent Microarray Analyzer, qui est une caméra numérique de très haute définition, refroidie à -50°C.

La réaction de luminescence proprement dite implique une déphosphorylation du Chemiluminescence Substrate par la phosphatase alcaline présente sur l'anticorps. Pour la lecture d'une puce, on scanne le code barre de la puce, on sélectionne le type de micropuces (humain, souris ou rat) adéquat, le type de marquage (RT ou RT coupée à une IVT), on nomme l'échantillon et on lance la lecture.

Les images sont ensuite analysées par le programme Applied Biosystems 1700 Chemiluminescent Microarray Analyzer. Le programme effectue une correction du bruit de fond, une intégration et normalisation des traits caractéristiques, un balisage des sondes problématiques, l'élimination des pixels artéfactuels et une estimation de l'incertitude du signal et du bruit de fond au niveau des pixels et des signaux afin de déterminer un signal normalisé.

Lors de l'export des données sous format .txt appelés raw data, on choisit les éléments nécessaires à l'analyse secondaire. Le plus souvent sont exportés le signal normalisé par la médiane des signaux de la puce, le coefficient de variance, calculée pour chaque probe à partir des quatre différentes images d'une même probe, et le Flag qui donne une valeur en fonction de différents critères pour les différents problèmes de la probe.

II.5. Analyse des données avec ace map

Les données brutes exportées depuis la plate-forme ont été analysées avec le logiciel ace.map. Les analyses nécessitent l'entrée des informations sur les données et la création de fichiers ma0, utilisés pour le reste de l'analyse. Les fichiers ma0 peuvent ensuite être utilisés pour mesurer des différences d'expression entre différentes conditions biologiques par l'intermédiaire de soustractions qui donnent les fichiers ma2. Les fichiers ma2 sont ensuite filtrés selon différentes critères et les listes de sondes filtrées sont contenues dans des fichiers ma3. Les analyses par ontologies des données

transcriptomiques ont été réalisées en utilisant les annotations fonctionnelles de chaque gène décrites dans Gene Ontology et PANTHER. Les sondes considérées comme significativement différentiellement exprimées entre deux conditions sont analysées en terme de voie de régulation, de processus biologique et de fonction moléculaire. La surreprésentation d'une ontologie dans un des groupes de sondes analysées est mesurée en terme de p-value, calculée par une loi binomiale et corrigée pour tenir compte de la dépendance entre les ontologies. Toutes les expériences ont été réalisées dans la mesure du possible en triplicats et la probabilité choisie et acceptée comme statistiquement significative a toujours été inférieure à $p \le 0,01$ ((Noth et al. 2005), (Noth et al. 2006), (Brysbaert et al. 2007), (Brysbaert et al. 2010), (Bécavin et al. 2011)).

II.6. Extraction des protéines cytoplasmiques et nucléaires de cellules transfectées.

Les cellules sont lavées avec du DPBS froid avant d'être trypsinisées. Le culot cellulaire obtenu après centrifugation des cellules est lavé au DPBST froid et est de nouveau centrifugé. Le culot cellulaire lavé est conservé dans la glace avant et pendant la lyse avec du tampon mono-détergent de lyse protéique.

L'extrait protéique est alors additionné de 20% de glycérol avant d'être congelé à -20°C ou à -80°C si on doit le conserver longtemps. L'extrait protéique peut être aussi directement dialysé pour être conservé en condition native.

II.7. Dialyse des extraits protéiques totaux

Préparation des tampons

La masse molaire de l'EDTA est de 292.2426 g.mol⁻¹. Pour préparer une solution stock d'EDTA (acide éthylène diamine tétracétique) concentrée à 0.5 M, il faut dissoudre 73.06 g dans 0.51 d'eau.

Pour préparer le tampon de dialyse 1X il faut préparer une solution contenant 0.1 M de potassium chloride ($K^+C\Gamma$), 20 mL d'Hepes 1 M et 0.2 mM EDTA, à laquelle on ajoute 200 mL de glycérol et la compléter à 11 avec de l'eau. On peut aussi préparer une solution de tampon de dialyse à 10X ne contenant pas de glycérol, solution à laquelle on ajoutera le glycérol et que l'on diluera 10 fois. Cette solution stock peut être conservée plusieurs mois à 4°C.

Dialyse

Les membranes de dialyse (Sigma-Aldricht, Cat. N° D9277) sont déionisées dans une solution d'EDTA à 0.05 M portée à ébullition pendant 10-15 minutes. Les membranes sont nouées à une extrémité et on y met les extraits protéiques avant de les refermer à l'autre extrémité. Les membranes sont placées dans 300 ml de tampon de dialyse 1X supplémenté en glycérol à 4°C et sous agitation pour une première dialyse de 45 minutes. On réalise ensuite deux autres dialyse de 45 minutes chacune pour être certain de bien éliminer tout le tampon de lyse. Les extraits protéiques dialysés sont ensuite conservé à -80°C.

II.8. Immunoprécipitation des protéines taggées

Préparation des tampons

Le kit d'immunoprécipitation des protéines taggées HA (Sigma-Aldrich, Cat N° IP0010) fourni les colonnes de purification, les billes d'agarose couplées à un anticorps anti-HA ainsi que le tampon d'immunoprécipitation (tampon IP) 10X. Il convient donc de préparer du tampon IP 1X et 0.1X en diluant le tampon 10X dans de l'eau ultrapure. Pour la purification anti-Flag, je n'ai pas utilisé les colonnes fournies mais uniquement l'anticorps anti-Flag couplé à des billes d'agarose (Sigma-Aldricht, Cat. N° A2220). J'ai utilisé le même tampon d'immunoprécipitation.

Protocole d'immunoprecipitation

Lors de l'immunoprécipitation anti HA, j'ai utilisé 10 uL de billes d'agarose pour 100 à 200 µl d'extrait protéique total, complété à 600 µl avec du tampon IP 1X. Lors de l'immunoprécipitation anti Flag, j'ai utilisé 15 µl de billes d'agarose couplé à l'anticorps anti-Flag préalablement lavées 3 fois avec le tampon d'immunoprécipitation IP 1X supplémenté avec du Tween 20 par centrifugation des billes à 5 000 g pendant 1 minute et à 4°C. 300 μ l de mélange 1 :1 extrait protéique : tampon d'immunoprécipitation IP 1X supplémenté avec du Tween 20 est ajouté aux billes et mis sous agitation head-over-tail à 4°C.

Le temps d'incubation varie de 1h à O/N ; pour ma part après avoir essayé divers temps d'incubation, j'ai opté pour un temps de réaction de 2 h afin de limiter les interactions non spécifiques.

Les billes sont ensuite centrifugées à 12 000 g pendant 15–30 secondes à 4°C pur éliminer le surnageant puis lavées avec 700 µl de tampon IP 1X 6 ou 7 fois pour éliminer les protéines non attachées aux billes d'agarose. Pour vérification j'ai conservé toutes ces surnageants. Le dernier lavage est réalisé avec du tampon IP à 0.1 X.

Lors de l'immunoprécipitation anti HA, j'ai tout d'abord tenté d'éluer les protéines HA à l'aide de 50 μ l de tampon d'élution contenant 0.1 M de glycine HCl, à pH 3.5 puis avec 50 μ l de tampon d'élution contenant 0.2 M Glycine, pH 2.2. Pour cela, on ajoute le tampon d'élution aux colonnes, on mélange gentiment les billes d'agarose dans le tampon, on laisse incuber 10 minutes à 4°C et on centrifuge les colonnes à 12 000 g pendant 15–30 secondes à 4°C. Le pH de chaque échantillon est ensuite tamponné par ajout de 1 μ l de 1.5 M Tris, à pH 9 par 20 μ l de tampon d'élution avant d'être stocké à -80°C. Enfin pour vérification, on réalise une dernière élution à l'aide de laemmli buffer 2X contenant du β -mercaptoethanol ; on laisse incuber 10 minutes à 4°C. Après vérification par western blot, il s'avère que seule l'élution en laemmli fonctionne.

Les protéines Flag sont éluées à l'aide à l'aide de 50 µl de tampon d'élution contenant 0.1 M de glycine HCl, à pH 3.5 selon le protocole précédent.

II. 9. Western blot

Préparation des gels d'acrylamide pour l'électrophorèse SDS-PAGE

Il est nécessaire de bien mélanger le gel avant de mettre l'APS et le Temed puis de couler rapidement le gel de migration. Pour niveler convenablement le niveau du gel de migration, déposer 200 μ l d'isopropanol à sa surface. Attendre 30 min que le gel polymérise complètement.

Le tampon de migration est constitué de Tris Glycine SDS (TGS) 1X (11 de TGS 10X). Le tampon de transfert est composé de 100 ml de TGS 10X auquel on ajoute 200 ml d'éthanol absolu et complété à 11 avec de l'eau. Le tampon de blocage est constitué de 5% de lait en poudre dissous dans du Phosphate Buffered Saline Tween-20 (PBST) soit 2.5 g de lait pour 50 ml de tampon de blocage.

Méthode

La concentration protéique des lysats cellulaires a été mesurée par la méthode de Bradford (BioRad, Cat. Number N° 500-0006) le kit est vendu avec le standard bovine serum albumin, (Cat. Number N° 500-0002). 50 mg d'extrait protéique total repris dans un volume total de 10 uL est mélangé à une solution de laemmeli contenant de l'urée 2X est dénaturée pendant 5 minutes à 95°C ou à température ambiante (pour la protéine CBP) avant d'être déposé sur un gel de poly-acrylamide de concentration choisie en fonction du poids moléculaire de la protéine d'intérêt pour être séparé par électrophorèse à 15 mA/gel dans du tampon de migration.

CBP (265 kDa) a été séparé dans un gel SDS contenant 8% d'acrylamide (SDS-PAGE) tandis que la protéine TDG (46 kDa) et la protéine RARα (51 kDA) ont été séparées sur un gel SDS contenant 10% d'acrylamide.

Les protéines sont ensuite transférées sur une membrane de nitrocellulose pendant 2 h à 175 mA et à 4°C ou O/N à 70 mA dans du tampon de transfert.

La membrane est alors incubée sous agitation pendant 30 minutes à température ambiante ou O/N à 4°C dans 10 ml de tampon de blocage, contenant du PBST et 5% de lait

en poudre. Cette étape permet de saturer la membrane en protéines non spécifiques et de limiter le bruit de fond. Après 3 lavages de 5 minutes au PBST, la membrane est incubée sous agitation 1 heure avec l'anticorps primaire. La membrane est alors lavée 2 fois pendant 5 minutes. La visualisation est par la suite permise grâce en incubant la membrane avec un anticorps secondaire lié à la peroxydase de raifort (horseradish peroxidase) utilisé en conjonction avec un agent luminescent l'ECL (Pierce, Cat. N° 32132X3) ; le produit de la réaction émet une luminescence proportionnelle à la concentration en protéine. La membrane est lavée 3 fois pendant 5 minutes avant d'être mise en présence d'ECL pendant 5 minutes en cassette. L'excès d'ECL est enlevé avant la révélation de la membrane.

II. 10. Microthermophorèse

Des extraits protéiques totaux de cellules HEK293 transfectées avec Y-CBP ou HA-TDG ainsi que de l'extrait de HEK293 transfectées avec Flag-RAR α ont été dialysé contre du tampon de dialyse 1X et la concentration des extraits protéiques ont été ajusté entre eux. J'ai réalisé des séries de 10 à 16 dilutions (1:1, 1:4; 1:8, etc.) des extraits protéiques de cellules HEK293 transfectées avec HA-TDG dans de l'extrait protéique de cellules HEK293 non transfectées (mocks) pour obtenir des concentrations décroissantes de TDG. Une quantité identique d'extrait de cellules HEK293 transfectées avec YFP-CBP a été ajouté à chaque dilution en présence ou non de l'extrait protéique FLAG-RAR α purifié par immunopurification ainsi qu'en présence ou non d'ATRA à 5.10⁻⁷ M ou d'éthanol. La réaction a lieu en condition native c'est-à-dire en présence de MgCl₂ à 20 mM.

II. 11. Cinétique d'activité glycosylase de TDG

<u>Matériels</u>

 Tampon de migration pour l'électrophorèse, TBE 10X (Tris, Borate (EDTA) 11 : 109 g de Tris, 55.6 g d'Acide Borique et 5.8 g EDTA dissous dans de l'eau. A autoclaver pendant 20 minutes. - DTT 10mM – 10 ml : 0,01545 g de DTT dissous dans 10 ml d'eau.

- Tampon pour la réaction de glysosylation 10 ml : 0.0596 g Hepes (25 mM), 50 μl EDTA à 0.2 M (1 mM), 1 mL DTT à 10 mM (1 mM) dissous dans 10 ml d'eau. Il faut ensuite ajuster le pH à 7.8 avec du KOH.
- Tampon charge 2X : 8 M d'urée (~4 ml), 20 mM EDTA, 250 µl de Tris dissous dans 10 ml d'eau. On ajuste le pH à 7,5 avec du HCl puis on ajoute 500 µl de glycérol.
- Marqueur de migration bleu : On mélange 1 ml tampon de charge 2X à 200 µl bleu
 6X.
- Préparation du gel acrylamide/urée à 18% 5ml : Pour préparer ce gel, dissoudre 2,4 g d'urée dans 1ml d'eau, 500 µl de TBE 10X et chauffer la préparation pour dissoudre l'urée. Ajouter 1,56 ml d'acrylamide 19:1 pour obtenir un gel à 18 % d'acrylamide. Ajouter tout de suite 8 µl d'APS 20% et 8 µl de TEMED, et couler immédiatement le gel.

Méthodes

- Annealing des oligos et dosage (à ne faire que la 1ère fois) : Pour le premier jeu d'expériences, nous avons utilisé des oligonucléotides à une concentration de 4.175 µg/ µl d'ADN double brin donc il faut essayer de se mettre à la même concentration à chaque nouveau lot d'oligonucléotides. Pour faire l'anneling des oligonucléotides simples brins en oligonucléotides doubles brins G/ T ou G/U, il faut mettre les 2 oligonucléotides simples brins pendant 10 min à 100°C dans un bain à sec puis on l'éteint et on laisse redescendre la température tout doucement.

- Réaction glycosylase : Nous avons utilisé 0.4 uL d'ADN double-brin à $4.175 \ \mu g/ \ \mu l$ et 0.5 μl de protéine TDG en présence ou non de 2.5 μl (5 équivalents) ou de 5 μl (10 équivalents) de protéine SUMO-1 pour un volume final de 20 μl par réaction c'est-à-dire pour 1 point du time-course, dans du tampon de glycosylation.

On incube les protéines avec les oligonucléotides à 37°C pendant 0, 5, 10, 30 et 60 minutes pour réaliser un time-course à 5 points.

Précipitation et hydrolyse de l'ADN : L'ADN est ensuite précipité par ajout de 2 μ l de NaCl à 3 M et de 60 μ l d'éthanol absolu froid pendant 1 heure à -80°C ou O/N à -20°C. Les échantillons sont ensuite centrifugés à 16 000 g, pendant 1h et à 4°C et le surnageant est éliminé très précautionneusement. Les culots d'ADN sont ensuite séchés à l'air libre ou à 37°C pendant 1h. Les culots d'ADN sont ensuite hydrolysés par ajout de 5 uL de NaOH à 0,01 N maximum à 50°C pendant 30 minutes. Il ne faut pas concentrer trop la solution de NaOH pour ne pas détruire la sonde fluorescente.

- Migration sur gel acrylamide : Pendant que la réaction d'hydrolyse de l'ADN se poursuit, le gel d'acrylamide à 18% est pré-migré à environ 100V pendant 30 minutes. Une fois l'hydrolyse des culots d'ADN totalement faite, on ajoute 5 µl de tampon de charge (2X), avant de vortexer et de centrifuger les culots pendant 2 minutes à 16 000 g. Les échantillons sont alors déposés sur le gel d'acrylamide pour être séparé par électrophorèse à 15 mA ou 100 V dans du tampon de migration jusqu'à mi-hauteur du gel. La visualisation du gel se fait immédiatement après la fin de la migration sous UV avec différents temps d'exposition.

Résultats

Lors de ma thèse, nous nous sommes posé différentes questions relatives au fonctionnement de TDG en tant qu'enzyme de réparation ainsi qu'en tant que facteur de transcription, seule ou au sein d'un complexe.

En effet, des études précédentes ont mises en évidence que TDG, découverte comme étant une enzyme appartenant à la voie de réparation de l'ADN Base Excision Repair ((Wiebauer et al. 1989), (Neddermann et al. 1993), (Hardeland et al. 2000)) possède d'autres fonctions que celle d'exciser les thymines et les uraciles lors des mésappariements guanine: thymine (G: T) et guanine: uracile dans les cellules. En effet, TDG peut interagir avec de différents facteurs de transcription dont ceux de la famille des récepteurs nucléaires tels que les récepteurs à l'acide rétinoïque RAR : RXR (Um et al. 1998) et les récepteurs aux œstrogènes (D. Chen et al. 2003) ou bien avec certains corégulateurs transcriptionnels tels que les histones acétyle transférases CBP/p300 (Tini et al. 2002) et SRC-1 (Lucey et al. 2005). Par ailleurs une étude récente montre que TDG possède une activité de régulation épigénétique de la transcription qui diffère de son activité de glycosylation. Lors de cette étude, des souris Tdg knockout (Tdg KO) présentaient une létalité embryonnaire précoce, aux alentours du jour embryonnaire 11.5 due à des hémorragies internes accompagnées de nécroses importantes ((Cortázar et al. 2011), (Cortellino et al. 2011)). Or ces défauts sont trop reproductibles d'un embryon à l'autre pour être seulement liés à un défaut de réparation des mésappariements G : U/T, mésappariements qui arrivent au hasard au cours de la vie des cellules.

Mon équipe s'est donc plus particulièrement intéressée à l'étude du complexe entre TDG et CBP/p300 et à l'étude de l'impact de ce complexe sur le transcriptome et l'épigénome.

Chapitre I _ TDG, un acteur de la réparation de l'ADN mais aussi un corégulateur transcriptionnel.

Lors de ma travaux, nous avons tout d'abord caractérisé la régulation de l'activité de glycosylase de TDG par son interaction covalente ou non, avec la protéine régulatrice SUMO-1 ((Hardeland et al. 2002), (Mohan et al. 2007)). Cette activité de réparation des mésappariements G : U/T est importante pour la survie des cellules et pour le développement des individus car elle permet de contrôler de façon indirecte la méthylation des cytosines. Or, comme je vous l'ai présenté en introduction, le contrôle de la méthylation des cytosines est un aspect essentiel de la régulation épigénétique de l'expression des gènes. Cette caractérisation s'est faite en collaboration avec le Dr. Smet-Nocca, aujourd'hui à l'Unité de glycobiologie structurelle et fonctionnelle (CNRS UMR 8576) de l'Université Lille 1. Cette étude a été publiée dans la revue BMC Biochemistry en février 2011 sous le titre « SUMO-1 regulates the conformational dynamics of thymine-DNA Glycosylase regulatory domain and competes with its DNA binding activity. »

Lors de cette étude, nous avons donc décrit les effets induits par l'interaction covalente et non-covalente de SUMO-1 sur la conformation dynamique de TDG, et plus particulièrement sur le changement de conformation du domaine de régulation (RD) et du domaine catalytique (CAT) de TDG ainsi que des régions N- et C-terminales de TDG. Nous avons aussi décrit la régulation de l'activité glycosylase de TDG via cette modification post-traductionnelle.

Pour cette étude, le Dr. Smet-Nocca a réalisé des mutants de sumoylation par mutagenèse dirigée et leur structure a été analysée en RMN et en spectroscopie dichroïque circulaire. Les résultats de cette étude ont été associés aux résultats fonctionnels obtenus par l'étude de cinétiques de glycosylation de TDG. Ces techniques ont aussi été utilisées pour la caractérisation d'autres mutants de TDG. Je vais donc vous présenter en introduction, les différents mutants créés et étudiés au sein du laboratoire que ce soit pour l'analyse de l'impact de SUMO-1 sur la structure et le fonctionnement de TDG que pour l'analyse transcriptomique que j'ai réalisée par la suite.

I.1. Mutagenèse dirigée envers la protéine TDG

Le mutant TDG-P65A présente une modification de la proline 65, située dans le domaine de régulation de TDG et permettant à TDG de modifier sa conformation active, en une alanine, avec la perte d'un point de rigidité à cet endroit. Je me suis par la suite servie de ce mutant pour étudier l'impact du complexe CBP/TDG sur la transcription des gènes dépendant des RARs comme je vous le présenterais au chapitre 2 des résultats.

Nous avons aussi muté les différents sites d'acétylation K59, K83, K84, K87 en remplaçant ces lysines par des alanines qui ne peuvent plus être reconnues par CBP. Ces mutations ont permis de créer respectivement les mutants TDG-K59A, TDG-K83/84/87A (TDG- Δ 3) et TDG-K59/83/84/87A (TDG- Δ 4).

Le mutant TDG-K330A présente le remplacement de la lysine 330, située au niveau du site de sumoylation de TDG, en une alanine ; ce mutant ne peut donc plus être sumoylé.

Les mutants TDG-D133A et TDG-E310Q sont deux mutants des deux sites potentiels de liaison à SUMO-1 ou SUMO Binding Motif (SBMs). Ces deux sites d'interaction avec SUMO-1 sont situés pour le SBM1 au niveau de la région N-terminal et le SBM2 au niveau de la région C-terminal du site catalytique de TDG. Le mutant TDG-D133A présente une mutation dans le site de liaison putatif SBM1 tandis que le mutant TDG-E310Q présente lui une mutation dans site putatif SBM2 (Figure 34, p91).



Figure 34 : Domaines de TDG (adapté de (Smet-Nocca et al. 2008))

I.2. Les fonctions de TDG sont-elles régulées par sumoylation

Les enzymes Thymine-DNA Glycosylases possèdent la capacité d'exciser les mésappariements G : U/T formés lors de la déamination spontanée ou non des bases cytosines non méthylées et méthylées. Ces enzymes scannent l'ADN à la recherche de ces mésappariements qu'elles reconnaissent grâce à leur site actif composé du domaine de régulation (RD) de l'activité BER et du domaine catalytique (CAT).

Le RD permet de réaliser la liaison spécifique à l'ADN au niveau des mésappariements G : U/T ainsi que la mise en place d'interactions intramoléculaires dynamiques avec le domaine CAT (Smet-Nocca et al. 2008) tandis que le domaine CAT est un domaine très conservé au cours de l'évolution et permettant de reconnaitre exclusivement les mésappariements G : U (Figure 17, p43). Les interactions RD / CAT sont modifiés quand TDG interagit avec son substrat, l'ADN, ce qui entraine une modification de l'activité glycosylase et du turnover enzymatique de TDG. Ces activités sont donc régulées par la conformation du RD de façon mismatch-dependent (Smet-Nocca et al. 2008).

En effet, de nombreuses études ont montré que TDG reconnait divers substrats en plus des bases thymine mal appariées, telles que les bases 3-N4-éthenocytosine, 5-fluorouracile, 5-methylcytosine (5-mC), pyrimidines, 5-fluorouracile (5-FU), 5-hydroxymethyluracile (HmU) ou bien 5-bromouracile (5-BU) (Hardeland et al. 2003) avec une affinité variable en fonction des substrats qu'elle rencontre. TDG présentera, par exemple, un turn-over plus lent si elle rencontre un mésappariement G : T ou G : U que si elle rencontre un mésappariement G : FU par exemple (Kunz et al. 2009). Ce résultat indique que le domaine catalytique de TDG n'est pas le seul domaine responsable de la spécificité de substrats et d'activité de TDG ; les domaines C- et N-terminaux de TDG sont impliqués aussi dans ces deux fonctions et permettent une plus grande spécificité de substrats et en contrepartie un turnover enzymatique plus faible que pour les autres UDG. Pour compenser ce faible turnover enzymatique, les cellules ont mises au point des voies de régulation de TDG et notamment la modification de sa conformation via des interactions covalentes ou non covalentes avec SUMO-1 (Smet-Nocca et al. 2008).

SUMO-1 est une petite protéine (97 acides aminés) appartenant à la famille des Small Ubiquitin-like Modifier ou SUMO proteins (Matunis et al. 1996); ce sont des protéines qui peuvent s'attacher de façon covalente (sumoylation) au niveau des résidus lysine ou interagir de façon non-covalente à d'autres protéines telles que TDG ou CBP et modifier ainsi leur activité. Ces deux mécanismes induisent des modifications posttraductionnelles impliquées dans la régulation de nombreux processus cellulaires et notamment dans la régulation de la transcription, de l'apoptose, du cycle cellulaire ainsi que dans la régulation de la stabilité des protéines. Par ailleurs, elles modifient la localisation cellulaire des protéines avec lesquelles elles interagissent en induisant notamment leur déplacement au niveau des compartiments nucléaires appelés promyelocytic leukemia protein (PML) oncogenic domains (PODs) ((Chalkiadaki et al. 2005), (Park et al. 2005)). Or les facteurs de transcription ainsi que les enzymes de réparation de l'ADN sont connus pour être associés à ces compartiments nucléaires ce qui suggère que les protéines SUMO sont impliquées dans la régulation de l'expression des gènes (Lin et al. 2001). I.2.a Impacts de l'interaction entre SUMO-1 et TDG sur la conformation et la dynamique de TDG

Lors de notre étude nous avons produit la protéine TDGwt et la protéine TDG-wt conjuguée à SUMO-1 au niveau de la lysine 330 (K330) grâce à des bactéries E. coli transformées avec le plasmide pGEX-TDGwt et au plasmide permettant la sumoylation de TDG pTE1/E2-SUMO-1 (Eilebrecht et al. 2010). Nous avons aussi produit un mutant de contrôle de cette liaison covalente, le mutant TDG-K330A, sous les mêmes conditions nous permettant ainsi d'avoir un contrôle négatif. Nous avons mis en évidence que la sumoylation de TDG par SUMO-1 dans ces conditions n'a lieu qu'au niveau de la lysine 330 (Eilebrecht et al. 2010) (Figure 34 p91).

Des études RMN précédentes de la protéine TDG non conjuguée à SUMO-1, avaient montré que seules les résonances issues des domaines N et C-terminaux (résidus 1-50 et 328-410 respectivement) de la protéine TDG pouvaient être détectées grâce à leur très grande flexibilité en solution ; les résonances du corps de la protéine n'étant pas détectables (Smet-Nocca et al. 2008).

Nous avons enregistré un spectre à deux dimensions HSQC 15N 1H d'une protéine uniformément marquée à l'azote 15N et à l'hydrogène 1H en absence et en présence de la protéine SUMO-1 non marqué. Ce type de spectre permet d'éditer le déplacement chimique de tous les protons liés à des atomes d'azote et d'hydrogène. Il constitue en quelque sorte l'empreinte de la protéine. Les déplacements chimiques des protons NH sont particulièrement sensibles à leur environnement (pH, température, proximité d'autres atomes dans l'espace). Ainsi, il est possible de visualiser la surface impliquée dans l'interaction grâce aux signaux modifiés par l'ajout de SUMO-1 (Calathea et al. 2008) (Figure 35, p94).



Figure 35 : Cartographie des variations de déplacements chimiques pour déterminer les régions d'interactions des protéines (Calathea et al. 2008).

A) Principe de la méthode : la protéine d'intérêt est uniformément marquée avec de l'azote 15N. Le spectre HSQC permet de corréler les déplacements chimiques protons et azote de chaque liaison NH peptidique. L'ajout du partenaire non marqué induit des variations de déplacements chimiques de certains signaux correspondant aux acides aminés en interaction avec celui-ci. Ces résidus sont reportés sur la structure de la protéine pour définir une région d'interaction.

B) Exemple d'application qui a permis de déterminer la région d'interaction d'un chaperon d'histones avec les histones.



Figure 36 : Spectre SQC ¹⁵N-¹H de TDG (en noir) et de TDG conjugué à SUMO-1 (en rouge) (Smet-Nocca et al. 2011).

Nous avons constaté que les résonnances associées à la région N-terminale de TDG (résidus 1-50) ne sont pas modifiées lors de la conjugation de TDG avec SUMO-1 au contraire de celles associées aux résidus 327-347, situés dans la région du site de sumoylation de la lysine 330 (Figure 36, p95). Ces résultats indiquent qu'il existe des modifications conformationnelles du domaine C-terminal de TDG lors de l'interaction covalente avec SUMO-1 tandis que le domaine N-terminal n'est pas impacté par cette interaction (Smet-Nocca et al. 2011).

Des études précédentes ont montré que SUMO-1 peut interagir avec TDG de façon non-covalente via deux sites de liaison distincts appelés respectivement SUMO-binding motif ou SBM1 (résidus 133-137) et SBM2 (résidus 308-311) situés tous les deux dans le domaine catalytique de TDG ((Baba et al. 2005), (Mohan et al. 2007), (Eilebrecht et al. 2010)). D'autres études avaient mis en évidence que les liaisons non-covalentes entre TDG et SUMO-1 pourraient être soit impliquées dans le processus de sumoylation de TDG luimême, en tant qu'états intermédiaires, soit dans les interactions fonctionnelles entre TDG et d'autres protéines sumoylées (Takahashi et al. 2005), (Mohan et al. 2007). Ces études avaient aussi montré que les interactions covalentes entre SUMO-1 et TDG permettaient de réduire la liaison de TDG à l'ADN tandis que les interactions entre TDG et l'ADN de même que la sumoylation de TDG empêchent la formation d'interactions intermoléculaires avec d'autres SUMO-1 (Mohan et al. 2007). Nous nous sommes particulièrement intéressés à l'étude des interactions intermoléculaires (non-covalentes) entre TDG et SUMO-1 via sa liaison au SBMs.

I.2.b Etude des SUMO-binding domains (SBMs)

Les études précédentes ayant montré l'existence de deux sites d'interaction avec SUMO-1 appelés SBMs, l'un situé au niveau de la région N-terminal et l'autre situé au niveau de la région C-terminal du site catalytique de TDG, nous voulions déterminer lequel de ces deux sites induisait les changements observés lors de notre étude RMN. Nous avons utilisé le mutant du site putatif SBM1, le mutant TDG-D133A, produit par l'équipe du Dr. Mohan (Mohan et al. 2007). Lors de leur étude, ils ont supposé que la perte de l'interaction TDG/SUMO-1 lors de la mutation de l'acide aspartique 133 en une alanine, induite la perte d'un SBM. Nous avons produit un mutant du SBM2, TDG-E310Q ainsi que le double mutant SBM1, SBM2, TDG-D133A/E310Q.

Lors de notre étude en RMN, nous avons montré que la mutation D133A située dans la séquence conservée DIVII du SBM1 entraine un repliement significatif de la protéine TDG et par conséquent son agrégation, conséquence que nous ne retrouvons pas pour la protéine TDG-wt ou pour le mutant TDG-E310Q, produites et étudiées dans les mêmes conditions (Figure 37, B et C, p97). Nous avons donc conclu que la caractérisation de l'impact de la mutation D133A de Mohan ne semble pas correcte. Nous pensons que la perte de l'interaction avec SUMO-1 observée lors de cette mutation est due à la perte du repliement de TDG et non à la perte d'un motif SBM.

Le mutant TDG-E310Q (SBM2) se comporte quant à lui comme la protéine TDG-wt autant lors de l'analyse en spectroscopie dichroïque circulaire (Figure 37 C, p97)



que lors de l'analyse RMN (Figure 37B, p97). Il semble donc que la mutation E310Q ne modifie pas trop fortement la conformation de TDG à l'inverse de la mutation D133A.

Figure 37 : Spectre HSQC ¹⁵N-¹H et spectre dichroïque circulaire de TDG-wt et des différents mutants (Smet-Nocca et al. 2011). (A, B) Spectre HSQC ¹⁵N-¹H et spectre dichroïque circulaire de TDG-wt (noir) et des différents

(A, B) Spectre HSQC ¹⁵N-¹H et spectre dichroïque circulaire de TDG-wt (noir) et des différents mutants (A) TDG-E310Q à 100 μ M (rouge) ou (B) TDG-D133A à 45 μ M (rouge) et TDG-D133A/E310Q à 50 μ M (bleu).

(C) Comparaison des spectres dichroïques circulaires TDG-wt (bleu), TDG-D133A (vert), TDG-E310Q (orange) and TDG-D133A/E310Q (rose).

(D) échantillons RMN des protéines marquées au ¹⁵N (1) TDG wild type (1 μl) et des mutants (2) TDG-E310Q (3 μl), (3) TDG-D133A (5 μl) et (4) TDG-D133A/E310Q (5 μl). (Smet-Nocca et al. 2011).

Nous nous sommes donc plus particulièrement intéressés à ce mutant et nous avons regardé les modifications de conformation du domaine C-terminal ou du domaine CAT de TDG-E310Q en présence de SUMO-1 (Smet-Nocca et al. 2011).



Figure 38 : Spectre HSQC ¹⁵N-¹H de TDG en présence ou en absence de SUMO-1 (Smet-Nocca et al. 2011).

A) Spectre HSQC ¹⁵N-¹H de TDG (noir) à 20 μM ou en présence de 120 μM of SUMO-1 (rouge). Les résonances de TDG-RD sont en gras et les résonances des résidus of the C-terminaux sont en italique.
(B) Spectre HSQC ¹⁵N-¹H de ¹⁵N-TDG-E310Q à 20 μM en absence de SUMO-1 (noir) ou en présence de 200 μM de SUMO-1 (rouge) et de ¹⁵N-SUMO-1 (vert).
(C) Merge des spectres HSQC ¹⁵N-¹H de ¹⁵N-TDG (noir) et de ¹⁵N-TDG-N (bleu) ce qui permet

(C) Merge des spectres HSQC ¹⁵N-¹H de ¹⁵N-TDG (noir) et de ¹⁵N-TDG-N (bleu) ce qui permet d'identifier les résonances des résidus de TDG-RD.

Les résultats obtenus (Figure 38, p98) montrent que la liaison de SUMO-1 à TDG est inhibée par la mutation E310Q (SBM2) et donc que la liaison de SUMO-1 au SBM situé en C-terminal (SBM2) est seule responsable des changements de conformations des domaines C et N-terminaux de TDG.

Nous avons constaté au cours de notre étude que la sumoylation de TDG agit plutôt sur la conformation du domaine C-terminal de TDG avec peu ou pas d'impact sur la conformation du TDG-RD tandis que l'interaction non-covalente de SUMO-1 à TDG via le SBM situé en C-terminal est en mesure de modifier structurellement les deux extrémités N- et C-terminales des régions de TDG et TDG sumoylée. En effet, nous avons constaté que SUMO-1 peut lier le SBM2 chez une protéine TDG déjà sumoylée (Figure 39, p99).



Figure 39 : Spectre HSQC ¹⁵N-¹H de TDG, de TDG sumoylée, TDG-E310Q et de TDG-E310Q sumoylée en présence ou absence de SUMO-1.

(A) Merge des spectres ¹⁵N-¹H de la protéine ¹⁵N-TDG (20 μ M) en présence de 200 μ M SUMO-1 (en noir), de la protéine sumoylée ¹⁵N-TDG (100 μ M) (en bleu) et de la protéine sumoylée ¹⁵N-TDG (20 μ M) en présence de 200 μ M SUMO-1 (en rouge). Les résonances des résidus C-terminaux sont en italique tandis que les résonances de TDG-RD sont en gras.

(B) Comparaison des spectres HSQC ¹⁵N-¹H du mutant ¹⁵N-TDG-E310Q sumoylée (en rouge) et non sumoylée (en noir) (100 μ M). Les résonances du domaine N-terminal de la protéine ¹⁵N-SUMO-1 sont indiquées par des flèches.

Nous avons de plus constaté que la liaison non covalente entraine un déplacement partiel du RD de TDG à partir de son interface CAT.

Ces résultats permettent de conclure que SUMO-1 n'adopte pas la même orientation lors de l'interaction non covalente que dans la protéine sumoylée, que cette liaison induit un déplacement partiel du RD de TDG à partir de son interface CAT entrainant un effet d'encombrement stérique au niveau du domaine CAT ; cette dynamique est en accord avec la localisation présumée de l'interface RD / CAT dans la protéine TDG (Smet-Nocca et al. 2008).

Nous nous sommes ensuite intéressés à l'impact de l'interaction covalente ou non entre SUMO-1 et TDG sur l'activité de cette dernière.

I.2.c Impacts des interactions covalentes ou non-covalentes sur l'activité glycosylase et le turnover de TDG

Des études précédentes ont montré que l'interaction intermoléculaire de SUMO-1 avec TDG est fortement diminuée par l'interaction de TDG avec son substrat, l'ADN (Mohan et al. 2007)? Qu'en est-il en présence de SUMO-1 ?

La présence de substrat d'ADN contenant une paire normale (G : C) ou une paire anormale (G : T ou G : U) peut déplacer de façon similaire le RD de TDG de son interface avec le domaine catalytique (Figure 40, p100), (**Erreur ! Source du renvoi introuvable.**, p101).



Figure 40 : SUMO-1 influence l'interaction entre TDG et son substrat, l'ADN (Smet-Nocca et al. 2011).
(A) Spectre HSQC ¹⁵N-¹H de TDG (20 μM) en absence de son substrat, l'ADN db (noir) ou en présence de 50 μM d'ADN db de 37-mer contenant un mésappariement G: T (rouge).
(B) Spectre HSQC ¹⁵N-¹H de TDG en présence de SUMO-1 (120 μM) et en absence d'ADN (noir) ou en présence d'ADN contenant un mésappariement G : T (50 μM) (rouge).



Les résonances associées aux résidus de TDG-RD sont notes en gras tandis que celles associées aux résidus du domaine C-terminal sont en italique.

Figure 41: Spectre HSQC ¹⁵N-¹H de TDG-wt et de TDG-N-terminal en présence de son substrat, l'ADN db contenant des dimères G: C, G: U ou G: T. Recouvrement des spectres HSQC ¹⁵N-¹H de TDG (20 μ M) (noir) et de TDG en présence de 37-mer d'ADN db (50 μ M) (rouge) contenant une paire G:C (A), une paire G:U (B) ou un mésappariement G:T (C). Le spectre du domaine N-terminal de TDG (TDG-N, résidus 1 à 111) en présence d'ADN db (bleu) (A) sert de référence.

Nos données montrent que les hétéroduplexes d'ADN contenant un mésappariement G : U ou un G : T entrainent une modification de TDG-RD de façon indépendante de la présence de SUMO-1 et empêche la liaison covalente de SUMO-1 au SBM2 (Figure 41, p101) ce qui rejoint nos premières constatations. Nous pouvons donc en conclure que la liaison à l'ADN via le domaine CAT de TDG modifie la conformation du SBM2 ou son accessibilité ce qui empêche toute interaction avec SUMO-1.



Figure 42 : Cinétique de l'activité glycosylase des différentes protéines TDG

(A, B) Cinétique de l'activité glycosylase de TDG sur des mésappariements G : T (A) ou G : U (B). La courbe de TDG-wt est en noir et en pointillés, celle de TDG sumoylées (carrés noirs) et de TDG- Δ N (losanges noirs) sont représentées par des lignes noires. La courbe de TDG en présence de 5 équivalents (carrés rouges) et 10 équivalents (carrés bleus) de SUMO-1 sont représentées par des lignes pointillées.

(C, D) comme pour (A, B) mais avec la protéine TDG-E310Q.

(E, F) Représentation graphique du turnover enzymatique de l'activité glycosylase de TDG au niveau des paires G: T (barres gris foncé) et G:U (barres gris clair) (E) et de l'activité glycosylase de TDG-E310Q (F). (G) Contrôle de l'intégrité des protéines avec un gel SDS-PAGE.

Nous avons observé une légère stimulation de l'activité glycosylase et du turnover de la protéine TDGwt et du mutant E310Q s'il y a sumoylation ou addition de SUMO-1 lors de la réaction de glycosylase des mésappariements G : T (Figure 42A, p102) alors même que la liaison intermoléculaire entre TDG et SUMO-1 n'est pas possible en présence d'ADN ou si il y a mutation du SBM terminal.

A l'opposé, l'activité G : U et le turnover enzymatique sont très sensibles à la sumoylation ou à l'addition de SUMO-1 pour les protéines TDG wild-type de façon dosedépendent (Figure 42B, p102). Cette réaction est spécifique à SUMO-1. En effet nous avons mimé l'ajout de SUMO-1 à l'expérience de glycosylase par l'ajout des mêmes quantités de BSA et nous n'observons pas ces résultats (Smet-Nocca et al. 2011).



Figure 43 : Cinétique de l'activité glycosylase des différentes protéines TDG en absence et en présence de SUMO-1 ou de BSA dans le tampon RMN (A, B) Mesure de l'activité réparatrice G: U de TDG (A) ou de TDG sumoylée (B) en présence ou en absence de 5 équivalents de SUMO-1 ou de BSA à pH 6.6. (C) Mesure de l'activité réparatrice G : T de TDG sumoylée en présence ou absence of 5 équivalents de SUMO-1 ou de BSA à pH 6.6.

Il semble donc que l'ajout de SUMO-1 permette d'augmenter la processivité G : T et G : U des protéines TDG sumoylées, même dans les conditions expérimentales de l'étude en RMN c'est-à-dire à pH 6.6 (Figure 43, p103). Ainsi l'interaction non-covalente avec SUMO-1, à l'instar de la sumoylation, agit positivement sur l'activité glycosylase G : U et améliore également bien que faiblement l'activité G : T, de façon dose-dépendante (Figure 44, p104).



Figure 44: Compétition entre le TDG-RD et SUMO-1 pour la liaison à l'ADN.

Spectre RMN bleu : TDG-N (20 μ M) en présence de 25-mer d'ADN db contenant un mésappariement G : T

Spectre RMN rouge : TDG-N (20 μM) en présence de 25-mer d'ADN db contenant un mésappariement G : T et d'un excès de SUMO-1 (80 μM)

Spectre RMN noir : 20 µM de TDG-N seule

Spectre RMN vert : 20 µM de TDG-N seule mis en présence de 80 d'ADN.

On observe un déplacement significatif du complexe RD/ADN par SUMO-1 au niveau des résidus E75, K78, S82, S85, S88 et S91.

L'absence de compétition au niveau des résidus Q55, A57, K64 and E69 indique qu'il y a une compétition partielle entre SUMO-1 et TDG-RD au niveau de la région 75-91.

Par ailleurs, nous avons étudié la capacité de SUMO-1 à interagir directement avec l'ADN et nous avons trouvé que SUMO-1 peut interagir de façon non spécifique avec l'ADN (Eilebrecht et al. 2010). Ce résultat explique l'existence d'une compétition entre SUMO-1 et TDG-RD pour la liaison à l'ADN. Cette compétition partielle entre SUMO-1 et TDG-RD pourrait donc déstabiliser suffisamment le complexe TDG/ADN avec, comme conséquence, une augmentation du turnover G : T / U.

Etant donné la faible affinité de TDG-N pour l'ADN (estimée à 100 à 500 μ M), une quantité substantielle de l'ADN libre (plus de 85%) doit se retrouve dans le mélange équimolaire TDG-N : ADN ce qui peut conduire à un grand nombre de complexes SUMO-1 : ADN non compétitifs. Dans le cadre de la protéine TDG entière, comme la présence d'un SBM favorise le recrutement de SUMO-1 et conduit à une augmentation significative de la concentration locale de SUMO-1 dans les environs de TDG-RD, la concurrence entre SUMO-1 et TDG-RD pourrait être plus prononcée. Nous avons montré qu'un tel mécanisme concurrentiel est effectivement réalisable (Eilebrecht et al. 2010).

I.2.d Impacts des interactions covalentes ou non-covalentes avec SUMO-1 sur l'interaction entre TDG et ses partenaires tels que CBP et RAR.

Comme je vous l'ai présenté en introduction, des études précédentes ont montré une association fonctionnelle entre TDG et le coactivateur transcriptionnel CBP ou CREBbinding protein (Tini et al. 2002). CBP et son proche homologue p300 sont des histones acétyle transférases qui permettent d'activer diverses voies régulées par des facteurs de transcription en contrôlant l'acétylation de la chromatine et des protéines associées à la chromatine, CBP/p300 peut donc être recrutée à l'ADN (Goodman et al. 2000). CBP a, par ailleurs, la possibilité de participer à la fois la régulation transcriptionnelle et la réparation de l'ADN en contrôlant l'activité de régulation de la transcription et de réparation de l'ADN de TDG via son acétylation. TDG est quant à lui un régulateur de la transcription dépendant de CBP/p300.

Nous nous sommes donc intéressés à l'impact des modifications posttraductionnelles de TDG par SUMO-1. Nous savons que la sumoylation de CBP entraine sa relocalisation aux PODs et l'activation de la transcription de ses gènes cibles. Par ailleurs, la sumoylation de TDG ou ses interactions non covalentes avec SUMO-1 permettent non seulement de réguler ses propriétés biochimiques (Hardeland et al. 2002) comme je vous l'ai présenté ci-dessus mais empêche sa relocalisation aux PODs où TDG peut interagir avec CBP (Mohan et al. 2007). Les PML oncogenic domains (PODs) sont des compartiments nucléaires riches en protéines Promyelocytic leukemia (PML) (Le et al. 1996), un facteur de transcription factor ayant une activité de suppresseur de tumeur ; ces protéines sont codées par le gène PML du chromosome 15. La translocation réciproque entre le gène PML et le gène du RARa entraine la formation de gènes chimériques RARca/PML et PML/RARa impliqués dans les leucémies promyélocytiques aigües (APL) (Borrow et al. 1990). Les protéines PML sont connues pour interagir avec le RARa (Zhong et al. 1999), avec CBP ((Zhong et al. 1999), (Matsuzaki et al. 2003), (Doucas et al. 1999)), avec TDG (Takahashi et al. 2005), avec SUMO-1 ((Kamitani et al. 1998), (Lin et al. 2001)) et avec de nombreuses autres protéines impliquées dans la régulation de l'expression des gènes.

Nous nous sommes donc intéressés à l'étude de l'impact de la sumoylation sur l'interaction CBP/TDG in vivo grâce à l'étude par imagerie confocale en FRET. Les cellules sur-exprimant la protéine CFP-CBP et la protéine YFP-TDG-wt présentent une légère diminution du temps de vie de fluorescence de CFP contrairement aux cellules sur-exprimant uniquement CFP-CBP (Figure 45, p107) (Spriet 2006). Ce résultat montre l'existence d'une interaction CBP/TDG *in vivo*.



Figure 45: Etude de l'effet de l'interaction CBP/TDG-wt (Spriet 2006). A) Distribution des temps de vie moyens mesurés sur chaque pixel des images de temps de vie de 30 cellules sur-exprimant CFP-CBP (bleu), CFP-CBP/YFP-TDG-wt (vert) et CFP-CBP/YFP-TDG-K330A (rouge).



J'ai confirmé ces résultats lors de ma thèse à l'aide d'une expérience de thermophorèse microscopique ou microthermophorèse (MST). La microthermophorèse est une technique basée sur les principes de la thermodynamique. La microthermophorèse est le mouvement dirigé de toute bioparticule placée dans un gradient de température microscopique obtenue grâce à des lasers (Figure 46, p108). Toute modification de la couche d'hydratation des biomolécules due à des changements de structure / conformation se traduit par une variation relative de mouvement le long du gradient de température mis en place dans le capillaire. Cette variation relative de mouvement est utilisée pour déterminer les affinités de liaison, la cinétique de liaison et la cinétique d'activité comme l'étude de la phosphorylation d'une protéine ou la liaison d'une biomolécule sur sa cible. La microthermophorèse permet de mesurer les interactions directement dans une solution. Elle a été développée par la société allemande, NanoTemper Technologies GmbH.


Figure 46: Principe de la thermophorèse microscopique (Jerabek-Willemsen 2011). A) Laser permettant d'induire un gradient de température à l'intérieur des capillaires de verre àù se trouve la molécule fluorescente étudiée et ses partenaires putatifs, dilués selon des dilutions en série de $\frac{1}{2}$.

B) Principe de la technique : on suit la fluorescence d'une molécule lors de sa diffusion dans un gradient de température du au laseret en présence de son/ses partenaires putatifs. Quand on allume le laser, on observe un saut de température et une diminution de la fluorescence de la molécule d'intérêt au cours du temps. Quand on éteint le laser, on retourne dans l'état de diffusion initial de la molécule fluorescente. Entre les deux, diffusion thermodynamique de la molécule fluorescente.

La fluorescence à l'intérieur d'un capillaire de verre est mesurée et normalisée à l'endroit chauffé en fonction du temps. Le laser infrarouge est allumé à l'instant t = 5 s et on observe un changement de fluorescence rapide due au saut de température puis une variation de la fluorescence en fonction de l'augmentation de la température dans le capillaire due au mouvement thermophorétique (échelle de temps ≈ 30 s) (NanoTemper) (Figure 46, p108).

Pour valider les résultats obtenus par l'analyse FRET du complexe CBP/TDG, j'ai donc réalisé une série d'expériences en microthermophorèse dans le modèle cellulaire Human Embryonic Kidney 293 ou HEK293. Pour cela j'ai transfecté des cellules avec soit le plasmide exprimant la protéine de fusion YFP-CBP, soit le plasmide exprimant la protéine de fusion HA-TDG-wt. Les extraits protéiques totaux YFP-CBP et HA-TDG-wt ont été dialysés contre du tampon de dialyse 1X que je vous ai décrit dans les matériels et méthodes. Les extraits totaux HA-TDG ont été dilué dans des extraits totaux de cellules non transfectées pour réaliser une gamme de concentration décroissante de la protéine HA-TDG-wt. Lors des expériences de microthermophorèse, une quantité choisie d'extrait protéique YFP-CBP est incubée, en présence de MgCL², avec chaque extrait de la gamme

de concentration HA-TDG-wt et placé par capillarité dans les capillaires de verre. Après tests, j'ai défini trois puissances du laser bleu à utiliser à chaque expérience de façon croissante : 20%, 40% et 60%. Les résultats obtenus à 40% de puissance du laser sont généralement les plus propres.



Figure 47: Microthermophorèse de CBP en fonction de doses croissantes de HA-TDG-wt. L'extrait cellulaire HA-TDG-wt a été dilué 10 fois dans de l'extrait cellulaire de cellules non transfectées pour obtenir une gamme de concentration. Cette dilution a été vérifiée par western-blot.

Cette expérience m'a permis de confirmer la présence d'un complexe CBP/TDG-wt dans des extraits cellulaires (Figure 47, p109).

Des expériences en FRET ont été réalisées avec le mutant de sumoylation TDG-K330A (Figure 48, p110 (Spriet 2006)) en parallèle de celles réalisées avec la forme sauvage de TDG.



Figure 48 : Etude de l'effet de la sumoylation sur l'interaction CBP/TDG (Spriet 2006) A) Distribution des temps de vie moyens mesurés sur chaque pixel des images de temps de vie de 30 cellules sur-exprimant CFP-CBP (bleu), CFP-CBP/YFP-TDG-wt (vert) et CFP-CBP/YFP-TDG-K330A (rouge). B et C) Proposition de modèle de l'interaction CBP/TDG.

On observe une légère diminution du temps de vie de fluorescence de CFP dans les cellules sur-exprimant la protéine CFP-CBP et la protéine YFP-TDG-K330A en comparaison avec les cellules sur-exprimant la protéine CFP-CBP et la protéine YFP-TDG-wt (Figure 48, p110).

Il y a donc augmentation de l'interaction CBP/TDG en absence de SUMO-1. La sumoylation de TDG est donc bien un moyen de contrôler la stabilité du complexe CBP/TDG. Ces résultats ont été complétés par une analyse transcriptomique du complexe CBP/TDG-wt et du complexe CBP/TDG-K330A réalisé au cours de ma thèse. Cette méthode d'étude de l'interaction entre CBP/TDG a été utilisée pour regarder l'impact de l'acétylation de TDG par CBP sur la stabilité du complexe CBP/TDG. Pour ce faire, les mutants d'acétylation TDG-K59A et TDG-delta4 ont été utilisés.



Figure 49 : Etude de l'effet de l'acétylation sur l'interaction CBP/TDG Distribution des temps de vie moyens mesurés sur chaque pixel des images de temps de vie de 30 cellules de chaque condition.

On n'observe une diminution du temps de vie de fluorescence de CFP dans les cellules sur-exprimant la protéine CFP-CBP et la protéine YFP-TDG-K59A en comparaison avec les cellules sur-exprimant la protéine CFP-CBP et la protéine YFP-TDG-wt (Figure 49, p111). Le temps de vie de la fluorescence étant inversement proportionnel à la force de l'interaction entre la protéine donneur (ici CFP-CBP) et la protéine accepteur (ici YFP-TDG), on peut donc en conclure qu'il y a augmentation de l'interaction CBP/TDG en absence du site d'acétylation K59A. On observe un résultat similaire avec le mutant TDG- Δ 3 (données non présentées). La mutation K59A et la mutation K83/84/87A (TDG- Δ 3) semblent donc stabiliser le complexe CBP/TDG (données non présentées).

On observe une augmentation du temps de vie de fluorescence de CFP dans les cellules sur-exprimant la protéine CFP-CBP et la protéine YFP-TDG- Δ 4 en comparaison avec les cellules sur-exprimant la protéine CFP-CBP et la protéine YFP-TDG-wt (données non présentées). Il y a donc diminution voire perte de l'interaction CBP/TDG en absence des quatre sites d'acétylation K59A, K83A, K84A et K87A.

L'acétylation de TDG par CBP est une autre voie de régulation de la stabilité du complexe CBP/TDG. Ces résultats ont eux aussi été complétés par une analyse transcriptomique du complexe CBP/TDG-wt au cours de ma thèse.

Nous avons donc démontré que la sumoylation empêche la formation du complexe CBP/TDG en modifiant la structure de TDG empêchant ainsi l'accès aux sites d'interaction avec CBP. Par ailleurs l'acétylation de TDG par CBP semble stabiliser le complexe CBP/TDG. Ces deux types de modifications sont donc deux mécanismes de régulation antinomiques de la stabilité du complexe CBP/TDG. Lors de ma thèse j'ai regardé l'impact de ces mutations sur l'activité transcriptomique du complexe CBP/TDG, résultats que je vous présenterais par la suite.

Chapitre II _ Le complexe tertiaire CBP/TDG/RARa peut-il expliquer, in vivo, l'activité régulatrice de TDG sur la transcription et notamment sur la transcription dépendante de l'acide rétinoïque ?

Des études précédentes à ma thèse ont orienté nos recherches vers l'étude de l'activité de corégulateur de la transcription de TDG au sein d'un complexe binaire tout d'abord, le complexe CBP/TDG puis tertiaire, le complexe CBP/TDG/RARα. TDG peut en effet, comme je vous l'ai présenté auparavant, interagir avec les facteurs de transcription RAR : RXR (Um et al. 1998), avec le corégulateur transcriptionnel CBP/p300 (Tini et al. 2002) ainsi qu'avec diverses autres protéines régulatrices de la transcription. De plus, des études précédentes ont montré que le phénotype des souris Tdg KO présentent un phénotype similaire à celui des embryons knock-out pour les facteurs de transcription RAR ou RXR (Mark et al. 2006), ou knock-out pour l'enzyme impliquée dans la synthèse de l'acide rétinoïque, retinaldehyde dehydrogenase (Vermot et al. 2003), ou knock-out Cbp ou p300-/_ ((Tanaka et al. 2000), (Yao et al. 1998)). Nous en déduisons que le phénotype de létalité est probablement liée à l'inactivation d'un stade pertinent du développement, de la transcription liée à la fonction de TDG (Cortellino et al. 2011).

Par ailleurs je vous ai présenté en première partie de mes résultats, l'effet de la sumoylation ou des interactions non covalentes avec SUMO-1 sur l'activité de réparation de l'ADN. Qu'en est-il de l'activité transcriptomique du complexe CBP/TDG/RAR ?

II. 1. Analyse du complexe CBP/TDG

Nous avons caractérisé le complexe CBP/TDG/RAR et son impact sur la transcription à l'aide d'une approche interdisciplinaire, mêlant des approches de biologie cellulaire et moléculaire et une approche transcriptomique à l'aide la plateforme Applied Biosystems 1700. J'ai donc utilisé des puces pangénomiques humaines couvrant environ 34 000 gènes et deux types de modèles cellulaires (cellules HEK et cellules COS-7) pour obtenir plus de 120 profils transcriptomiques. Pour cela j'ai transfecté les cellules avec les plasmides EYFP et YFP-CBP ou YFP-TDG-wt ou YFP-TDG-K330A ou YFP-TDG-Δ4 et

j'ai ou cotransfecté les cellules avec les plasmides YFP-CBP etYFP-TDG-wt ou YFP-TDG-K330A ou YFP-TDG- Δ 4 pendant 48h (paragraphe I.1 de la partie résultats, p90). Les ARNc issus de ces différentes conditions ont été hybridés sur des puces pangénomiques humaines de la technologie Applied Biosystems 1700 et analysés avec le logiciel ace.map.

I.1.a Effets de TDG et de CBP sur le transcriptome des cellules

TDG est connue comme un enzyme de réparation mais aussi comme un corégulateur de la transcription sous la dépendance de CBP.

Les gènes différentiellement exprimées entre les cellules Cos-7 sur-exprimant TDG-wt et les cellules Cos-7 non transfectées ont été sélectionnées sur une probabilité d'expression différentielle entre les conditions inférieure à 0,05 ($p\leq0,05$).

On observe 2 509 gènes ayant une probabilité d'expression différentielle p \leq 0,05 entre les cellules Cos-7 sur-exprimant TDG-wt et les cellules Cos-7 non transfectées. La majorité de ces gènes présentent un Log Q > 0 (Figure 50 A, p115). On constate donc que TDG a une action de corégulateur positif de la transcription des gènes des cellules COS-7.



Figure 50 : Effets de TDG-wt sur la transcription des gènes des cellules COS-7.

Comparaison entre des cellules COS-7 sur-exprimant TDG-wt et des cellules COS-7 non transfectées : A) Profil d'expression des gènes

B et C) Classification des gènes en ontologies de fonctions moléculaires

D) Classification des gènes en ontologies Pathways

Par ailleurs, j'ai classé les gènes statistiquement différentiellement exprimés entre les deux conditions en ontologies de fonction moléculaire et j'ai constaté que les gènes des fonctions moléculaires suivantes : Ribosomal protein, Nucleic acid binding, mRNA splicing factor, mRNA processing factor et Kinase sont sur-représentés dans la soustraction cellules Cos-7 sur-exprimant TDG-wt versus cellules Cos-7 non transfectées que ce qu'on observerait dans des conditions normales (Figure 50 B et C, p115). TDG-wt semble donc avoir un impact majeur sur la régulation des gènes notamment via la régulation de l'expression des ARN messagers et de leur épissage. On constate aussi que l'ontologie kinases présente plus de gènes activés observés que le nombre attendu tandis qu'elle présente moins de gènes réprimés que le nombre attendu. TDG-wt a donc un fort impact sur l'activation des kinases dans les cellules. Les kinases sont des enzymes appartenant au groupe des transférases. Elles sont souvent activées par des récepteurs membranaires et elles permettent la catalyse des réactions de phosphorylation par l'ajout d'un ion phosphate à une molécule cible, appelée le substrat ; ce substrat peut être une protéine, un lipide, un sucre ou encore une autre kinase. Cette réaction de phosphorylation permet de transmettre des signaux et donc d'induire une réponse appropriée de la cellule ; par exemple, les kinases de la famille des MAPK jouent un rôle dans la division, la croissance et la prolifération cellulaires.

J'ai aussi regardé les effets de CBP sur la transcription des gènes en comparant les cellules Cos-7 sur-exprimant CBP et des cellules Cos-7 non transfectées et en filtrant sur une probabilité d'expression différentielle inférieure ou égale à 0,05. J'ai obtenu 2 818 gènes en majorité activés (Figure 51A, p117). Ces gènes ont été aussi classés selon une ontologie en fonctions moléculaire. J'ai constaté que les gènes des fonctions moléculaires suivantes : Nucleic acid binding, Ribosomal protein, Isomerases et Histones sont sur-représentés (Figure 51B et C, p117). On constate aussi que l'ontologie Histone présente plus de gènes activés que le nombre attendu tandis que les ontologies Kinase et Transferase présentent plus de gènes réprimés que le nombre attendu. CBP semble donc avoir un impact majeur sur la régulation des gènes notamment via son rôle d'histone acétyl-transferase ; CBP semble aussi avoir un impact sur l'inactivation de certaines voies de signalisation dans les cellules.



Figure 51 : Effets de CBP sur la transcription des gènes des cellules COS-7. Comparaison entre des cellules COS-7 sur-exprimant CBP et des cellules COS-7 non transfectées : A) Profil d'expression des gènes, B et C) Classification des gènes en ontologies de fonctions moléculaires.

Mon analyse transcriptomique me confirme donc les rôles de régulateurs transcriptionnels de CBP et TDG-wt dans les cellules COS-7. Cette étude montre, en outre, la possibilité de transfecter des cellules issues de singes (les COS-7) avec des protéines humaines et d'hybrider les ARN issus de cette expérience sur des puces pangénomiques humaines. Par ailleurs, je me suis intéressée à l'impact de TDG sur la transcription des gènes dépendant de l'acide rétinoïque.

En effet, comme je vous l'ai présenté auparavant, TDG interagit avec RAR α de façon ligand-indépendant. Par ailleurs, une étude précédente a mis en évidence l'impact de TDG sur la différenciation des progéniteurs neuronaux dépendant à l'acide rétinoïque. J'ai donc observé dans mes analyses l'impact de TDG-wt sur la transcription des gènes dépendant de l'acide rétinoïque en cofiltrant la soustraction cellules COS-7 traitées à

l'acide rétinoïque versus cellules non traitées ($p \le 0.05$) avec la soustraction cellules COS-7 sur-exprimant TDG-wt versus cellules COS-7 non transfectées. J'ai obtenu 2 666 gènes dont la probabilité d'expression différentielle est inférieure ou égale à $p \le 0.05$ dans la soustraction cellules COS-7 traitées à l'acide rétinoïque en comparaison avec les cellules non traitées et qui sont présentes dans la soustraction cellules COS-7 sur-exprimant TDG-wt comparées aux cellules COS-7 non transfectées (Figure 52A, p118).



Figure 52 : Effets de l'acide rétinoïque sur les gènes dépendant de TDG-wt sur la transcription des gènes. A) Profil d'expression des gènes entre les soustractions « cellules COS-7 traitées à l'acide rétinoïque versus cellules non traitées » et « cellules COS-7 sur-exprimant TDG-wt versus cellules COS-7 non transfectées ».

B) Profils d'expression des gènes entre les soustractions « cellules HEK293 traitées à l'acide rétinoïque versus cellules non traitées » et « cellules HEK293 sur-exprimant TDG-wt et traitées à l'acide rétinoïque versus cellules HEK293 sur-exprimant TDG-wt et non traitées » et « cellules HEK293 sur-exprimant TDG-P65A et traitées à l'acide rétinoïque versus cellules HEK293 sur-exprimant TDG-P65A et non traitées ».

On observe un pattern d'expression similaire avec les cellules HEK293 lorsque l'on compare la soustraction cellules HEK293 traitées à l'acide rétinoïque versus cellules HEK293 non traitées à la soustraction cellules HEK293 sur-exprimant TDG-wt et traitées à l'acide rétinoïque versus cellules HEK293 sur-exprimant TDG-wt et non traitées et à la soustraction soustraction cellules HEK293 sur-exprimant TDG-P65A et traitées à l'acide rétinoïque versus cellules HEK293 sur-exprimant TDG-P65A et non traitées. Ces soustractions ont été filtrées sur une probabilité p \leq 0.05 (Figure 52B, p118). J'ai utilisé lors de cette analyse le mutant TDG-P65A dont je vous présenterais les caractéristiques spécifiques au paragraphe II. 2. b de la partie résultats (p124).

On constate que les gènes activés par le traitement à l'acide rétinoïque des cellules sont en majorités réprimés par la surexpression de TDG-wt dans ces mêmes cellules. Et réciproquement (Figure 52, p118). TDG semble donc avoir un impact négatif sur la transcription des gènes dépendant de l'acide rétinoïque dans nos deux modèles. Le mutant TDG-P65A a un impact opposé à la forme sauvage de TDG sur les gènes régulés par l'acide rétinoïque.

I.1.b Effets de CBP/TDG sur le transcriptome des cellules

Lors de mon analyse des effets du complexe CBP/TDG sur le transcriptome, nous nous attendions à voir les gènes activés dans la condition CBP + TDG-wt, fortement surexprimés dans la condition CBP + TDG-K330A et réprimés dans les conditions CBP + TDG- Δ 4 et CBP + TDG-P65A. En effet, comme je vous l'ai présenté précédemment (paragraphe I. 2. d de la partie résultats, p105), la mutation K330A stabilise le complexe CBP/TDG tandis que les mutations Δ 4 (K59/K83/84/87A) et P65A déstabilisent le complexe CBP/TDG pour différentes raisons.

Malheureusement je n'ai pas vu un tel pattern d'expression lors de mon analyse. En effet, on constate que les gènes surexprimés par CBP/TDG-wt le sont tout autant par la condition CBP/TDG-delta4 et CBP/TDG-P65A où le complexe CBP/TDG est déstabilisé (Figure 53, p120).



Figure 53: Effets de CBP en présence de TDG-wt et des mutants sur la transcription des gènes des cellules Cos-7. Profil d'expression des gènes entre les soustractions « cellules sur-exprimant CBP et TDG versus cellules non transfectées » pour les protéines TDG-wt, TDG-K330A, TDG-K59A, TDG-Δ4 et TDG-P65A.

Par ailleurs l'analyse en ontologies à l'aide de la classification panther et de la correction de Bonferroni appliqué par le logiciel ace.map n'ayant pas donné de résultats concluants, j'ai donc regardé de façon globale la fonction des gènes statistiquement différentiellement exprimés dans les différentes conditions ($p \le 0.05$) à l'aide d'un travail bibliographique. J'ai classé les gènes significativement différentiellement exprimés ($p\le 0.05$) selon 19 catégories : *mRNA transcription and regulation (TF), chromatin packaging and remodeling, nucleoside, nucleotide and nucleic acid metabolism, mRNA splicing and pre-mRNA processing, cell cycle, apoptosis, oncogenesis, cell proliferation and differentiation, receptors, signal transduction, transport, cell structure and motility, cell adhesion and cell communication, development, mesoderm development, ectoderm*

development, protein metabolism and modification, immunity and defense et *others* (Figure 54, p121).



Figure 54 : Analyse par classification en ontologie fonctions moléculaires des gènes régulés par le complexe CBP/TDG.

On remarque que les catégories *mRNA* transcription and regulation (TF), Nucleoside, nucleotide and nucleic acid metabolism, signal transduction et development sont fortement sur-représentées par rapport au % représentatif de cette catégorie attendu lors d'un titage aléatoire des gènes. Les catégories transport, protein metabolism and modification, others représentent chacune environ 15% des gènes différentiellement exprimés et les catégories Cell proliferation and differentiation, receptors, cell adhesion and cell communication et immunity and defense représentent elles environ 10% des gènes différentiellement exprimés.

La forte proportion de gènes impliqués dans le développement m'a incité à faire une sous-classification des gènes impliqués dans le développement (Figure 55, p122). Au cours de cette étude approfondie des gènes régulés par le complexe CBP/TDG, j'ai constaté un enrichissement en gènes issus de l'ectoderme et du mésoderme. Ces deux feuillets embryonnaires sont respectivement impliqués dans le développement et la fonction du système nerveux et dans le développement et la fonction du système musculaire et cardio-vasculaire. Les gènes développementaux régulés par le complexe CBP/TDG, dans mon analyse, sont pour la moitié d'entre eux des gènes impliqués dans le système nerveux (Figure 55, p112). Nos résultats rejoignent donc les résultats de Cortázar publiés en 2011 dans la revue Nature (Cortázar et al. 2011) de Cortellino publiés dans la revue Cell (Cortellino et al. 2011).



Figure 55 : Sous classification selon l'origine développementale des gènes issus de la catégorie des gènes développementaux (à gauche) et des facteurs de transcription (à droite).

Par ailleurs, j'ai regardé l'expression d'un petit groupe de gènes impliqués dans le développement du système nerveux dans mes analyses (Figure 56, p122).



Figure 56 : Gènes impliqués dans le développement du système nerveux. En rouge sont représentés les gènes activés par le complexe CBP/TDG tandis qu'en bleu sont représentés les gènes réprimés par le complexe CBP/TDG.

On constate que la majorité de ces gènes sont surexprimés (en rouge) avec un fold change dont le Log Q est supérieur ou égal à 1. Les gènes réprimés (en bleu) présentant un fold change dont le Log Q est supérieur à 1 sont peu nombreux. Ces gènes ci-dessus sont donc majoritairement surexprimés dans les cellules transfectées avec CBP/TDG que dans les cellules non transfectées. Ces résultats tendent vers un rôle du complexe CBP/TDG dans la régulation des gènes dépendant de l'acide rétinoïque (RA) ; en effet l'acide rétinoïque est un acteur de l'induction de la mise en place et du maintien du système nerveux au cours du développement (Rhinn et al. 2012) et chez l'adulte (Puttagunta et al. 2011). La perturbation de la signalisation de RA chez l'adulte induit une dégénérescence des neurones moteurs, le développement de la maladie d'Alzheimer et, éventuellement, le développement de la maladie de Parkinson (Maden et al. 2007).

II.2. Analyse du complexe tertiaire CBP/TDG/RARa

II.2.a Existence d'un complexe CBP/TDG/RARa

Les études précédentes ((Um et al. 1998), (Tini et al. 2002)) ont donc orienté nos recherches vers l'étude des gènes dépendant à l'acide rétinoïque régulés par le complexe CBP/TDG. Nous avons montré à l'aide d'une étude en FRET la présence d'une interaction entre le complexe CBP/TDG et les RAR. Pour cette étude, nous avons traité les cellules avec des doses croissantes d'acide rétinoïque (Figure 57, p124).

On constate une décroissance du temps de vie de fluorescence de la CFP dans les cellules sur-exprimant la protéine CFP-CBP et la protéine YFP-TDG-wt en présence de doses croissantes d'acide rétinoïque. L'acide rétinoïque est dissous dans de l'éthanol 100%. Pour éliminer tout impact de l'éthanol sur le temps de vie de fluorescence de la CFP, on a réalisé une mesure de la fluorescence émise par le complexe CBP/TDG-wt en présence d'éthanol absolu (EtOH).



Figure 57: Etude de l'effet de l'acide rétinoïque sur l'interaction CBP/TDG. A) Le temps de vie de la fluorescence de CFP a été mesuré en présence de doses croissantes d'acide rétinoïque all trans (ATRA), B) Comparaison des temps de vie de fluorescence de CFP-CBP/YFP-TDG-wt en présence d'ATRA et de CFP-CBP/YFP-TDG-K330A.

On constate que le traitement à l'éthanol ne modifie en rien la fluorescence du complexe CBP/TDG. L'acide rétinoïque semble donc stabiliser le complexe CBP/TDG de façon dose-dépendante à l'ATRA.

Lors de ma thèse, j'ai donc regardé *in vivo* la présence d'un complexe CBP/TDG/RARα. Pour cela, j'ai utilisé un nouveau modèle de cellules, les cellules Human Embryonic Kidney 293 ou HEK293 ; ces cellules appartiennent à une lignée cellulaire humaine présentant de nombreux avantages. Tout d'abord elles possèdent une croissance cellulaire très rapide ainsi qu'une très grande efficacité de transfection. La culture de ces cellules était très bien connue et maitrisée au laboratoire. Enfin des études ont mis en évidence certaines propriétés de neurones immatures dans des cultures de HEK293. Pour cela j'ai utilisé le mutant TDG-P65A caractérisé par le Dr. Smet-Nocca (paragraphe I.1 de la partie résultats, p90). Ce mutant m'a permis de disloquer le complexe CBP/TDG/RAR. Je vais vous présenter les caractéristiques de ce mutant.

II.2.b Le mutant TDG-P65A, un mutant de fonctionnalité

Le mutant TDG-P65A présente donc une mutation au niveau de la proline 65, proline située dans le domaine de régulation de TDG et proche des sites de liaison à CBP.

Les prolines sont des acides aminés qui servent aux organismes vivants à assembler les protéines. De plus les prolines ne peuvent former de liaison hydrogène du fait de son amine secondaire ce qui déstabilise les hélices α ou forme des bifurcations dans les feuillets β . Ces propriétés des prolines nous ont poussés à vérifier la structure du mutant TDG-P65A par RMN (Figure 58, p125).



Figure 58: Analyse RMN du mutant TDG-P65A : Résonnances de TDG-wt (en noir) et de TDG-P65A (en rouge). A gauche on observe la partie N-terminale des deux protéines tandis qu'à droite on observe la protéine entière.

On constate (Figure 58, p125) que les résonnances du mutant TDG-P65A se superposent presque parfaitement aux résonnances de la forme sauvage de TDG tant dans le contexte de la protéine entière que dans le contexte de la partie N-terminale de la protéine. Les quelques différences observées sont dues aux modifications de l'environnement électronique des acides aminés voisins de la proline 65. TDG-P65A possède une structure similaire à la forme sauvage de TDG. Il n'y a donc pas de changement conformationnel des domaines N et C-terminaux ni du domaine de régulation de TDG sous l'effet de la mutation P65A.



Figure 59 : Etude de l'activité glycosylase de la protéine TDG-P65A (losange bleu) comparée à celle de TDG-wt (carré rose), de TDG-RD-CAT (carré vert) c'est-à-dire d'une protéine très similaire à TDGwt et de TDG-N-CAT (carré rouge) c'est-à-dire ne présentant plus le domaine d'activité G : T A et B) Activité G : T et G : U ; C et D) activité G : T et G : U des protéines TDG-P65A (losange bleu), TDG-RD-CAT (carré vert) et TDG-N-CAT (carré rouge) comparées à TDG-wt ; E) Comparaison entre les activités G : T et G : U pour les protéines TDG-wt (carré rose), TDG-P65A (losange bleu), TDG-RD-CAT (carré vert) et TDG-N-CAT (carré rouge). La protéine TDG-P65A est plus active que la protéine wt, notamment sur la réparation G : U mais possède un turnover enzymatique similaire à celui de TDG-wt. La mutation P65A ne modifie donc pas le turnover de TDG et modifie peu son activité glycosylase (Figure 59, p126).

Par ailleurs, une première analyse en microscopie (Expériences de FRET) avait montré que la mutation P65A disloquait le complexe CBP/TDG même en présence de RA (données non présentées). Ces résultats ont été confirmés in vivo par des études en microthermophorèse dans des cellules HEK293.

Pour cela j'ai transfecté des cellules HEK293 avec soit le plasmide exprimant la protéine de fusion YFP-CBP, soit le plasmide exprimant la protéine de fusion HA-TDG-wt, soit le plasmide exprimant la protéine de fusion HA-TDG-P65A. Les extraits protéiques totaux YFP-CBP, HA-TDG-wt et HA-TDG-P65A ont été dialysés contre du tampon de dialyse 1X que je vous ai décrit au paragraphe II.8 de la partie matériels et méthodes. Les extraits totaux HA-TDG-wt et HA-TDG-P65A ont été dilués dans des extraits totaux de cellules non transfectées pour réaliser une gamme de concentration décroissante de la protéine HA-TDG-wt et de la protéine HA-TDG-P65A. Lors des expériences de microthermophorèse, une quantité choisie d'extrait protéique YFP-CBP est incubée, en présence de MgCL², avec chaque extrait de la gamme de concentration HA-TDG-wt ou HA-TDG-P65A.



Figure 60: Microthermophorèse de YFP-CBP en fonction de doses croissantes de HA-TDG et de HA-TDG-P65A. Comparaison entre la courbe d'affinité de la protéine HA-TDG-wt pour YFP-CBP (en noir) et celle de la protéine HA-TDG-P65A pour YFP-CBP (en gris).

On constate que dans les mêmes conditions de transfection, de préparation des extraits protéiques HA-TDG-wt et HA-TDG-P65A et de microthermophorèse, le complexe CBP/TDG-wt est plus stable (Kd relatif de dissociation plus fort) que le complexe CBP/TDG-P65A (Figure 60, p128). Le mutant TDG-P65A est donc un bon outil pour déstabiliser le complexe CBP/TDG et en étudier les effets. J'ai utilisé la technique de la microthermophorèse pour étudier la stabilité du complexe CBP/TDG en présence du RAR α activé ou non par l'acide rétinoïque. En effet, des études précédentes ayant montré que le RAR α pouvait interagir avec TDG et avec CBP respectivement de façon ligand-indépendant et ligand-dépendent, il semblait intéressant d'étudier la stabilité du complexe CBP/TDG en présence du RAR α activé ou non par l'acide rétinoïque.

Pour cela j'ai transfecté des cellules HEK293 avec soit le plasmide exprimant la protéine de fusion YFP-CBP, soit le plasmide exprimant la protéine de fusion HA-TDG-wt, soit le plasmide exprimant la protéine de fusion HA-TDG-P65A, ou le plasmide exprimant la protéine de fusion Flag-RARα. Les extraits protéiques totaux Flag-RARα ont été purifiés par immunoprécipitation contre le tag Flag. Les extraits protéiques totaux YFP-

CBP, HA-TDG-wt, HA-TDG-P65A et les extraits purifiés Flag-RAR α ont été dialysés contre du tampon de dialyse 1X que je vous ai décrit au paragraphe II.7 de la partie matériels et méthodes (p81). Les extraits totaux HA-TDG-wt et HA-TDG-P65A ont été dilués dans des extraits totaux de cellules non transfectées pour réaliser une gamme de concentration décroissante de la protéine HA-TDG-wt et de la protéine HA-TDG-P65A. Lors des expériences de microthermophorèse, une quantité choisie d'extrait protéique YFP-CBP est incubée, en présence de MgCL², avec chaque extrait de la gamme de concentration HA-TDG-wt ou de la gamme de concentration HA-TDG-P65A en présence ou en absence d'extrait purifié Flag-RAR α activé ou non à l'aide d'acide rétinoïque all trans (ATRA) à 5.10⁻⁷ M. On sait que le mutant TDG-P65A a permis de casser le complexe CBP/TDG. On sait aussi que RAR α doit être activé par la liaison de son ligand, l'ATRA, pour pouvoir interagir avec CBP. On pose comme hypothèse que

- (i) dans la condition CBP/TDG-wt/RARα activé par l'ATRA (RARα*), on trouve dans le capillaire (Figure 61, p129) :
 - YFP-CBP, lié avec HA-TDG-wt et avec F-RARα* lui-même lié à HA-TDGwt ; ce complexe présente une vitesse de migration lente de par sa taille
 - YFP-CBP, lié uniquement avec F-RARα* ou avec HA-TDG-wt et donc qui présente une vitesse de migration légèrement plus rapide
 - YFP-CBP, HA-TDG-wt et F-RARα* qui n'ont pas réussi à se lier à YFP-CBP et qui présentent donc la vitesse de migration la plus rapide.
 - HA-TDG-wt lié à F-RARα* mais ces interactions ne jouent pas sur la décroissance de fluorescence de YFP-CBP.



Figure 61: Schéma des complexes protéiques présents dans les capillaires lors de l'étude du complexe YFP-CBP/HA-TDG-wt/flag-RARα en présence de 5.10⁻⁷ M d'ATRA. La fluorescence est émise, après activation par un laser bleu, par la protéine CBP taggée avec le fluorophore YFP.

- (ii) dans la condition CBP/TDG-P65A/RARα*, il y a dans le capillaire (Figure 62, p130):
 - YFP-CBP, lié avec F-RARα* lui-même lié avec HA-TDG-P65A, présente une vitesse de migration lente ;
 - YFP-CBP, lié uniquement avec RARα*, présente une vitesse de migration plus rapide que dans le cas précédent ;
 - YFP-CBP, HA-TDG-P65A et RARα* libres dans la solution. YFP-CBP présente dans ce contexte la vitesse de migration la plus rapide.
 - HA-TDG-P65A lié à F-RARα mais ces interactions ne jouent pas sur la décroissance de fluorescence de YFP-CBP.



Figure 62: Schéma des complexes protéiques présents dans les capillaires lors de l'étude du complexe YFP-CBP/HA-TDG-P65A/flag-RARα en présence de 5.10⁻⁷ M d'ATRA. La fluorescence est émise, après activation par un laser bleu, par la protéine CBP taggée avec le fluorophore YFP.

- (iii) dans la condition CBP/TDG-wt/RARα, il y a donc en solution dans le capillaire(Figure 63, p131) :
 - YFP-CBP, lié avec HA-TDG-wt lui-même lié à F-RARα, présente une vitesse de migration lente ;
 - YFP-CBP, lié uniquement avec HA-TDG-wt et qui présente donc une vitesse de migration légèrement plus rapide;
 - YFP-CBP, HA-TDG-wt et F- libres dans la solution. YFP-CBP présente la vitesse de migration la plus rapide.

- HA-TDG-wt lié à F-RARα mais ces interactions ne jouent pas sur la décroissance de fluorescence de YFP-CBP.



Figure 63: Schéma des complexes protéiques présents dans les capillaires lors de l'étude du complexe YFP-CBP/HA-TDG-wt/flag-RARa. La fluorescence est émise, après activation par un laser bleu, par la protéine CBP taguée avec le fluorophore YFP.

- (iv) dans la condition CBP/TDG-P65A/RARα, il y a en solution dans le capillaire (Figure 64, p131) :
 - YFP-CBP est ni lié à HA-TDG-P65A ni à F-RARα et présente donc une vitesse de migration rapide ;
 - HA-TDG-P65A est lié ou non à F-RARα mais ces interaction ne jouent pas sur la décroissance de fluorescence de YFP-CBP.



Figure 64: Schéma des complexes protéiques présents dans les capillaires lors de l'étude du complexe YFP-CBP/HA-TDG-P65A/flag-RARα. La fluorescence est émise, après activation par un laser bleu, par la protéine CBP taggée avec le fluorophore YFP.

Dans les mêmes conditions de microthermophorèse, le complexe CBP/TDGwt/RAR α activé par l'ajout de 5.10⁻⁷ M d'ATRA est plus stable que le complexe CBP/TDG-P65A/RAR α activé par l'ATRA tandis que le CBP/TDG-wt/RAR α en absence d'ATRA est plus stable que le complexe CBP/TDG-wt et que les complexes CBP/TDG-P65A/RAR α en absence d'ATRA (RAR α) et CBP/TDG-P65A (Figure 65, p132).



Figure 65: Microthermophorèse de YFP-CBP en fonction de doses croissantes de HA-TDG et de HA-TDG-P65A et d'une concentraion fixe de flag-RARα activé ou non par l'ajout d'atRA à 5.10⁻⁷ M. Comparaison entre la courbe d'affinité de la protéine HA-TDG-wt pour YFP-CBP (A) et celle de la protéine HA-TDG-P65A pour YFP-CBP (B). En noir, complexe YFP-CBP/HA-TDG, en vert, complexe YFP-CBP/HA-TDG en présence de RARα et en présence de RARα activé par l'ajout d'ATRA (en rouge)

TDG-P65A permet donc de déstabiliser le complexe CBP/TDG/RARα de deux façons différentes, en fonctions de la présence ou de l'absence d'ATRA.

Par ailleurs, des expériences d'activation Gal4 de la protéine RARα par la forme sauvage ou les formes mutées de TDG ont été réalisées.



Figure 66 : Expériences de transactivation Gal4 du RARα par TDG-wt, TDG-K330A et TDG-P65A en absence (noir) ou en présence d'acide rétinoïque all trans (orange). On suit ici l'activité d'un gène rapporteur, ici la luciférase, activé lorsque le gène d'intérêt, ici RARα, est exprimé.

La condition de contrôle Gal4-RAR α présente, en présence d'ATRA, une activation basale (RAR α). En présence de HA-TDG-wt, l'activation de Gal4-RAR α est augmentée d'un facteur 2.1 par rapport à l'activation basale de Gal4-RAR α en présence d'ATRA. Les mutants HA-TDG-K330A et HA-TDG-P65A présentent une augmentation plus faible de l'activation de Gal4-RAR α que la protéine TDG sauvage. Malgré cette diminution de l'activation de Gal4-RAR α , ils peuvent tout de même activer Gal4-RAR α en présence d'ATRA de l'ordre d'un facteur 1.6 et 1.5 respectivement (Figure 66, p133). TDG-K330A et TDG-P65A ont conservé la capacité à interagir et activer le RAR α .

Les résultats de l'analyse en RMN de la structure du mutant TDG-P65A, des expériences de cinétique de l'activité glycosylase de ce mutant de TDG ainsi que ces résultats ont donc orienté mon choix vers ce mutant pour l'analyse transcriptomique du complexe CBP/TDG/RARa.

II.2.c Analyse transcriptomique du complexe CBP/TDG/RARa

Pour mon analyse transcriptomique de l'impact complexe CBP/TDG en présence d'ATRA dans les cellules HEK293, j'ai cotransfecté les cellules avec les plasmides YFP-CBP et YFP-TDG-wt ou YFP-TDGP65A pendant 24 h avant de traiter la moitié des conditions de transfections avec 5.10⁻⁷ M d'atRA. Les ARNc issus de ces différentes conditions ont été hybridés sur des puces pangénomiques humaines de la technologie Applied Biosystems 1700 et analysé avec le logiciel ace.map comme lors de mon premier transcriptome. J'ai tout d'abord regardé l'expression des gènes régulés par le traitement des cellules avec 5.10⁻⁷ M d'ATRA (Figure 67A, p135). Le traitement semble activer majoritairement les gènes des cellules HEK293 (p≤0.005. J'ai ensuite conservé cette liste de gènes et j'ai regardé leur expression lorsque les cellules ont été cotransfectées avec CBP et TDG-wt ou TDG-P65A. J'ai représenté ces résultats par des scatter plots dans lesquels l'axe des X représente l'expression des gènes dans la condition « cellules traitées à l'ATRA » et l'axe des Y représente l'expression des gènes dans la condition « cellules cotransfectées avec CBP et TDG-wt » ((Figure 67B, p135) ou dans la condition « cellules cotransfectées avec CBP et TDG-P65A » ((Figure 67 C, p128). Enfin la Figure 67 C (Figure 67C, p135) représente la superposition des deux scatter plots précédents. J'ai par ailleurs, représenté es résultats par des histogrammes (Figure 68, p136).



Figure 67: impact des complexes CBP/TDG sur la régulation des gènes dépendant de l'ATRA dans les cellules HEK293.

A) profil transcriptomique des gènes statistiquement (p≤ 0.005) régulés par 5.10⁻⁷ M d'ATRA

B-C) Scatter plot entre les gènes régules par l'ATRA et ceux régules par le complexe CBP/TDGwt (noir) ou par CBP/TDG-P65A (rouge) à $p \le 0.005$.

Le log2-fold change de l'expression des gènes statistiquement différentiellement exprimés entre la condition "cellules HEK293 sur-exprimant CBP et TDG traitées à l'ATRA" et "cellules HEK293 sur-exprimant CBP and TDG" (y-axis) est comparé au log2-fold change de l'expression des gènes statistiquement différentiellement exprimés entre la condition "cellules traitées avec 5.10⁻⁷ M ATRA" et "cellules non traitées" (x-axis). Les deux conditions de surexpression CBP/TDG-wt (B) et CBP/TDG-P65A (C) sont comparées à l'expression des gènes des cellules traitées à l'ATRA.

D) Overlap entre les deux scatter plots précédents.



Figure 68: Impact du complexe CBP/TDG sur la régulation des gènes dépendant à l'atRA dans les cellules HEK293.

A) Les gènes différentiellement exprimés entre les cellules traitées à l'ATRA et les cellules non traitées (en orange) ont étaient utilisé comme un ensemble de gènes à étudier. L'expression des gènes de ce ensemble a été étudiée dans la condition comparant les cellules sur-exprimant CBP et TDG-wt et traitées à l'ATRA aux cellules sur-exprimant CBP et TDG-wt (en noir);

A) L'expression de ce sous-ensemble de gènes a été étudiés dans les conditions « cellules surexprimant CBP et TDG-wt et traitées à l'ATRA versus cellules sur-exprimant CBP et TDG-wt » (en noir) et « cellules sur-exprimant CBP et TDG-P65A et traitées à l'ATRA versus cellules sur-exprimant CBP et TDG-P65A » (en rouge).

On constate que la surexpression de CBP et de TDG-wt, en présence d'ATRA semble éteindre ou du moins inverser l'expression des gènes sous la dépendance de l'acide rétinoïque tandis que la surexpression de CBP et de TDG-P65A, en présence d'ATRA, on observe un retour à la « normale» de la majorité des gènes sous la dépendance de l'acide rétinoïque (Figure 67, Figure 68, p135-136). La régulation de la transcription est donc affectée par la perte de l'interaction CBP/TDG.



Figure 69 : Impact de CBP, TDG et du complexe CBP/TDG sur la régulation des gènes dépendant à l'ATRA dans les cellules HEK293

A) Profils d'expression du sous-ensemble des gènes modifiés par l'ATRA (en jaune), et par l'ATRA et CBP (en noir) à $p \le 0.005$. B) Profils d'expression du sous-ensemble des gènes modifiés par l'ATRA (en jaune), et par TDG-wt (en noir) à $p \le 0.005$. C) Profils d'expression du sous-ensemble des gènes modifiés par l'ATRA (en jaune), et par TDG-P65A (en rouge foncé) à $p \le 0.005$. D) Profils d'expression du sous-ensemble des gènes modifiés par l'ATRA (en jaune), et par tDG-P65A (en jaune), et par le complexe CBP/TDG-wt (en bleu) à $p \le 0.005$. E) Profils d'expression du sous-ensemble des gènes modifiés par l'ATRA (en jaune), et par le complexe CBP/TDG-wt (en jaune), et par le complexe CBP/TDG-P65A (en rose) à $p \le 0.005$.

On constate que la surexpression de CBP, en présence d'ATRA, semble éteindre ou du moins inverser l'expression des gènes sous la dépendance de l'acide rétinoïque tout comme semble le faire la surexpression de TDG-wt, en présence d'ATRA. La surexpression de TDG-P65A, en présence d'ATRA, semble induire une forte « augmentation » de l'expression des gènes sous la dépendance de l'ATRA. On constate enfin que la surexpression de CBP et de TDG-wt, en présence d'ATRA, induit la « répression » de l'expression des gènes sous la dépendance de l'ATRA, induit la surexpression de CBP et de TDG-P65A, en présence d'ATRA, induit un pattern d'expression des gènes proche de celui observé lors de la surexpression de CBP, en présence d'ATRA (Figure 69, p137).

Or ce n'est pas ce à quoi je m'attendais à observer. Je m'attendais à voir les gènes dépendant à l'ATRA surexprimés par la présence du complexe CBP/TDG. La dislocation du complexe CBP/TDG via l'utilisation du mutant TDG-P65A permet donc un retour au pattern d'expression de la surexpression de CBP, en présence d'acide rétinoïque. La régulation de la transcription est donc affectée par la perte de l'interaction CBP/TDG (TDGP65A) mais aussi par les mutants de fonctionnalité TDG-K330A, TDG-K59A, TDG- Δ 3 et TDG- Δ 4. Il existe donc bien un complexe CBP/TDG fonctionnel *in vivo*, stabilisé par l'acide rétinoïque de façon dose-dépendante tandis que la régulation des gènes dépendant de l'ATRA est affectée par la perte de l'interaction CBP/TDG. Nos résultats mettent aussi en avant la possibilité d'un autre composant impliqué dans le complexe CBP/TDG/RAR.

Discussion

Il est intéressant de se demander pourquoi une protéine découverte comme étant une enzyme de réparation possède une activité de régulateur de la transcription. Je me suis donc posée différentes questions relatives au fonctionnement de TDG en tant qu'enzyme de réparation ainsi qu'en tant que facteur de transcription, seule ou au sein d'un complexe avec le facteur de transcription RAR α (Um et al. 1998) et le cofacteur transcriptionnel CBP/p300 (Tini et al. 2002).

Nous avons montré que l'activité réparatrice et transcriptionnelle de TDG peuvent être régulées de différentes manières. Tout d'abord TDG interagir de façon covalente et non-covalente avec la protéine régulatrice SUMO-1 (Mohan et al. 2007). SUMO-1 est une petite protéine de 97 acides aminés appartenant à la famille des Small Ubiquitin-like Modifier (SUMO) proteins (Matunis et al.1996). Les protéines SUMO peuvent s'attacher de façon covalente (sumoylation) au niveau des résidus lysines ou interagir de façon noncovalente via des motifs protéiques spécifiques. Les interactions covalentes et non covalentes de SUMO-1 sur d'autres protéines telles que TDG ou CBP induisent des modifications post-traductionnelles qui régulent leur activité et contrôlent de nombreux processus cellulaires et notamment dans la régulation de la transcription, de l'apoptose, du cycle cellulaire ainsi que dans la régulation de la stabilité des protéines. Les protéines SUMO modifient aussi la localisation cellulaire des protéines avec lesquelles elles interagissent en induisant notamment leur déplacement au niveau des compartiments nucléaires appelés promyelocytic leukemia protein (PML) oncogenic domains (PODs) ((Chalkiadaki et al. 2005), (Park et al. 2005)). Or les facteurs de transcription tels que CBP et RARα ainsi que les enzymes de réparation de l'ADN tels que TDG sont connus pour être associés à ces compartiments nucléaires ce qui suggère que les protéines SUMO sont impliquées dans la régulation de la formation de ces complexes au niveau de ces compartiments cellulaires (Lin et al. 2001). Nous nous sommes donc intéressés à la régulation de l'activité de réparation et à l'activité transcriptionnelle de TDG par SUMO-1. Cette étude a été publiée dans la revue BMC Biochemistry en février 2011 sous le titre « SUMO-1 regulates the conformational dynamics of thymine-DNA Glycosylase regulatory domain and competes with its DNA binding activity. ».

Comme je vous l'ai présenté en introduction, des études précédentes ont montré une association fonctionnelle entre TDG et le coactivateur transcriptionnel CBP ou CREBbinding protein (Tini et al. 2002). CBP et son proche homologue p300 sont des histones acétyle transférases qui permettent d'activer diverses voies régulées par des facteurs de transcription (Goodman et al. 2000). CBP a, également la possibilité de participer à la fois la régulation transcriptionnelle et à la réparation de l'ADN en contrôlant l'activité de régulation de la transcription et de réparation de l'ADN de TDG via son acétylation.

I _ La régulation de la conformation, de la dynamique et des fonctions de TDG par SUMO-1

Lors de cette étude le Dr. Smet-Nocca a mis au point, par mutagenèse dirigée basée sur les connaissances structurelles de TDG (Hardeland 2002, Baba 2006, Mohan 2007), des mutants permettant l'étude de l'interaction covalente et non-covalente entre TDG et SUMO-1. Elle a tout d'abord muté la lysine 330 (K330) en une alanine (TDG-K330A) pour supprimer le site de sumoylation de TDG, site situé dans le domaine catalytique de TDG (Hardeland et al. 2002). La protéine TDG-wt conjuguée à SUMO-1 au niveau de la lysine 330 (TDG-S) a été obtenue grâce à des bactéries E. coli cotransformées avec le plasmide pGEX-TDGwt et le plasmide pTE1/E2-SUMO-1 (Eilebrecht et al. 2010). La comparaison entre la structure de TDG-sumoylée et de TDG-K330A a mis en évidence que la sumoylation de TDG par SUMO-1 a bien lieu au niveau de la lysine 330.

SUMO-1 peut réguler la conformation et la dynamique de TDG

Des études RMN précédentes de la protéine TDG-wt avaient montré que seules les résonances issues des domaines N et C-terminaux (résidus 1-50 et 328-410 respectivement) (Smet-Nocca et al. 2008) de la protéine TDG pouvaient être détectées grâce à leur très grande flexibilité en solution ; les résonances du corps de la protéine n'étant pas détectables. Des études RMN similaires ont été réalisées sur les protéines TDG-wt, TDG-S et TDG-K330A. Lors de notre étude, nous avons constaté que les résonnances associées à la région N-terminale de TDG (résidus 1-50) ne sont pas

modifiées entre la forme non conjuguée (TDG-wt) et la forme conjuguée (TDG-S) de TDG tandis que les résonnances associées aux résidus 327-347, situés dans la région du site de sumoylation de la lysine 330 (Smet-Nocca et al. 2011) sont modifiées entre ces deux formes. Ces résultats indiquent qu'il existe des modifications conformationnelles du domaine C-terminal de TDG lors de l'interaction covalente avec SUMO-1 tandis que le domaine N-terminal, région qui est impliquée dans l'excision des mésappariements G : T, n'est pas impactée par cette interaction. L'activité de réparation des mésappariements G : U/T est une fonction importante pour la survie des cellules et pour le développement des individus car elle permet de contrôler de façon indirecte la méthylation des cytosines, un des mécanismes essentiel de la régulation épigénétique de l'expression des gènes.

Lors de cette étude nous nous sommes aussi intéressés aux liaisons non covalentes existant entre TDG et SUMO-1. On constate que TDG-wt en présence de SUMO-1 présente des modifications des résonnances associées à la région N-terminale du domaine de régulation de TDG (RD ou TDG-RD) et à la région C-terminale de TDG. Or des études précédentes ont montré que SUMO-1 peut interagir avec TDG de façon non-covalente via deux sites de liaison distincts appelés respectivement SUMO-binding motif 1 ou SBM1 (résidus 133-137) et SBM2 (résidus 308-311). Ces deux sites sont situés dans le domaine catalytique de TDG ((Baba et al. 2005), (Mohan et al. 2007)). Par ailleurs d'autres études avaient mis en évidence que les liaisons non-covalentes entre TDG et SUMO-1 pourraient être soit impliquées dans le processus de sumoylation de TDG lui-même, en tant qu'états intermédiaires, soit dans les interactions fonctionnelles entre TDG et d'autres protéines sumoylées (Takahashi et al. 2005). Nos observations ont montré que SUMO-1 n'induisait pas de changements directs de conformation du N-terminal de TDG via ses interactions non covalentes ce qui est en conformité avec études précédentes ((Baba 2005), (Mohan et al. 2007), (Takahashi et al. 2005)). Nous nous sommes donc intéressés aux sites de sumoylation de TDG ou SBMs

Pour cette étude, le Dr. Smet-Nocca a réalisé les mutants des deux sites présumés de liaison à SUMO-1 ou SUMO Binding Motif par mutagenèse dirigée basée sur les connaissances précédentes de la structure de TDG (Baba 2002, Mohan 2007). Pour déterminer lequel de ces deux sites induisait les changements observés lors de notre étude RMN, nous avons donc produit un mutant du SBM1, TDG-D133A et un mutant du SBM2,

TDG-E310Q ainsi qu'un double mutant SBM1, SBM2, appelé TDG-D133A/E310Q. Le SBM1 est situé au niveau de la région N-terminale de TDG tandis que le SBM2 est situé au niveau de la région C-terminal du site catalytique de TDG. La structure de ces mutants a été analysée en RMN et en spectroscopie dichroïque circulaire.

La mutation D133A située dans la séquence conservée DIVII du SBM1 entraine la perte du repliement correct de la protéine TDG et son agrégation (Smet-Nocca et al. 2011), conséquence que nous ne retrouvons pas pour la protéine TDG-wt ou pour le mutant TDG-E310Q, produites et étudiées dans les mêmes conditions. Le mutant TDG-D133A (SBM1) ne peut donc plus interagir ni avec SUMO-1, ni avec quoi que ce soit d'autre, la protéine ne présentant plus une structure tertiaire fonctionnelle. De plus le motif IVII, à l'exception du résidu D133, n'est pas accessible au solvant dans les structures du domaine catalytique de TDG non modifiée par SUMO-1 et modifiées par SUMO-1. Notre étude a donc montré que SUMO-1 ne peut pas interagir pas avec le mutant TDG-D133A et que ce site de liaison à SUMO-1 appelé SBM-1, situé en terminal du domaine CAT de TDG, ne semble pas en être un.

Lors de l'étude en RMN du mutant TDG-E310Q (SBM2), nous n'avons pas observé de grande différences entre lui et la forme sauvage de TDG (Smet-Nocca et al. 2011). Il semble donc que la mutation E310Q ne modifie pas trop fortement la conformation de TDG à l'inverse de la mutation D133A. Nous nous sommes donc plus particulièrement intéressés à ce mutant et nous avons regardé les modifications de conformation du domaine C-terminal ou du domaine CAT de TDG-E310Q en présence de SUMO-1 (Smet-Nocca et al. 2011). Les résultats obtenus montrent que la liaison de SUMO-1 à TDG est inhibée par la mutation E310Q (SBM2) et donc que la liaison de SUMO-1 au SBM situé en C-terminal (SBM2) est seule responsable des changements de conformations des domaines C et N-terminaux de TDG. En résumé, il n'y a donc qu'un seul vrai SBM chez TDG, le SBM2.

L'interaction intermoléculaire (non-covalente) TDG/SUMO-1 via le site de sumoylation SBM2 entraine des réarrangements structurels (Smet-Nocca et al. 2011). Nous avons aussi constaté que SUMO-1 peut lier le SBM2 chez une protéine TDG déjà sumoylée (Smet-Nocca et al. 2011). La connaissance du mutant E310Q de TDG nous

permet d'affiner notre étude des effets de la sumoylation sur la structure et la dynamique de TDG.

Nous avons constaté au cours de notre étude que la sumoylation de TDG agit plutôt sur la conformation du domaine C-terminal de TDG avec peu ou pas d'impact sur la conformation du TDG-RD tandis que l'interaction non-covalente de SUMO-1 à TDG via le SBM situé en C-terminal est en mesure de modifier structurellement les deux extrémités N- et C-terminales des régions de TDG et TDG sumoylée. Nous avons de plus constaté que la liaison non covalente entraine un déplacement partiel du RD de TDG à partir de son interface CAT. Ces résultats permettent de conclure que SUMO-1 n'adopte pas la même orientation lors de l'interaction non covalente que dans la protéine sumoylée, que cette liaison induit un déplacement partiel du RD de TDG à partir de son interface CAT entrainant un effet d'encombrement stérique au niveau du domaine CAT ; cette dynamique est en accord avec la localisation présumée de l'interface RD / CAT dans la protéine TDG (Smet-Nocca et al. 2008).

SUMO-1 peut réguler l'activité de réparation de l'ADN de TDG

Nous nous sommes ensuite intéressés à l'impact des interactions, covalentes ou non, entre SUMO-1 et TDG sur l'activité de cette dernière. La sumoylation de TDG agit plutôt sur la conformation du domaine C-terminal de TDG et que l'interaction non-covalente de SUMO-1 à TDG via le SBM2 entraine un déplacement partiel du RD de TDG à partir de son interface CAT.

Le RD est le domaine qui permet de réaliser la liaison spécifique de TDG à l'ADN au niveau des mésappariements G : U/T ainsi que la mise en place d'interactions intramoléculaires dynamiques avec le domaine catalytique de TDG (CAT ou TDG-CAT) (Smet-Nocca et al. 2008). Le domaine CAT est quant à lui un domaine très conservé au cours de l'évolution des DNA glycosylases et permet de reconnaitre exclusivement les mésappariements G : U (

Figure 17, p43). Les interactions RD/CAT sont modifiées quand TDG interagit avec son substrat, l'ADN, ce qui entraine une modification de l'activité glycosylase et du turnover enzymatique de TDG. De nombreuses études ont montré que TDG reconnait
divers substrats en plus des bases thymine mal appariées, telles que les bases 3-N4éthenocytosine, 5-methylcytosine (5-meC), pyrimidine, 5-fluorouracile (5-FU), 5hydroxymethyluracile (5-HmU) ou bien 5-bromouracile (5-BrU) (Hardeland et al. 2003) avec une affinité variable en fonction des substrats qu'elle rencontre. TDG présentera, par exemple, un turnover plus lent si elle rencontre un mésappariement G : T ou G : U que si elle rencontre un mésappariement G : 5-FU par exemple (Kunz et al. 2009). L'activité de glycosylase de TDG est donc régulée par la conformation du RD de façon dépendant des mésappariements (Smet-Nocca et al. 2008). Ce résultat couplé aux modifications du domaine C-terminal lors des interactions covalentes avec SUMO-1 et aux modifications de la région N-terminale du domaine de régulation de TDG (RD ou TDG-RD) et de la région C-terminale de TDG indiquent que le domaine catalytique de TDG n'est pas le seul domaine responsable de la spécificité de substrats et d'activité de TDG. Les domaines C- et N-terminaux de TDG sont impliqués aussi dans ces deux fonctions et permettent une plus grande spécificité de substrats et en contrepartie un turnover enzymatique plus faible que pour les autres UDG. Pour compenser ce faible turnover enzymatique, les cellules ont mises au point des voies de régulation de TDG et notamment la modification de sa conformation via des interactions covalentes ou non covalentes avec SUMO-1 (Smet-Nocca et al. 2008), (Smet-Nocca et al. 2011). Nous nous sommes donc demandé comment SUMO-1 régule-t-il l'activité glycosylase de TDG ?

Des études précédentes ont montré que l'interaction intermoléculaire de SUMO-1 avec TDG est fortement diminuée par l'interaction de TDG avec son substrat, l'ADN (Mohan et al. 2007). La présence de substrat d'ADN contenant une paire normale (G: C) ou une paire anormale (G: U/T) peut déplacer de façon similaire le RD de TDG de son interface avec le domaine catalytique (Smet-Nocca et al. 2011) en présence ou en absence de SUMO-1. Les hétéroduplex d'ADN contenant un mésappariement G: U/T entrainent donc une modification de TDG-RD de façon indépendante de la présence de SUMO-1 et empêche la liaison covalente de SUMO-1 au SBM2 (Smet-Nocca et al. 2011) ce qui rejoint nos premières constatations. Nous pouvons donc en conclure que la liaison à l'ADN via le domaine CAT de TDG modifie la conformation du SBM2 ou son accessibilité ce qui empêche toute interaction avec SUMO-1. Lors d'interactions covalentes et non covalentes de SUMO-1 à TDG, nous avons observé une légère stimulation du turnover et de l'activité glycosylase G : T de la protéine TDGwt (Smet-Nocca et al. 2011). A l'opposé, le turnover enzymatique et l'activité G: U sont très sensibles à la sumoylation ou à l'addition de SUMO-1 pour les protéines TDG wild-type de façon dose-dépendant (Smet-Nocca et al. 2011) de façon spécifique à SUMO-1. Il semble donc que l'ajout de SUMO-1 permette d'augmenter la processivité G : U/T de TDG. En contrepartie, la présence d'ADN entraine la rupture des interactions SBM2/SUMO-1 (Smet-Nocca et al. 2011).

Nous avons étudié la capacité de SUMO-1 à interagir directement avec l'ADN et nous avons trouvé que SUMO-1 peut interagir de façon non spécifique avec l'ADN (Eilebrecht et al. 2010). Ce résultat explique l'existence d'une compétition entre SUMO-1 et TDG-RD pour la liaison à l'ADN. Cette compétition partielle entre SUMO-1 et TDG-RD pourrait donc déstabiliser suffisamment le complexe TDG / ADN avec, comme conséquence, une augmentation du turnover G : T / U.

Pour résumer en quelques mots l'action de SUMO-1 sur TDG, nous pouvons dire que les interactions covalentes et non covalentes avec SUMO-1 induisent un déplacement de l'interface RD/CAT. Ce changement de conformation permet à TDG de se détacher de l'ADN tandis que la compétition entre le RD de TDG et SUMO-1 pour la liaison à l'ADN permet à l'ADN de se détacher de TDG. Ces deux mécanismes augmentent le turnover de TDG et son activité glycosylase G : U/T.

SUMO-1 peut modifier la capacité de TDG à interagir avec ses différents partenaires

Des études précédentes ont montré une association fonctionnelle entre TDG et le coactivateur transcriptionnel CBP ou CREB-binding protein (Tini et al. 2002). CBP et son proche homologue p300 sont des histones acétyltransférases qui permettent d'activer diverses voies régulées par des facteurs de transcription en contrôlant l'acétylation de la chromatine et des protéines associées à la chromatine ; CBP/p300 peut donc être recrutée à l'ADN (Goodman and Smolik 2000). CBP a, par ailleurs, la possibilité de participer à la fois la régulation transcriptionnelle et la réparation de l'ADN en contrôlant l'activité de régulation de la transcription et de réparation de l'ADN de TDG via son acétylation. TDG est quant à lui un régulateur de la transcription dépendant de CBP/p300. Par ailleurs, nous savons que la sumoylation de TDG ou ses interactions non covalentes avec SUMO-1

permettent non seulement de réguler ses propriétés biochimiques (Hardeland et al. 2002) comme je vous l'ai présenté ci-dessus mais empêche sa relocalisation aux PODs où TDG peut interagir avec CBP (Mohan et al. 2007). Les PML oncogenic domains (PODs) sont des compartiments nucléaires riches en protéines Promyelocytic leukemia (PML) (Le Yang et al 1996), un facteur de transcription ayant une activité de suppresseur de tumeur ; ces protéines sont codées par le gène PML du chromosome 15. Les protéines PML sont connues pour interagir avec le RAR α (Zhong 1999), avec CBP ((Matsuzaki et al. 2003), (Doucas et al. 1999)), avec TDG (Takahashi et al. 2005), avec SUMO-1 ((Kamitani et al. 1998), (Lin et al. 2001)) et avec de nombreuses autres protéines impliquées dans la régulation de l'expression des gènes.

Nous nous sommes demandés quel est l'impact de l'interaction TDG/ SUMO-1 sur les autres fonctions de TDG telle que la capacité à activer CBP (Tini et al. 2002). Pour cela nous avons utilisé le TDG-K330A qui a perdu la capacité à interagir avec SUMO-1. Nous avons analysé l'impact de la sumoylation sur l'interaction CBP/TDG in vivo grâce une étude en microthermophorèse ainsi qu'une étude du mutant TDG-K33OA par imagerie confocale en utilisant la technique du FRET (Figure 34, p 88) (Spriet 2006)). Lors de l'étude FRET, les cellules sur-exprimant la protéine CFP-CBP et la protéine YFP-TDG-wt présentent une légère diminution du temps de vie de fluorescence de la CFP contrairement aux cellules sur-exprimant uniquement CFP-CBP (Spriet 2006)). L'analyse de la dynamique de l'interaction entre CBP et TDG lors d'expériences de microthermophorèse YFP-CBP et HA-TDG-wt montre la présence d'une forte interaction entre CBP et TDGwt. Ce résultat valide bien l'existence d'un complexe CBP/TDG complexe *in vivo*.

Des expériences de FRET ont été réalisées avec le mutant de sumoylation TDG-K330A en parallèle de celles réalisées avec la forme sauvage de TDG. On observe une légère diminution du temps de vie de fluorescence de la CFP dans les cellules surexprimant la protéine CFP-CBP et la protéine YFP-TDG-K330A en comparaison avec les cellules sur-exprimant la protéine CFP-CBP et la protéine YFP-TDG-wt. Il y a donc augmentation de l'interaction CBP/TDG en absence de SUMO-1. La sumoylation de TDG est donc bien un moyen de contrôler la stabilité du complexe CBP/TDG. Nous savons que CBP peut moduler l'activité de TDG par son acétylation au niveau des lysines K59, K83, K84, K87 (Tini et al 2002) et que SUMO-1 déstabilise le complexe CBP/TDG. Nous nous sommes donc intéressés à l'impact de l'acétylation de TDG par CBP sur la stabilité du complexe CBP/TDG. Nous avons émis comme hypothèse que l'acétylation et l'interaction, covalente ou non, avec SUMO-1 sont deux mécanismes contradictoires qui contrôlent la formation et la stabilité du complexe CBP/TDG dans la cellule. Pour cela, nous avons analysé les résultats obtenus en FRET pour les mutants d'acétylation TDG-K59A, TDG-K83, K84, K87A (TDG- Δ 3) et TDG-K59, K83, K84, K87A (TDG- Δ 4). Ces mutants ont été mis au point, par mutagenèse dirigée basée sur les connaissances structurelles de TDG (Tini et al 2002) par le Dr. Smet-Nocca. Lorsque les différentes lysines sont mutées et donc ne peuvent plus permettre l'acétylation de TDG par CBP, que se passe-t-il pour la complexe CBP/TDG ?

On observe une diminution du temps de vie de fluorescence de CFP dans les cellules sur-exprimant les protéines CFP-CBP et YFP-TDG-K59A d'une part et dans les cellules sur-exprimant les protéines CFP-CBP et YFP-TDG- Δ 3 d'autre part et ce par comparaison avec les cellules sur-exprimant les protéines CFP-CBP et YFP-TDG-wt (Figure 49, p102) (Spriet 2006)). Il y a donc augmentation de l'interaction CBP/TDG en absence du site d'acétylation K59A ou des trois sites d'acétylation K83A, K84A, K87A. En effet, malgré ces mutations, CBP peut toujours acétyler au moins une ou des lysines, permettant de lier CBP et TDG. La mutation K59A et les mutations K83A, K84A, K87A (TDG- Δ 3) stabilisent le complexe CBP/TDG en limitant l'encombrement stérique dû aux trop nombreuses acétylations mais ne permet pas de déstabiliser le complexe CBP/TDG car il reste au moins un site acétylé donc a moins une liaison entre CBP et TDG (Figure 70, p148).



Figure 70: Schéma de l'impact des mutants d'acétylation sur la stabilité du complexe CBP/TDG. En bas, j'ai représenté le complexe CBP/TDG-wt avec la protéine TDG acétylée au niveau de ses quatre sites d'acétylation. Puis j'ai représenté les différents mutants d'acétylation par ordre croissant de déstabilisation du complexe CBP/TDG.

On observe une augmentation du temps de vie de fluorescence de CFP dans les cellules sur-exprimant la protéine CFP-CBP et la protéine YFP-TDG- Δ 4 en comparaison avec les cellules sur-exprimant la protéine CFP-CBP et la protéine YFP-TDG-wt (données non publiées). Il y a donc diminution voir perte de l'interaction CBP/TDG en absence de tous les sites d'acétylation de TDG. L'acétylation de TDG par CBP est une autre voie de régulation de la stabilité du complexe CBP/TDG.

Notre hypothèse concernant les rôles contradictoires de l'acétylation et l'interaction, covalente ou non avec SUMO-1 sur la stabilité du complexe CBP/TDG est validée par cette étude. Il est donc intéressant de comprendre le rôle de ce complexe dynamique dans la cellule. Ces résultats ont donc été complétés par une analyse transcriptomique du complexe CBP/TDG réalisée au cours de ma thèse.

II_ Etude in vivo des complexes CBP/TDG et CBP/TDG/RARa dans la régulation de la transcription.

Etude in vivo du complexe CBP/TDG dans la régulation de la transcription.

Lors de ma thèse, j'ai donc tout d'abord regardé la présence *in vivo* d'un complexe CBP/TDG. Pour étudier le rôle de ce complexe sur le transcriptome, il nous fallait pouvoir le disloquer sans pour autant modifier les fonctions des deux composants, CBP et TDG. Pour cela, nous nous sommes orientés vers le mutant TDG-P65A caractérisé par le Dr. Smet-Nocca.

Le mutant TDG-P65A présente une mutation au niveau de la proline 65, proline située dans le domaine de régulation de TDG et proche des sites de liaison à CBP. Les prolines sont des acides aminés qui ne peuvent plus former de liaison hydrogène du fait de leur amine secondaire. Elles peuvent donc déstabiliser les hélices α ou former des bifurcations dans les feuillets β . Par ailleurs, la proline 65 est située au milieu des sites d'acétylation de TDG par CBP et donc dans la région permettant l'interaction CBP/TDG. Ce sont ces propriétés de la proline 65 qui nous ont poussé à vérifier la structure du mutant TDG-P65A par RMN.

On constate que les résonnances du mutant TDG-P65A se superposent presque parfaitement aux résonnances de la forme sauvage de TDG tant dans le contexte de la protéine entière que dans le contexte de la partie N-terminale de la protéine. Les quelques différences observées sont dues aux modifications de l'environnement électronique des acides aminés voisins de la proline 65. Il n'y a pas de changements conformationnels des domaines N et C-terminaux ni du domaine de régulation de TDG sous l'effet de la mutation P65A. TDG-P65A possède donc une structure similaire à la forme sauvage de TDG.

La protéine TDGP65A présente une activité glycosylase G : U/T ainsi qu'un turnover enzymatique identique à la protéine wt. Enfin, une première analyse en microscopie (expériences de FRET) avait montré que la mutation P65A disloquait le complexe CBP/TDG même en présence de RA (données non présentées). Ces résultats ont été confirmés in vivo par des études en microthermophorèse dans des cellules HEK293. Lors de ces expériences de microthermophorèse, j'ai choisi de marquer la protéine CBP avec le fluorophore en utilisant le plasmide codant pour la protéine YFP-CBP car CBP est le composant le plus volumineux et donc dont le taux de transfection sera le plus faible. On constate que dans les mêmes conditions de transfection, de préparation des extraits protéiques et de microthermophorèse, le mélange YFP-CBP/HA-TDG-P65A présente un jump de fluorescence décalé vers les concentrations plus élevées de HA-TDG. Cela veut dire qu'il faut une concentration beaucoup plus élevée (10 fois) de TDG-P65A que de TDG-wt pour obtenir la même vitesse de migration de YFP-CBP. TDG-P65A présente donc un Kd relatif de dissociation plus élevé donc une moins grande stabilité sous la forme d'un complexe CBP/TDG. Le mutant TDG-P65A est donc un bon outil pour déstabiliser le complexe CBP/TDG et en étudier les effets sur le transcriptome sans modifier la structure ou l'activité glycosylase de TDG. J'ai donc orienté mon choix vers ce mutant lors de mon analyse transcriptomique du complexe CBP/TDG.

Des études précédentes à ma thèse ont montré que TDG peut modifier l'activité du corégulateur transcriptionnel CBP/p300 (Tini et al. 2002). Une publication de 2011 a validée notre hypothèse du rôle de corégulateur épigénétique de la transcription de TDG dans un modèle souris (Cortázar et al. 2011). Nous avons validé de notre côté l'hypothèse du rôle de TDG sur la transcription dans des modèles « humains ». La comparaison des profils transcriptomiques de cellules COS-7 et HEK293 sur-exprimant d'YFP-TDG et des profils transcriptomiques de cellules non transfectées avec un filtrage sur une d'expression différentielle entre les deux conditions inférieure montre l'existence de 2 509 gènes chez les cellules COS-7 (p \leq 0,05) et de 2 032 gènes chez les cellules HEK293 (p \leq 0,005) (Figure 50, p115); la majorité de ces gènes présentent un Log Q > 0. TDG a donc une action de corégulateur positif de la transcription des gènes. La classification en ontologies de ces

gènes montre que les fonctions moléculaires suivantes *« Ribosomal protein, Nucleic acid binding, mRNA splicing factor, mRNA processing factor* et *Kinase »* sont sur-représentés. TDG-wt semble donc avoir un impact majeur sur la régulation des gènes notamment via la régulation de l'expression des ARN messagers et de leur épissage. L'ontologie *« Kinases »* présente plus de gènes activés observés que le nombre attendu tandis qu'elle présente moins de gènes réprimés que le nombre attendu. TDG-wt a donc un fort impact sur l'activation des kinases dans les cellules. Or les kinases sont des enzymes appartenant au groupe des transférases souvent activées par des récepteurs membranaires. Les kinases permettent la catalyse des réactions de phosphorylation par l'ajout d'un ion phosphate à son substrat qui peut être une protéine, un lipide ou un sucre. Cette réaction de phosphorylation permet de transmettre des signaux et donc d'induire une réponse appropriée de la cellule. TDG est joue donc un rôle dans la régulation de l'expression des ARN messagers et de leur épissage, la division, la croissance et la prolifération cellulaires. TDG semble donc impliqué dans le développement embryonnaire ; c'est un corégulateur de la transcription tout autant qu'une enzyme de réparation de l'ADN.

La comparaison entre les profils transcriptomiques des cellules sur-exprimant CBP et des cellules non transfectées montre 2 818 gènes ($p\leq0,05$) chez les cellules COS-7 et 3 020 chez les cellules HEK293 ; la majorité de ces gènes sont là aussi activés (Figure 51A, p117) et les fonctions moléculaires des gènes correspondant peuvent être classées dans les ontologies suivantes : «*Nucleic acid binding, Ribosomal protein, Isomerases* et *Histones* ». Ces fonctions sont sur-représentées par rapport à des conditions normales (Figure 51B et C, p117). L'ontologie «*Histone* » présente plus de gènes activés que le nombre attendu tandis que les ontologies «*Kinase* et *Transferase* » présentent plus de gènes réprimés que le nombre attendu. CBP semble donc avoir un impact majeur sur la régulation des gènes notamment via son rôle d'histone acétyltransférase. CBP semble aussi avoir un impact sur l'inactivation de certaines voies de signalisation dans les cellules. CBP a donc lui-aussi une action de corégulateur positif de la transcription des gènes.

Cette analyse transciptomique confirme les rôles de régulateurs transcriptionnels de CBP et TDG-wt *in vivo*. Je me suis donc demandé au cours de mes travaux de doctorat quel est l'impact du complexe dynamique CBP/TDG sur la régulation de la transcription.

Pour cela j'ai regardé les profils transcriptomiques des cellules sur-exprimant CBP et TDG-wt ou CBP et TDG-P65A. Nous avions posé comme hypothèse au début de mon travail de doctorat que le complexe CBP/TDG sur le transcriptome induisait plus fortement les gènes sous la dépendance de CBP et/ou de TDG-wt. Nous nous attendions donc à voir les gènes activés dans la condition CBP/TDG-wt fortement surexprimés dans la condition CBP/TDG-K330A et réprimés dans les conditions CBP/TDG-P65A. En effet, comme je vous l'ai présenté précédemment, la mutation K330A stabilise le complexe CBP/TDG tandis que la mutation P65A déstabilise le complexe CBP/TDG. Malheureusement je n'ai pas vu un tel pattern d'expression lors de mon analyse. On observe une majorité de gènes activées par la surexpression de CBP et de TDG-wt. On observe aussi que ce profil transcriptomique est similaire à celui obtenu lorsque l'on surexprime CBP et TDG-K330A c'est-à-dire lorsque l'on stabilise le complexe CBP/TDG.

Par ailleurs l'analyse en ontologies à l'aide de la classification panther et de la correction de Bonferroni appliquée par le logiciel Ace.map n'ayant pas donné de résultats concluant pour ces profils transcriptomiques, j'ai donc regardé de façon globale la fonction des gènes statistiquement différentiellement exprimés dans les différentes conditions (p ≤0.05) à l'aide d'un travail bibliographique. J'ai classé les gènes significativement différentiellement exprimés (p ≤0.05) selon 19 catégories (Figure 54, p121) : "mRNA transcription and regulation (TF), chromatin packaging and remodeling, nucleoside, nucleotide and nucleic acid metabolism, mRNA splicing and pre-mRNA processing, cell cycle, apoptosis, oncogenesis, cell proliferation and differentiation, receptors, signal transduction, transport, cell structure and motility, cell adhesion and cell communication, development, mesoderm development, ectoderm development, protein metabolism and modification, immunity and defense et others". J'ai remarqué que les catégories « mRNA transcription and regulation (TF), Nucleoside, nucleotide and nucleic acid metabolism, signal transduction et development » sont fortement sur-représentées (autour de 20%) par rapport à leur poids lors d'un tirage aléatoire de 2000 gènes. Les catégories « transport, protein metabolism and modification et others » représentent chacune environ 15% des gènes modifiés par le complexe CBP/TDG et les catégories « Cell proliferation and differentiation, receptors, cell adhesion and cell communication et immunity and defense » représentent elles environ 10% des gènes. La modification de la stabilité du complexe

CBP/TDG dans les cellules ne semble donc pas impacter la régulation de la transcription de façon globale. Deux hypothèses s'ouvrent à nous :

- le taux de transfection n'étant pas forcément homogène entre CBP et TDG à cause de la taille respective de ces protéines, peut-être ne voyons-nous que l'effet de TDG ou de ses mutants ponctuels ;
- (ii) CBP et TDG ne sont pas seuls au sein du complexe fonctionnel de régulation de la transcription.

La surexpression de TDG-wt et son mutant TDG-P65A induit des différences significatives entre les profils transcriptomiques obtenus. L'exemple de la mutation P65A semble démontrer que l'hypothèse (ii) est la plus plausible.

J'ai aussi constaté un enrichissement de mes ontologies avec des gènes issus de l'ectoderme et du mésoderme. Ces deux feuillets embryonnaires sont respectivement impliqués dans le développement et la fonction du système nerveux et dans le développement et la fonction du système musculaire et cardio-vasculaire. Cela m'a donc incité à faire une sous-classification des gènes impliqués dans le développement régulés par le complexe CBP/TDG, dans mon analyse. Ils sont pour la moitié d'entre eux des gènes impliqués dans le système nerveux (Figure 55, p122). Nos résultats rejoignent donc les résultats publiés dans la revue Nature (Cortázar et al. 2011) et dans la revue Cell (Cortellino et al. 2011).

J'ai ensuite regardé les profils d'expression d'un petit groupe de gènes impliqués dans le développement du système nerveux (Figure 56, p122). On constate que la majorité de ces gènes sont surexprimés (en rouge) avec un fold-change dont le Log Q est supérieur ou égal à 1. Les gènes réprimés (en bleu) présentant un fold-change dont le Log Q est supérieur à 1 sont peu nombreux. Ces gènes ci-dessus sont donc majoritairement surexprimés dans les cellules transfectées avec CBP/TDG. Ces résultats tendent vers un rôle du complexe CBP/TDG dans la régulation de la mise en place et du maintien du système nerveux au cours du développement et chez l'adulte. Nous savons que l'acide rétinoïque est un acteur de l'induction de la mise en place et du maintien du système nerveux au cours du développement et chez l'adulte et que la perturbation de la

signalisation de RA chez l'adulte induit une dégénérescence des neurones moteurs, le développement de la maladie d'Alzheimer et, éventuellement, le développement de la maladie de Parkinson. De plus, une étude récente a mise en évidence l'impact de TDG sur la différenciation des progéniteurs neuronaux dépendant à l'acide rétinoïque (Cortázar et al. 2011). Ces résultats tendent vers un rôle du complexe CBP/TDG dans la régulation des gènes dépendant de l'acide rétinoïque (RA).

Etude *in vivo* du complexe CBP/TDG/RARα dans la régulation de la transcription.

Comme je l'ai noté auparavant, TDG interagit avec RARα (Um 1998). Des études avaient déjà commencé au laboratoire pour voir si le RA influençait le complexe CBP/TDG. Ces études de la stabilité de l'interaction entre CBP et TDG en présence de doses croissantes d'acide rétinoïque de forme trans (all trans retinoic acid ou ATRA) ont été réalisées par FRET. On constate une décroissance du temps de vie de fluorescence de la CFP dans les cellules sur-exprimant la protéine CFP-CBP et la protéine YFP-TDG-wt en présence de ces doses croissantes ATRA (Figure 57, p116). L'acide rétinoïque semble donc stabiliser le complexe CBP/TDG de façon dose-dépendante. L'ATRA étant dissout dans de l'éthanol 100%, nous avons utilisé le traitement à l'éthanol comme contrôle négatif pour éliminer tout impact de l'éthanol sur le temps de vie de fluorescence de la CFP. On constate que le traitement à l'éthanol ne modifie en rien la fluorescence du complexe CBP/TDG. L'augmentation de l'interaction entre CFP-CBP et YFP-TDG est donc bien induite par la présence d'ATRA.

On sait que le mutant TDG-P65A a permis de casser le complexe CBP/TDG. On sait aussi que RAR α doit être activé par la liaison de son ligand, l'ATRA (RAR α *) pour pouvoir interagir avec CBP. On pose donc comme hypothèses que le complexe CBP/TDGwt/RAR α * est plus stable que le complexe CBP/TDG-wt/RAR α et CBP/TDG-P65A/RAR α *, que le complexe CBP/TDG-wt/RAR α est plus stable que le complexe CBP/TDG-P65A/RAR α .

On constate que dans les mêmes conditions de microthermophorèse, le complexe CBP/TDG-wt/RAR α activé par l'ajout de 5.10⁻⁷ M d'ATRA est plus stable que le complexe CBP/TDG-P65A/RAR α activé par l'ATRA (RAR α *) tandis que le CBP/TDG-wt/RAR α en absence d'ATRA est plus stable que le complexe CBP/TDG-wt et que les complexes CBP/TDG-P65A/ RAR α en absence d'ATRA (RAR α) et CBP/TDG-P65A. Cette expérience m'a donc permis de mettre en évidence la présence d'un complexe CBP/TDG/RAR α . J'ai par ailleurs montré que le mutant TDG-P65A permet de déstabiliser le complexe CBP/TDG/RAR α en empêchant l'interaction entre CBP et TDG. On peut ensuite jouer sur la présence ou non d'ATRA pour renforcer ou non l'interaction entre

CBP et RAR α . Des expériences d'activation Gal4 de la protéine RAR α par la forme sauvage ou les formes mutées de TDG ont été réalisées et ont montré que le mutant TDG-P65A peut activer, bien que plus faiblement (1.5 fois) l'activité luciférase du gène rapporteur sous la dépendance de la protéine de fusion Gal4-RAR α . Les résultats de l'analyse en microthermophorèse ainsi que l'étude de l'activité luciférase sous le contrôle du mutant TDG-P65A ont donc conformé mon choix d'utiliser ce mutant pour l'analyse transcriptomique du complexe CBP/TDG/RAR α .

J'ai tout d'abord observé dans mes analyses l'impact de TDG-wt sur la transcription des gènes dépendant de l'acide rétinoïque en cofiltrant la soustraction « cellules COS-7 traitées avec 5.10⁻⁷ M d'ATRA contre cellules non traitées » avec la soustraction « cellules COS-7 sur-exprimant TDG-wt contre cellules COS-7 non transfectées » et je me suis intéressée à l'impact de TDG sur la transcription des gènes dépendant de l'acide rétinoïque. Lorsque j'ai filtré la soustraction « cellules COS-7 traitées avec 5.10^{-7} M d'ATRA contre cellules non traitées » à p ≤ 0.05 , j'ai obtenu 2 666 gènes. Cette liste de gènes a été utilisée dans la soustraction «cellules COS-7 sur-exprimant TDGwt contre cellules COS-7 non transfectées » pour regarder le comportement des gènes associés en présence de TDG-wt seul (Figure 52, p118). On constate que les gènes activés par le traitement à l'acide rétinoïque des cellules sont en majorités réprimés par la surexpression de TDG-wt dans ces mêmes cellules et réciproquement. J'ai par ailleurs comparé les gènes statistiquement (p≤0,005) différentiellement exprimés dans les soustractions «HEK293 traitées à l'ATRA contre HEK293 non traitées » et «HEK293 sur-exprimant TDG-wt contre HEK293 non transfectées». On remarque le même patron d'expression inversé entre l'acide rétinoïque et TDG. TDG semble donc avoir un impact négatif sur la transcription des gènes dépendant de l'acide rétinoïque dans nos deux modèles.

Ces différents résultats (rôle de CBP et de TDG sur la transcription des gènes dans les cellules COS-7 et dans les cellules HEK293, modification des profils transcriptomiques de l'ATRA en présence de TDG-wt) nous ont donc conforté dans notre hypothèse, à savoir que TDG agit sur le transcriptome au sein d'un complexe transcriptionnel comprenant CBP et/ ou RAR α .

Nous avons donc comparé les gènes statistiquement ($p \le 0,005$) différentiellement exprimés dans les soustractions « HEK293 sur-exprimant CBP et TDG-wt traitées à l'ATRA contre HEK293 sur-exprimant CBP et TDG-wt non traitées » et « HEK293 surexprimant CBP et TDG-P65A traitées à l'ATRA contre HEK293 sur-exprimant CBP et TDG-P65A non traitées » Je m'attendais à voir les gènes dépendant à l'ATRA surexprimés par la présence du complexe CBP/TDG. Or ce n'est pas ce que je constate dans mes analyses. La présence du complexe CBP/TDG (cotransfection CBP/TDG-wt) semble éteindre ou du moins inverser l'expression des gènes sous la dépendance de l'acide rétinoique tandis que lorsque le complexe CBP/TDG est dissocié (cotransfection CBP/TDG-P65A), on observe un retour à la « normale » de la majorité des gènes sous la dépendance de l'acide rétinoïque. La régulation de la transcription des gènes dépendant de l'ATRA est donc affectée par la perte de l'interaction CBP/TDG (Figure 67, Figure 68, Figure 69, p135 à 137).

Conclusion

Lors de mes travaux de doctorat, j'ai continué la caractérisation totale ou partielle des mutants de fonctionnalité TDG-P65A et TDG-K330A, à l'aide d'une approche interdisciplinaire. Le mutant TDG-K330A a permis à notre équipe d'étudier le rôle de SUMO-1 sur la conformation de TDG et sur sa régulation de l'activité glycosylase des ponts reliant les bases mal-appariées à leur sucre. SUMO-1 permet de changer la structure de TDG de la forme « ouverte », liée à l'ADN, à la forme « fermée » qui s'attache moins fortement à l'ADN. La compétition pour la liaison à l'ADN entre SUMO-1 et le domaine catalytique de TDG semble être un autre moyen d'augmenter le turnover de TDG.

J'ai aussi montré que le complexe CBP/TDG, stabilisé par l'acide rétinoïque de façon dose-dépendante, possède bien une activité transcriptionnelle. Le complexe CBP/TDG régule de façon négative l'expression des gènes sous la dépendance de l'acide rétinoïque. L'expression de ces gènes est « réactivée » lorsque l'on casse l'interaction entre CBP et TDG grâce à l'utilisation du mutant TDG-P65A.

La mutation P65A est le remplacement de la proline 65 en une alanine ; TDG perd donc un point de flexibilité cis/trans. Par ailleurs ce mutant est particulièrement intéressant car cette mutation ponctuelle ne modifie ni sa structure si son activité glycosylase lié à la conformation de son domaine catalytique et de régulation. Les prolines sont un des substrats des isomérases appelées peptidyl-prolyl cis/trans isomerases (PPIases) qui catalysent l'isomérisation cis/trans des ponts peptidiques peptidyl-prolyl. On peut donc supposer que la modification de l'activité transcriptionnelle de ce mutant par rapport à la forme sauvage de TDG est liée à la perte de ce site d'isomérisation post-traductionnel induit par une PPIase.

Enfin, j'ai montré au cours de mes travaux la présence d'un complexe CBP/TDG/RARα où TDG interagit avec autant avec CBP qu'avec RARα.



Figure 71 : Schéma récapitulatif des complexes CBP/TDG-wt/RARα et CBP/TDG-P65A/RARα A gauche, schéma du complexe CBP/TDG-wt/RARα et à droite schéma du complexe CBP/TDG-P65A/RARα.

Perspectives

Je pense qu'il serait intéressant de continuer à caractériser les différents mutants de TDG et notamment de comparer la force d'interaction entre CBP et TDG pour les mutants TDG-K330A, TDG-K59A, TDG- Δ 3 en présence et en absence du récepteur à l'acide rétinoïque α . Il serait aussi intéressant de comparer la perte de l'interaction entre TDG- Δ 4 et TDG-P65A en présence et en absence du récepteur à l'acide rétinoïque α . Pour cela nous pourrions utiliser la technique de microthermophorèse.

L'analyse transcriptomique des mutants TDG-K330A, TDG- K59A, TDG- Δ 3, TDG- Δ 4 en présence d'acide rétinoïque doit aussi être approfondie grâce à des études de transcriptomique.

De nombreuses équipes travaillent avec des Mouse Embryonic Fibroblasts (MEFs), des cellules embryonnaires de souris et des progéniteurs neuronaux Tdg KO. Il serait intéressant de suivre les profils transcriptomiques de ces cellules après avoir réactivé TDG à l'aide de plasmides codant pour TDG-wt ou pour les différents mutants de TDG. Par ailleurs, il serait intéressant de regarder l'impact du complexe CBP/TDG chez ces cellules notamment sur les gènes du développement embryonnaire précoce. Pour étudier le développement embryonnaire précoce, il serait aussi possible de faire des souris chimères Tdg ⁷/. Tdg ⁺/₊ et de suivre l'évolution des cellules issues des différentes couches embryonnaire. Je pense que ces embryons seraient viables au moins jusqu'au stade embryonnaire E11,5 comme le sont les souris Tdg ⁷/. Regarder s'il existe des KO conditionnels pour Tdg est aussi une autre option envisageable.

Enfin des études ont montré que SUMO-1 induit le changement de localisation de TDG, CBP et des RAR au sein du noyau et notamment dans les PODs, une étude à l'aide d'hybridization in situ permettrait de voir si la perte de la fonctionnalité d'un domaine de TDG (TDG-K330A, TDG-K59A, TDG- Δ 3, TDG- Δ 4 ou TDG-P65A) en présence des deux autres composants induirait des

modifications de leur localisation respective dans le noyau au cours du cycle cellulaire.

Annexes

Lors de ma thèse, j'ai apporté mes compétences dans l'acquisition de profils transcriptomiques à divers projets dont je vais vous présenter les travaux en découlant.

Projet N°1 _ Acquisition de profils transcriptomiques pour l'étude des mécanismes de réponse à une infection par Streptococcus pneumoniae.

Trois papiers portent sur l'étude des mécanismes de réponse à l'infection par *Streptococcus pneumoniae* chez la souris.

Ces études ont été réalisées dans des modèles murins permettant l'étude des mécanismes de défense de l'organisme face à une infection bactérienne à *Streptococcus pneumoniae*. En effet, l'utilisation de plus en plus importante des antibiotiques couplée à l'efficacité pas toute à fait complète des vaccins entraine une augmentation des maladies bactériennes notamment en milieu hospitalier. Il est donc important de mieux comprendre les mécanismes de défense de l'organisme face à ces infections. *Streptococcus pneumoniae* est une bactérie qui cause des pneumonies, des méningites et entraine dans de nombreux cas une septicémie.

Pour ces projets, j'ai réalisé les étapes d'amplification, de marquage et l'hybridation des cRNA obtenus sur des puces pangénomiques de souris Applied Biosystems (Cat. N° 4345065) comme présenté lors du matériels et méthodes. J'ai par ailleurs effectué la lecture de ces puces permettant l'obtention de profils transcriptomiques.

I.1. La signalisation via TLR5 stimule la production naturelle d'IL-17 et d'IL-22 par l'activation des cellules immunitaires CD3 (neg) CD127 + de la rate et de la muqueuse.

Le premier papier a été publié dans le journal Journal of Immunology en juillet 2010 sous le titre « TLR5 signaling stimulates the innate production of IL-17 and IL-22 by CD3(neg)CD127+ immune cells in spleen and mucosa. » par Van Maele L, Carnoy C,

Cayet D, Songhet P, Dumoutier L, Ferrero I, Janot L, Erard F, Bertout J, <u>Leger H</u>, Sebbane F, Benecke A, Renauld JC, Hardt WD, Ryffel B et Sirard JC.

Lors de l'immunité adaptative, les lymphocytes Th17 produisent les cytokines IL-17 et IL-22 ; des cytokines qui stimulent les défenses des muqueuses et la réparation des tissus infectés. Van Maele et al. ont montré l'activation des gènes codant pour les cytokines Il-17 et IL-22 par la Flagelin, un agoniste des TLR5, dans les tissus lymphoïques, les intestins et les poumons chez la souris. Cette publication a mis en avant l'activation par les TLR des cellules CD3^{neg}CD127⁺ et la production des cytokines Th17related. Ces deux mécanismes semblent cruciaux la défense des muqueuses de l'hôte lors de la réponse immunitaire immédiate.

Dans cette étude, il a été observé que la Flageline agoniste de TLR5 induit la transcription rapide et transitoire des gènes codant l'IL-17 et l'IL-22 dans les tissus lymphoïdes, les intestins et les poumons. Cette réponse innée renforce temporairement l'expression de gènes associés à la signature antimicrobienne Th17. Cette étude a par ailleurs permis d'identifier la source des cytokines Th17-related comme étant un sousensemble de cellules CD3negCD127 + ; parmi ces cellules, il existe des cellules exprimant CD4 et ressemblant à des cellules inductrices de tissus lymphoïdes. Cette étude a aussi permis de démontrer le rôle essentiel des cellules dendritiques lors de l'expression de cytokines liées Th17 et ce qui explique la stimulation des cellules innées par ces cellules dendritiques. Ces données montrent donc que l'activation des cellules CD3negCD127 + induite par TLR et que la production de cytokines Th17-related peut être cruciale pour les défenses précoces contre l'invasion pathogène des tissus de l'hôte (ici la souris).

En conclusion, l'étude a démontré que les cellules LinnegCD127 + induisent une réaction rapide à la signalisation TLR, et sont une source innée d'IL-17A, d'IL-17F, et d'IL-22 suite à cette signalisation. Cette étude a permis d'émettre l'hypothèse que cette réaction immunitaire se produit lors de la pénétration microbienne dans la lamina propria (tissu conjonctif lâche situé sous les épithéliums qui tapissent notamment les muqueuses digestives, respiratoires ou urogénitales). La réaction immunitaire stimule à ce moment-là, des effecteurs innés pour éliminer localement l'infection. Des populations de cellules

similaires ont été identifiés chez l'homme (14, 16), il reste donc à voir si la stimulation des TLR peut aussi favoriser l'activation de ces cellules immunitaires innées chez l'Homme.

TLR5 Signaling Stimulates the Innate Production of IL-17 and IL-22 by CD3^{neg}CD127⁺ Immune Cells in Spleen and Mucosa

Laurye Van Maele,^{*,†,‡,§,¶,1} Christophe Carnoy,^{*,†,‡,§,¶,1} Delphine Cayet,^{*,†,‡,§,¶,1} Pascal Songhet,^{||} Laure Dumoutier,[#] Isabel Ferrero,^{**} Laure Janot,^{††,‡‡} François Erard,^{††,‡‡} Julie Bertout,[¶] Hélène Leger,^{§,§§,¶¶} Florent Sebbane,^{*,†,‡,§,¶} Arndt Benecke,^{§,§§,¶¶} Jean-Christophe Renauld,[#] Wolf-Dietrich Hardt,^{||} Bernhard Ryffel,^{††,‡‡} and Jean-Claude Sirard^{*,†,‡,§,¶}

In adaptive immunity, Th17 lymphocytes produce the IL-17 and IL-22 cytokines that stimulate mucosal antimicrobial defenses and tissue repair. In this study, we observed that the TLR5 agonist flagellin induced swift and transient transcription of genes encoding IL-17 and IL-22 in lymphoid, gut, and lung tissues. This innate response also temporarily enhanced the expression of genes associated with the antimicrobial Th17 signature. The source of the Th17-related cytokines was identified as novel populations of CD3^{neg}CD127⁺ immune cells among which CD4-expressing cells resembling lymphoid tissue inducer cells. We also demonstrated that dendritic cells are essential for expression of Th17-related cytokines and so for stimulation of innate cells. These data define that TLR-induced activation of CD3^{neg}CD127⁺ cells and production of Th17-related cytokines may be crucial for the early defenses against pathogen invasion of host tissues. *The Journal of Immunology*, 2010, 185: 1177–1185.

oll-like receptors are key players in innate immunity and are essential for sensing microbial components and triggering the host defense (1). At the luminal interface, the TLR response is mediated by the epithelium and mainly consists of neutrophil recruitment and activation (2). After microbes cross the epithelium, sensing occurs within the lamina propria. However, the nature of the TLR-mediated innate cells and defense factors that are triggered by microbial desequestration has yet to be defined.

¹L.V.M., C.C., and D.C. contributed equally to this work.

Received for publication January 15, 2010. Accepted for publication May 11, 2010.

Recent studies highlighted the contribution of IL-17A, IL-17F, and IL-22 to defensive reactions within the mucosa (3-6). These cytokines help orchestrate innate immunity by stimulating epithelial cells to produce defense molecules, matrix proteases, and tissue repair molecules (7, 8). The source of IL-17A, IL-17F, and IL-22 varies. During an adaptive response, the lymphocytes that differentiate into Th17 cells are the main producers of cytokines (9). IL-17A can rapidly be produced during innate responses to bacteria or microbial molecular patterns by γδ T lymphocytes in a TLR4dependent manner, NKT cells activated with α -galactosylceramide, or lymphoid tissue inducer (LTi)-like cells following stimulation with the TLR2/Dectin-1 agonist zymosan (10-12). NK-like and LTi-like innate lymphocytes expressing IL-7Ra, NKp46, the transcription factor RORyt, and eventually CCR6 are sources of IL-22 and/or IL-17 in mucosa under steady-state conditions (13-18). Interestingly, microbial flora-colonizing mucosa are required to switch on lasting IL-17 and IL-22 production (15, 17). In absence of these innate lymphocytes, infectious colitis is exacerbated, suggesting an operational role of IL-22 and IL-17 in the gut's innate immunity (15, 17). However, the link between TLR-mediated signaling, Th17-related cytokine production by innate immune cells, and mucosal defenses has not been defined.

The ability of TLR5 signaling to induce mucosal production of IL-17 and IL-22 and thereby promote antimicrobial defense has never been investigated. TLR5 detects flagellins—the main protein of bacterial flagella (19). Flagellins are expressed by bacteria, particularly pathogenic bacteria, in the gut and the lung and activate epithelial TLR5 signaling (19–21). Flagellin expression is switched off as soon as bacteria translocate into the lamina propria (22). Detection of flagellin molecules represents therefore an alarm signal for subepithelial invasion and/or disruption of the epithelial barrier function. TLR5 signaling is rapidly induced in the lamina propria dendritic cells (DCs) of the small intestine (23). In the current study, we show that flagellin activates (via DCs) the splenic and mucosal production of IL-17 and IL-22 and the subsequent expression of target genes. This TLR5-mediated response was associated with a unique population of immune cells express-

^{*}Institut Pasteur de Lille, Centre d'Infection et d'Immunité de Lille; [†]Institut National de la Santé et de la Recherche Médicale Unité 1019; [†]Centre National de la Recherche Scientifique Unité Mixte de Recherche 8204; [§]Université Lille Nord de France; [¶]Institut de Recherche Interdisciplinaire, Centre National de la Recherche Scientifique Unité de Service et de Recherche 3078; [¶]Institut Fédératif de Recherche 142, Lille; ^{§§}Institut des Hautes Études Scientifiques, Bures-sur-Yvette; ^{††}University of Orléans and ^{‡‡}Centre National de la Recherche 6218, Institut de Transgenose, Orléans, France; [∥]Institute of Microbiology, Eidgenössiche Technische Hochschule Zurich, Zurich; **Ludwig Institute for Cancer Research, Lausanne Branch, University of Lausanne, Epalinges, Switzerland; and [#]Ludwig Institute for Cancer Research, Brussels Branch, de Duve Institute, Université catholique de Louvain, Brussels, Belgium

This work was supported by the Institut National de la Santé et de la Recherche Médicale (to. C.C., L.V.M., D.C., and J.-C.S.), the Institut Pasteur de Lille, the Université Lille Nord de France, and the Région Nord Pas de Calais (ARCir Europe). W.-D.H. and J.-C.S. are funded by the European Community (Grant INCO-CT-2006-032296).

Microarray data were deposited in the publicly available database (http://mace.ihes. fr) with accession number 2844328654.

Address correspondence and reprint requests to Dr. Jean-Claude Sirard, Center for Infection and Immunity of Lille, Institut National de la Santé et de la Recherche Médicale Unité 1019, Centre National de la Recherche Scientifique Unité Mixte de Recherche 8204, Institut Pasteur de Lille, 1, rue du Professeur Calmette, 59019 Lille Cedex, France. E-mail address: jean-claude.sirard@inserm.fr

The online version of this article contains supplemental material.

Abbreviations used in this paper: Ct, cycle threshold; DC, dendritic cell; DTX, diphtheria toxin; γ_c , common γ -chain; LTi, lymphoid tissue inducer; ns, nonsignificant; WT, wild-type.

Copyright © 2010 by The American Association of Immunologists, Inc. 0022-1767/10/\$16.00

ing CD127 but not CD3 that resembles LTi cells, LTi-like, or NK-like innate lymphocytes. Our findings suggest that CD3^{neg}CD127⁺ innate immune cells may be instrumental to the host's mucosal defense through the early production of Th17-related cytokines.

Materials and Methods

Mice

Specific pathogen-free mouse strains C57BL/6J, C57BL/6J-Ly5.1, BALB/ c, and $Tcrb^{-/-}$, $Tcrd^{-/-}$, $Tcrb^{-/-}Tcrd^{-/-}$, $Tlr5^{-/-}$ (24); $Myd88^{-/-}$ (25); transgenic animals for pre-TCR α , Cd11c-DTR-EGFP (Itagx-DTR/EGFP) (26); $Rag2^{-/-}Il2rg^{-/-}$ backcrossed on C57BL/6J mice; $Cd1d^{-/-}$ backcrossed on BALB/c background; and C.B-17 *scid* (SCID) mice were purchased from Charles River Laboratories (Wilmington, MA), The Jackson Laboratory (Bar Harbor, ME), or Janvier (St. Berthevin, France) or bred in an accredited establishment (number A59107; Institut Pasteur de Lille, Lille, France; Transgenose Institute Centre National de la Recherche Scientifique, Orleans, France; RCHCI at Eidgenössiche Technische Hochschule Zurich, Zurich, Switzerland; Ludwig Institute for Cancer Research, Brussels, Belgium; and Ludwig Institute for Cancer Research, Lausanne, Switzerland). Animals (6–16 wk old) were used according to national regulations and ethical guidelines.

For bone marrow (BM) chimera, recipient mice were irradiated (1000–1500 rad) and reconstituted 2–24 h later with BM cells (4–20 × 10⁶ cells i.v.). These mice were used at 10–16 wk posttransplantation, and the degree of chimerism was assessed by measuring CD45.1 and CD45.2 surface expression by leukocytes. The current protocol yielded 96.7% reconstitution for *Cd11c-DTR/EGFP*→C57BL/6 (wild-type [WT]) in spleen and 98.2% for *Thr5^{-/-}*→WT and 97.2% for WT→*Tlr5^{-/-}* in lung. Depletion of CD11c⁺ cells was achieved by injecting i.p. diphtheria toxin (DTX) as described previously (26). Depletion of γ 8 T cells (~90% depletion) and NK cells (~72% depletion) was performed by injecting i.p., 24 h prior to flagellin treatment, 100 µg mAb specific for TCR8-chain (GL3 clone) or NK1.1 (PK136) or irrelevant mAb HB152 as control.

Flagellin administration

LPS-depleted flagellin FliC from *Salmonella typhimurium* (5 μ g) produced as previously described (21), ultrapure LPS from *Escherichia coli* (serotype 0111:B4, 5 μ g; InvivoGen, Toulouse, France), or phosphorothioate CpG oligonucleotide (5'-TCCATGACGTTCCTGATGCT-3', 5 μ g; Eurogentec, Angers, France) diluted in PBS were injected i.v. or i.p to mice.

Flow cytometry and sorting

Spleens were digested with collagenase D (0.5 mg/ml; Roche, Basel, Switzerland) and DNase I (40 μ g/ml; Sigma-Aldrich, St. Louis, MO) during 10 min at 37°C. Cells were stained for CD127-FITC, CD45.1-FITC, CD45.2-PE, PE-conjugated lineage-specific Abs (CD3, B220, Gr1, CD11b, and CD11c), MHCII-PE, NK1.1-PercP-Cy5.5, CD4-APC, CD11c-APC, and CD45.1-Pacific Blue (BD Biosciences, San Jose, CA; BioLegend, San Diego, CA; and eBioscience, San Diego, CA) and sorted on a BD FACSAria.

DC culture

DCs were differentiated from BM as described previously (25). On day 7 or 11, BMDCs were stimulated for 2 h and analyzed.

Determination of cytokine production

CCL20 and IL-22 (R&D Systems, Minneapolis, MN) and IL-17A (eBioscience) levels were measured by ELISA in serum, and tissue homogenates were prepared with T-PER Reagent (Pierce, Rockford, IL) supplemented with protease inhibitors (Roche).

Gene expression

Total RNA was extracted with the Nucleospin RNA II kit (Macherey-Nagel, Hoerd, France) and reverse transcribed with the High-Capacity cDNA Archive Kit (Applied Biosystems, Foster City, CA). A preamplification of cDNA from sorted cells was performed prior to real-time PCR using the PreAmp kit (Applied Biosystems). cDNA was amplified using SYBR Green-based real-time PCR (Supplemental Table I) or commercial TaqMan assays (Applied Biosystems). For high-throughput analysis, TaqMan Low Density Arrays (Applied Biosystems) were used. Analysis was carried out using Real-Time StatMiner software (Integromics, Granada, Spain). Relative mRNA levels $(2^{-\Delta\Delta Ct})$ were determined by comparing 1) the cycle thresholds (Ct) for the gene of interest and *Actb* (Δ Ct) and 2) Δ Ct values for treated and control groups ($\Delta\Delta$ Ct). Ct upper limit was fixed to 33 cycles.

Microarrays

Total RNA (2 μ g) was processed on the Mouse Whole Genome Arrays version 2.0 (Applied Biosystems) (27). Data were analyzed using the NeONORM method, and heat maps were created as described previously (27, 28). Gene Ontology was analyzed using the Panther Protein Classification System (www.pantherdb.org).

Statistical analysis

The Mann-Whitney U test and the Graphpad Prism software 5.0 were used in analyses. The Limma test with Benjamini-Hochberg false discovery rate (FDR) correction was used for high-throughput PCR with TaqMan Low Density Arrays. Results were considered significant for p < 0.05 indicated by an asterisk. Results are expressed as arithmetic means \pm SD.

Results

Systemic TLR5 signaling enhances II17a, II17f, and II22 gene expression in lymphoid tissues

To establish whether TLR stimulation promotes the rapid expression of the Th17-related cytokines, mice were treated i.p. or i.v. with a TLR4 agonist (LPS), a TLR5 agonist (flagellin), or a TLR9 agonist (CpG). Gene expression in spleen and lymph nodes was then monitored (Fig. 1A, 1B). Flagellin administration triggered within 2 h ~1000-fold increase of *Il22* mRNA levels. Similarly, *Il17a* and *Il17f* gene expression was upregulated. A TLR5-mediated, Th17-related innate response was also observed in the mediastinal and inguinal lymph nodes and, to a lesser extent, in the liver (Fig. 1, Supplemental Fig. 1). LPS was initially shown to promote *Il22* expression in many tissues (29). We found that LPS also enhanced the Th17-related innate response but to a lesser extent than flagellin did (Fig. 1A, 1B, Supplemental Fig. 1). Although TLR9-mediated signaling activated the response in lymph nodes, it was devoid of any effect in the spleen.

The flagellin-dependent response was transient and peaked at 2 h; mRNA levels returned to baseline levels after 24 h (Fig. 1C, 1D). Gene profiling showed that the expression of genes specific for TLR-, IL-17R-, and IL-22R-mediated signaling was significantly enhanced in spleen (Fig. 1E). These genes encode pleiotropic and Th17-promoting cytokines (TNF- α , IL-1 β , and IL-6), chemokines that are specific for neutrophils, monocytes, and lymphocytes (CXCL-1, -2, -5, -9, and -10), antimicrobial molecules like CAMP and HAMP, lipocalin 2, S100A9, and tissue remodeling proteases matrix metalloproteinases 3 and 13. Strikingly, the expression of chemokine CCL20, which is specific for the recruitment of DC precursors, Th17 lymphocytes, LTi-like or NK-like cells (12, 14, 30, 31), was significantly upregulated. The transcription of the IFN- γ encoding gene was also upregulated by flagellin treatment. However, we did not observe any change in the expression of genes coding for IL-21, ROR γ t, or TGF- β (i.e., other factors involved in Th17 differentiation). As shown in Fig. 1F, IL-22 levels rose significantly in serum and spleen from flagellin-treated animals, whereas IL-17A was hardly detectable in serum but rose 3-fold in the spleen. In conclusion, TLR5 signaling in lymphoid tissues promotes the rapid production of the innate cytokines IL-17A, IL-17F, and IL-22—a pattern that resembles a Th17-related innate response.

TLR5-mediated innate responses require common γ -chaindependent immune cells

TLR5 is expressed by monocyte/macrophage/DC lineages, NK cells, CD4⁺ lymphocytes, and radioresistant stromal cells but not B lymphocytes (25, 32–35). To define the cells involved in the early production of IL-17 and IL-22, we used BM chimera expressing or not *Tlr5* and tested their ability to respond to flagellin. As shown in Fig. 2A and 2E, TLR5-competent hematopoietic cells were required to trigger flagellin-mediated *Il17* and *Il22* gene expression. By using SCID, *pre-TCRa* animals harboring enhanced number of $\gamma\delta$



FIGURE 1. TLR5 signaling promotes a swift, transient, Th17-related response. C57BL/6 mice (n = 3-4) were treated i.p. or i.v. with PBS, flagellin, LPS, or CpG. Tissues were assayed 2 h later (or as indicated) for cytokine production or mRNA levels. Spleen (*A*) and mesenteric lymph node (*B*) gene expression in response to i.v. injection of TLR agonists. Kinetic analysis of transcription in mesenteric (*C*) and mediastinal (*D*) lymph nodes after i.v. flagellin injection. *E*, Spleen transcriptional signature after i.p. flagellin injection. mRNA levels are expressed relative to the PBS group. *F*, Cytokine production in spleen extract (per whole spleen) and serum (per ml) after i.p. flagellin injection. The results are representative of two to four experiments. Results are given as means \pm SD. *p < 0.05 in Mann-Whitney *U* test relative to PBS group. In *E*, all genes display p < 0.05 in Limma test.

lymphocytes, $Tcrb^{-/-}$ and $Tcrd^{-/-}$ mice, we demonstrated that neither B cells nor TCR $\alpha\beta$ -/TCR $\gamma\delta$ -expressing T lymphocytes were required for flagellin-mediated *Il17* and *Il22* expression (Fig. 2*B*, Supplemental Fig. 2*A*, 2*B*, 2*D*). In contrast, the response was impaired



FIGURE 2. The IL receptor γ_c is required for the TLR5-mediated, Th17related innate response. Mice (n = 3-4) were treated i.p. with PBS or flagellin. Lymphoid tissues and serum were sampled 2 h after injection for the quantification of mRNA levels and ELISA. *A*, Response in mediastinal lymph nodes from C57BL/6 (WT) \rightarrow *Tlr5^{-/-}* and *Tlr5^{-/-}* \rightarrow WT BM chimera. The spleen response in SCID (*B*) and *Rag2^{-/-}Il2rg^{-/-}* (*C*) mice. *D*, The flagellin-mediated liver response. mRNA levels are expressed relative to the PBS group. *E*, IL-22 production in serum. Results are given as means \pm SD. **p* < 0.05 in Mann-Whitney *U* test relative to PBS group.

in $Rag2^{-/-1/2}rg^{-/-}$ mice that have almost normal DCs but lack B and T lymphocytes as well as NK, NKT, LTi, NK-like, and LTi-like cells, all of which depend on the IL receptor common γ -chain (γ_c) encoded by ll2rg gene (Fig. 2C, 2E) (12, 15, 17, 36). The impairment was not a collapse of TLR5 signaling, because the liver was still responsive to flagellin (Fig. 2D). Our experiments using genetically deficient animals and depleting Abs suggested that CD1d-restricted NKT cells or NK cells were not drivers of the TLR5-mediated response (Supplemental Fig. 2C, 2D). Therefore, our data showed that the cells expressing Th17-related cytokines after flagellin stimulation require the IL receptor γ_c for differentiation or activation but are not NK, NKT, or TCR $\gamma\delta$ -expressing innate lymphocytes.

CD3^{neg}CD127⁺ LTi-like cells produce IL-17 and IL-22 in response to TLR5 signaling

We next sought to determine which γ_c -dependent innate immune cells are involved in TLR5-mediated response. Recent work has suggested that innate lymphocytes expressing IL-7R (i.e., the CD127 or IL-7R α -chain and the CD132 or γ_c) and LTi/NK cell markers are sources of IL-22 and IL-17 (12–17). To determine whether innate immune cells produce Th17-related cytokines, splenic cells from mock- and flagellin-treated animals were sorted on the basis of lineage markers (CD11b, CD11c, Gr1, CD3, and B220), NK1.1, CD4, and CD127 (Fig. 3A). After TLR5 stimulation, Lin^{neg}NK1.1^{neg}CD127⁺CD4^{neg} and Lin^{neg}NK1.1^{neg}CD127⁺



FIGURE 3. Lin^{neg}CD127⁺ innate cells upregulate *ll17* and *ll22* expression following TLR5 stimulation. C57BL/6 (*A*–*C*) or SCID (*D*, *E*) mice (*n* = 3–4) were treated i.p. with PBS or flagellin. Spleens were sampled after 2 h for cell sorting and quantification of mRNA levels. *A*, Sorting scheme. Cells were sorted as Lin⁺CD4⁺, Lin⁺NK1.1⁺, and Lin^{neg}NK1.1^{neg}CD127⁺CD4⁺ cells and Lin^{neg}NK1.1^{neg}CD127⁺CD4^{neg} cells. *B*, Regulation of *ll17f* and *ll22* expression following flagellin stimulation. mRNA levels and statistical significance are expressed relative to PBS. *C*, Levels of transcripts in sorted cells following flagellin stimulation. mRNA levels and statistical significance are expressed relative to Lin⁺CD4⁺cells. The data correspond to the mean of three independent sorting. Results are given as means ± SD. **p* < 0.05 in Mann-Whitney *U* test. *D* and *E*, Spleens cells were used such as (input) or sorted as CD3^{neg}CD4⁺CD127⁺, DC (CD11c⁺MHCII⁺), or NK (NKp46⁺). In *D*, mRNA levels in flagellin-treated animals are expressed relative to PBS. In *E*, mRNA levels levels are expressed relative to input cells.

CD4⁺ cells were found to strongly upregulate *Il17f* and *Il22* expression ~80- and 500-fold, respectively (Fig. 3B). These subsets account for ~0.5-1% of Lin^{neg}NK1.1^{neg} cells and 0.02-0.1% of splenocytes in a C57BL/6 mouse. Similar observations were done using CD3^{neg}CD127⁺CD4⁺ cells from SCID mice (Fig. 3D). Both subsets were found to express CD45, indicating a hematopoietic origin and their classification as immune cells (Supplemental Fig. 3A). We also found that the CD3^{neg}CD127⁺CD4⁺ and CD3^{neg} CD127⁺CD4^{neg} cells are mostly absent in the Rag2^{-/-}Il2rg^{-/-} mouse (Supplemental Fig. 3D). This is consistent with previous observations (12, 17). In spleen, the frequency of Lin^{neg}NK1.1^{neg} CD127⁺CD4⁺ matched the number of LTi cells (12, 37). We also detected upregulation of Il22 transcript levels in CD4 lymphocytes (Lin⁺CD4⁺) and NK cells (Lin⁺NK1.1⁺) after flagellin administration. However, the Lin^{neg}NK1.1^{neg}CD127⁺CD4⁺ cells were the most potent producers, because Il17f and Il22 mRNA levels were 100- to 1000-fold higher in this population than in Lin^{neg}NK1.1^{neg} CD127⁺CD4^{neg}, CD4, or NK cells (Fig. 3C). Like the NK-like or LTi-like cells, Lin^{neg}NK1.1^{neg}CD127⁺CD4⁺ innate lymphocytes

TLR-MEDIATED INNATE IMMUNE CELL ACTIVATION

express *Ccr6*, but we were unable to detect any mRNA for *Rorc* or *Ncr1* encoding ROR γ t and NKp46. In addition, we observed that, upon flagellin stimulation, *Ifng* expression was mainly upregulated in NK cells rather than Lin^{neg}CD127⁺ cells (data not shown). Our data provide evidence that Lin^{neg}CD127⁺ innate immune cells, and especially the CD4⁺ fraction, which resembles LTi cells, can rapidly produce IL-17 and IL-22 cytokines following TLR activation.

TLR5-mediated activation of Th17-related innate responses requires DCs

To determine whether the TLR5-mediated upregulation of Th17related innate response is a direct innate lymphocyte activation process or requires DC stimulation, DTX-mediated ablation of CD11c⁺ cells was performed in a *Cd11c-DTR-EGFP* BM chimera. The DTX treatment depleted 93.8 \pm 2.2% of CD11c⁺MHCII^{high} DCs (Fig. 4*A*). In addition, we found that DTX treatment did not eliminate the CD3^{neg}CD4⁺CD127⁺ cells in the spleen of a *Cd11c-DTR-EGFP* chimera and did not alter the TLR5-mediated production of IL-22 (Supplemental Fig. 3). As shown in Fig. 4*B*, systemic administration of flagellin to DC-depleted animals resulted in impaired *Il17f or Il22* transcription, compared with controls. Thus, our experiments demonstrated that DCs are necessary for TLR5mediated expression of Th17-related cytokines.



FIGURE 4. DCs regulate TLR5-dependent, Th17-related innate responses. *Cd11c-DTR-EGFP* \rightarrow C57BL/6 chimera mice (n = 4) were injected i.p. with 100 ng DTX or untreated (no DTX). *A*, Depletion of splenic DCs was assessed 24 h later by flow cytometry. *B* and *C*, The flagellin-specific transcriptional response. Twenty-four hours after DTX injection, mice were treated i.p. with PBS or flagellin. Spleen was sampled after 2 h for the quantification of transcripts specific for Th17-related cytokine genes (*B*) and DC-specific cytokine genes (*C*). mRNA levels are expressed relative to the PBS \rightarrow no DTX group. *D*, BM-derived DCs from C57BL/6 mice were treated or not with LPS or flagellin (1 µg/ml) for 2 h, and mRNA levels were determined relative to mock DCs. Statistical significance was expressed relative to no DTX (*B*, *C*) or relative to the mock DCs (*D*). Results are given as means \pm SD. *p < 0.05 in Mann-Whitney *U* test.

Th17 differentiation depends on tissue-derived TGF-B and IL-1B, and IL-6 produced by DCs and the maintenance of Th17 phenotype have been associated to DC-derived IL-23 (9). In vivo, DC cell ablation was found to attenuate the upregulation of Il6 transcription in response to flagellin (Fig. 4C). We did not observe any alteration in Illb, Ill2b coding for the p40 chain of IL-12 or IL-23, and *Tgfb* gene expression (data not shown). Intriguingly, Il23a (coding for the p19 chain of IL-23) transcription was enhanced by DTX treatment; it is possible that CD11c^{low}MHCII^{neg} cells having infiltrated the spleen in DTX-treated mice support Il23a upregulation (Fig. 4A, 4C). In this study, we found that flagellin promoted the expression of both Il12b and Il23a in BM-derived DCs (Fig. 4D). In response to flagellin, DCs can therefore produce both IL-6 and IL-23, with the subsequent expression of an innate, Th17-related signature. However, in vivo, IL-23 alone appears unable to induce gene activation. In response to flagellin, DC-mediated IL-6 production could therefore promote the Th17-related signature by Lin^{neg}CD127⁺ cells.

Previous studies suggested that DCs produce IL-22 (4, 38). In this study, we found that flagellin promoted the expression of *Il12b* in BM-derived DCs but did not have any effect on *Il17a*, *Il17f*, and *Il22* transcription (Fig. 4D). LPS, which strongly increased *Il12b* and *Il23a* transcription, did not upregulate expression of *Il17a* and *Il17f* in DCs but did enhance *Il22* mRNA levels ~8-fold, compared with untreated cells. Analysis of transcription in splenic DCs sorted from mice treated with flagellin also showed that DCs were not potent source of IL-22 (Fig. 3D, 3E). Taken as a whole, our data suggested that DCs are not a major source of TLR5-mediated, Th17-related cytokine production but contribute to induction of the latter.

TLR5 signaling triggers an intestinal, Th17-related, innate response

Because Th17-related cytokines are important in the control of inflammation and infection in the mucosa (3-6, 38, 39), we next assessed the impact of flagellin administration on intestinal tissues. Flagellin strongly enhanced the production of IL-22 within 2 h of administration from the duodenum to the proximal colon; the level then returned to the baseline at 8 h (Fig. 5). CCL20 synthesis was also strongly induced in the small intestine (Supplemental Fig. 4A). In contrast, gut IL-17A production changed moderately following flagellin treatment. In any intestine segments, TLR5 signaling strongly enhanced transcription of the Il17 and Il22 genes and those encoding antimicrobial peptides, neutrophil-specific chemokines and growth factors, and tissue remodeling/repair molecules (Fig. 5C, Supplemental Fig. 4B-E). A similar pattern was also observed in lung tissue, but regression to baseline levels was not observed at 8 h, suggesting that the kinetic may be different in the respiratory tract (Supplemental Fig. 5). In conclusion, systemic flagellin administration promotes mucosal Th17-related innate responses.

We used the same type of analysis than for the response in spleen and lymph nodes to investigate the effects of flagellin on mucosa. The gut and pulmonary flagellin-mediated Th17-like innate response was stronger in WT \rightarrow *Tlr5^{-/-}* than in *Tlr5^{-/-}* \rightarrow WT animals (Fig. 5*D*, Supplemental Fig. 5). These findings indicated that TLR5-competent hematopoietic cells are also instrumental in the mucosal response. DC depletion impaired the upregulation of the intestinal Th17-related signature (Supplemental Fig. 4*F*). Last, we compared Th17-related gene expression in *Rag2^{-/-}Il2rg^{-/-}* and WT animals (Fig. 5*E*). As previously reported, steady-state levels of *Il22*, *Il17*, and *Ifng* transcripts in the ileum were significantly lower in *Rag2^{-/-}Il2rg^{-/-}* than in WT animals (17). After TLR5 stimulation, transcription of these genes was not enhanced in the *Rag2^{-/-}Il2rg^{-/-}* intestine. Taken as a whole, our data suggest that TLR5 signaling elicits a mucosal Th17-related response by innate immune cells that require the γ_c and help from DCs.

Finally, the expression of many genes was upregulated in the gut of $Rag2^{-/-}Il2rg^{-/-}$ animals, suggesting that TLR5 signaling promotes the activation of non-lymphoid cells (Fig. 5*E*). Experiments with the $Tlr5^{-/-}$ chimera reinforced these conclusions because radioresistant cells, likely structural/stromal cells, contributed to TLR5-mediated responses, especially in lung tissue (Fig. 5*D*, Supplemental Fig. 5).

The flagellin-induced transcriptional signature has features of Th17- and TLR-mediated responses

To further characterize the response to flagellin, we performed a microarray time course analysis of gene expression in the distal ileum of treated mice. Biological processes including signaling and defense pathways, cytokine- and chemokine-mediated immunity, NF-kB signaling, granulocyte- and macrophage-mediated immunity, cell proliferation and differentiation, and apoptosis, were significantly modulated 2-8 h after flagellin administration (Supplemental Fig. 6, Supplemental Table II). On the basis of these data, we arbitrarily defined five groups of genes according to their potential role in the gut immune response (Table I). The first group of genes encodes modulators of IL-17 an IL-22 production as well as differentiation of Th17, LTi, LTi-like, NK, and NK-like cells. The second class was involved in the positive or negative transcriptional control of IL-17R (i.e., C/EBPB and C/EBPb), IL-22R (i.e., SOCS-1, SOCS-3, and STAT5A) and TLR (i.e., activating transcription factor 3 and I- $\kappa B\alpha$) signaling. The third group includes TLR, IL-17R, and IL-22R signaling target genes coding for antimicrobial molecules or factors regulating epithelium barrier function. Interestingly, our analysis also identified antiviral genes. In contrast to previous studies of Th17 cytokine-mediated signaling, we did not find any transcriptional modulation of the mucin and β-defensin genes. The fourth group corresponds to genes encoding factors involved in recruitment, development, or function of various immune cells. Relative to controls, chemokines specific for innate immune cells like neutrophils, NK, NKT, or LTi cells, T and B lymphocytes, or monocytes were significantly enriched-suggesting that several cell types may enter the tissues and participate in the immune response. The last group includes genes required for TLR4 or TLR2 signaling (i.e., LBP, CD14, and MAL). In conclusion, intestinal gene expression profiling showed that flagellin promotes a transient immune response involving regulators and effectors of both TLR- and Th17mediated immunity.

Discussion

The way in which TLR signaling activates the host's innate defenses during mucosal invasion by pathogens is subject to debate. In the current study, we showed that TLR5 signaling induces systemic and mucosal innate expression of the Th17-related IL-17 and IL-22 cytokines by stimulating Lin^{neg}CD127⁺ cells in a DC-dependent manner. Overall, our data suggest that Lin^{neg}CD127⁺ cells may play a major role as innate lymphocytes in the early orchestration of a TLR-dependent, protective response to mucosal invasion by pathogens.

Th17-related cytokines contribute to adaptive immunity in response to various inflammatory and infectious diseases (5, 6, 30, 39); however, their impact on the early phase of infection is poorly understood. The effect of TLR signaling on IL-17 and IL-22 was previously suggested because their synthesis was enhanced after administration of TLR2 and TLR4 agonists (12, 29). Very recently, TLR5 signaling was also associated to such response (40). In this study, we analyzed the immune response to the TLR5 activator



FIGURE 5. TLR5 signaling promotes γ_c -dependent, Th17-related innate responses in the gut. Mice (n = 3-4) were treated i.p. with PBS or flagellin, and gut segments (5 cm for proximal duodenum, jejunum, and distal ileum; 3 cm for proximal colon and the whole cecum) were sampled at the indicated times for quantification of cytokine production and mRNA levels. *A* and *B*, Cytokine production in C57BL/6 (WT) animals. Tissue segments were homogenized within 1 ml lysis buffer. The results are representative of two experiments. *C*, Time course analysis of the transcriptional response in

TLR-MEDIATED INNATE IMMUNE CELL ACTIVATION

flagellin, with a focus on mucosa. Our rationale was that because flagellin expression is specifically restricted to luminal compartment, its desequestration is likely to be an alarm signal for mucosal invasion (22). We found that systemic flagellin administration promotes the swift, intense, transient production of IL-17A, IL-17F, and IL-22 and factors controlling Th17 differentiation. For example, flagellin modulates expression of the genes coding for the aryl hydrogen receptor repressor, the aryl hydrogen receptor nuclear translocator-like factor ARNTL and the activating transcription factorlike factor BATF-all of which are involved in Th17 differentiation (41, 42). Moreover, the flagellin-induced innate response and the Th17 adaptive response share many effectors, such as chemokines, antimicrobial peptides, antiapoptotic factors, and tissue remodeling factors. Therefore, we hypothesized that flagellin-mediated response enables the rapid and transient recruitment of systemic and mucosal defenses.

Our results suggest that the Lin^{neg}CD127⁺ cells, especially the CD4⁺ fraction, have a pivotal role in the TLR-mediated response via the production of IL-17 and IL-22. These cells are similar to LTi (CD3^{neg}CD127⁺CD4⁺), LTi-like (RORyt⁺Lin^{neg}CD127⁺CD4⁺), and NK-like $(ROR\gamma t^+NKp46^+Lin^{neg}CD127^+NK1.1^{+/neg})$ cells (12, 14-17). LTi-like/NK-like cells constitutively produce Th17related cytokines in a process that depends on gut flora, γ_c , and RORyt (15, 17). Moreover, the LTi-like cells were shown to upregulate the production of IL-17 and IL-22 in response to microbial products. The development of Lin^{neg}CD127⁺CD4⁺ and Lin^{neg} CD127⁺CD4^{neg} cells identified in this study requires the γ_c . These subsets also express the CCR6 but not the NKp46 encoding gene, suggesting a common ontogeny with LTi and LTi-like cells (12, 13). We were unable to detect any expression of RORyt in the Lin^{neg} CD127⁺ cells; expression below our assay's detection threshold is one possible explanation for this failure. Furthermore, Rorc expression was downregulated in gut after flagellin injection. Interestingly, we noted enhanced intestinal expression of the gene encoding NFIL3-a factor that is essential for NK cell development (43). Recent studies demonstrated that increased expression of IL-7 enhances the number of LTi cells (37) and that deficiency in IL-7 affects the number of NK-like IL-22 expressing population (44). Additional work will be needed to define the ontogeny and transcriptional factors involved in the differentiation of Lin^{neg} CD127⁺ cells.

DCs have an important role in integrating microbial signals and activating immune cells like Th17 lymphocytes (9). When DCs were depleted, the Th17-related innate response to flagellin was impaired, indicating that DCs are necessary for the activation of Lin^{neg}CD127⁺ cells. Similarly, transcription of *ll6* was attenuated, which suggests that, in Th17 differentiation, DC-derived IL-6 like may contribute to Lin^{neg}CD127⁺ cell activation (9). Our findings suggested that IL-1 β , IL-23, or TGF- β (or at least the amounts produced by DCs) are not required for Lin^{neg}CD127⁺ cell activation. However, IL-23 might be important for an alternative activating pathway for production of IL-17 or IL-22 by LTi-like and NK-like cells because 1) cells are activated in vitro by supplementing the culture medium with IL-23 and 2) CD3^{neg}CD4⁺

WT distal ileum. *D*, Contribution of TLR5 signaling in the hematopoietic compartment is essential for the Th17-related signature. mRNA levels 2 h after flagellin treatment are expressed as $WT \rightarrow Tlr5^{-/-}$ relative to $Tlr5^{-/-}$ \rightarrow WT mice. *E*, Contribution of the γ_c to intestinal response. mRNA levels are expressed relative to PBS-treated WT mice. In *A*, *B*, and *D*, all data displayed statistically significant changes, except where denoted as "ns" for "nonsignificant." Results are given as means \pm SD. Mann-Whitney *U* test relative to PBS group was used in *A* and *B*. Limma test was used in *C*–*E*.

		Expression		
Group	Genes	2 h	8 h	Contribution to
Ι	Ahrr	_	§	Transcriptional regulation of <i>IL-22</i> gene
	Arntl	0	_	Transcriptional regulation of <i>IL-22</i> gene
	Batf	+	0	Master regulation of Th17 differentiation
	Il1b, Il6, Il1r1, Tnfsf15	+	0	Activation of Th17 differentiation
	Il17f, Il22	+	+	Th17 cell phenotype
	Nfil3	+	+	Regulation of NK cell differentiation
	Pparg	0	—	Inhibition of Th17 differentiation
	Rorc	_	_	Master regulation of Th17, Treg, LTi, LTi-, and NK-like cell differentiation
II	Atf3, Irf5, Map3k8, Nfkb1	+	0	Activation of IL-17R and TLR signaling
	Cebpb, Cebpd	+	+	Activation of IL-17R signaling
	Dusp1, Nfkbia, Nfkbie, Zfp36	+	0	Repression of IL-17R and TLR signaling
	Irf4	+	+	Activation of IL-17R and TLR signaling
	Socs1, Socs3	+	+	Activation of IL-22R signaling
	Stat5a	+	0	Activation of IL-17R and IL-22R signaling
III	Adm, Areg, Ereg, Nrg1, Sprr1a, Sprr2i, Tff2	+	+	Epithelial barrier function
	Bcl3, Birc3	+	0	Inhibitor of apoptosis
	Hamp, lipocalin 2, S100a8, S100a9	+	+	Antimicrobial activity
	Isg15, Isg20, Oas2, Oas3	+	+	Antiviral activity
	Ncf1, Nox1, Sod2	+	0	Antimicrobial activity via superoxides
	Retlna, Retlng	+	0	Inflammation control
	Rnase1,S100a11	+	+	Antimicrobial activity?
	Saa2	0	+	Acute phase protein
	Saa3	+	+	Acute phase protein
	Serpine1, Serpina3g, Timp1	+	+	Inflammation control
IV	Adamts4, Adamts3, Adamts8	+	0	Tissue remodeling
	Ccl2, Ccl3, Ccl4	+	+	Monocyte recruitment
	Ccl17, Ccl22	+	0	T cell recruitment
	Cxcl1, Cxcl2, Cxcl5	+	+	Granulocyte recruitment
	Ccl7, Cxcl9, Cxcl10, Cxcl11	+	0	Immune cells recruitment
	Cxcl13	0	+	B and LTi cell recruitment
	Ifng, Ifn-regulated genes	+	0	Activation of immunity
	Ltb	+	0	Lymphotoxin β receptor signaling in stromal cells
	Mmp3, Mmp13	+	+	Tissue remodeling
	Tnfrsfla, Tnfrsflb	+	0	TNFRI and TNFRII signaling
	Vcam1	+	+	LTi-mediated organization of lymphoid tissues
V	Cd14	+	+	LPS binding and TLR4 signaling
	Lbp	0	+	LPS binding and TLR4 signaling
	Mal	0	+	Activation of TLR2 and TLR4 signaling
	Tlr2, Pglyprp1	+	0	Bacterial cell wall signaling

Table I. Flagellin-induced gene expression in the ileum displays signatures that are reminiscent of Th17- and TLR-mediated responses

Microarray gene expression data from the distal ileal segments of mice (n = 3) treated i.p. with flagellin for 2 and 8 h were compared with data from mock animals in a time course analysis. All genes differentially expressed and the corresponding method and annotation are included as Supplemental Fig. 5 and Supplemental Table II.

o, similar expression relative to mock animals; +, upregulation relative to mock animals; -, downregulation relative to mock animals; §, data from a comparison between mock and 2-h flagellin-treated animals.

cells express IL-23R (12, 13, 15, 45). In the intestine, TLR5 signaling activates lamina propria DCs, which then promote Th17 differentiation (23). In contrast, intestinal DCs do not respond to TLR4 stimulation (23, 46). Flagellin treatment enhanced the transcription of the genes coding for CD14, LPB, MAL, and TLR2 (Supplemental Fig. 2, Table I). These findings suggest that responsiveness to TLR2 and TLR4 agonists may be reactivated or amplified after TLR5 stimulation, allowing the production of a second wave of effectors.

The relevance of TLR5 signaling in defense has recently been assessed. Flagellin-mediated protection of rodents and nonhuman primates against lethal irradiation was associated with CSF3-mediated granulopoiesis and the antiapoptotic effect of superoxide dismutase 2 (47). Flagellin treatment has been linked to resistance against inflammatory colitis and gut infections (40, 48). The TLR5-induced circulating and local production of IL-17/IL-22 may be the main driving force behind these protective effects. The contribution of IL-22 was recently suggested as instrumental in the control of infection with enterococci (40). In response to IL-

17R and IL-22R signaling, epithelial and stromal cells produce antimicrobial peptides (RegIII), CXC chemokines, and growth factors (CSF3) for neutrophils, all of which are involved in mucosal protection (3–7, 38). Flagellin treatment prompted the expression of similar factors. In addition, our study identified other potential effectors of the TLR/IL-17R/IL-22R axis, such as antiviral molecules (ISG15, ISG20, OAS2, and OAS3), acute-phase proteins (SAA2, SAA3, and PTX3), and superoxide-mediated killing (NCF1, NOX1, and superoxide dismutase 2).

The Lin^{neg}CD127⁺CD4⁺ cells described in this paper resemble the LTi cells that are instrumental in the development of secondary/tertiary lymphoid tissues (such as Peyer's patches or isolated lymphoid follicles) (13, 49). Intestinal LTi cells express lymphotoxin β and CCR6 and CXCR5, the receptors for CCL20 and CXCL13, respectively (13, 49). CXCL13 and CCL20 are produced by epithelial/stromal cells and are involved in the clustering of lymphocytes, DCs, and LTi to form lymphoid follicles (50, 51). Ectopic epithelial expression of CXCL13 increases the number of LTi cells (which produce IL-22 constitutively) and isolated lymphoid follicles (13). Our work revealed that flagellin upregulated gut expression of genes coding for lymphotoxin β , CXCL13, and CCL20. We previously showed that flagellin triggers CCL20 production in intestinal epithelial cells (31). Hence, TLR5 signaling in both epithelial and hematopoeitic cells may increase the development of secondary/tertiary lymphoid tissues. Recent studies showed that TLR2, TLR4, and Nod1 are involved in the development of lymphoid follicles (50, 51). Pattern recognition receptors in general and TLRs in particular may have a pivotal role in simultaneously conditioning the antimicrobial environment and new ectopic sites for the development of mucosal adaptive immunity.

In conclusion, the current study found that Lin^{neg}CD127⁺ cells constitute a rapidly reacting, innate source of IL-17A, IL-17F, and IL-22 in response to TLR signaling. We hypothesize that this immune reaction occurs during microbial penetration into the lamina propria and stimulates innate effectors to locally clear the infection. Similar cell populations have been identified in humans (14, 16), and so, it remains to be seen whether TLR stimulation can promote activation of these innate immune cells.

Acknowledgments

We thank Michel Simonet and François Trottein for critical reading of the manuscript, Shizuo Akira for $Tlr5^{-/-}$ mice, David Dombrowicz for $Rag2^{-/-1}ll2rg^{-/-}$ mice, and Christelle Faveeuw and Laxmi Koodun for technical assistance.

Disclosures

The authors have no financial conflicts of interest.

References

- Beutler, B., Z. Jiang, P. Georgel, K. Crozat, B. Croker, S. Rutschmann, X. Du, and K. Hoebe. 2006. Genetic analysis of host resistance: Toll-like receptor signaling and immunity at large. *Annu. Rev. Immunol.* 24: 353–389.
- Sansonetti, P. J. 2004. War and peace at mucosal surfaces. Nat. Rev. Immunol. 4: 953–964.
- Aujla, S. J., Y. R. Chan, M. Zheng, M. Fei, D. J. Askew, D. A. Pociask, T. A. Reinhart, F. McAllister, J. Edeal, K. Gaus, et al. 2008. IL-22 mediates mucosal host defense against Gram-negative bacterial pneumonia. *Nat. Med.* 14: 275–281.
- Zheng, Y., P. A. Valdez, D. M. Danilenko, Y. Hu, S. M. Sa, Q. Gong, A. R. Abbas, Z. Modrusan, N. Ghilardi, F. J. de Sauvage, and W. Ouyang. 2008. Interleukin-22 mediates early host defense against attaching and effacing bacterial pathogens. *Nat. Med.* 14: 282–289.
- Conti, H. R., F. Shen, N. Nayyar, E. Stocum, J. N. Sun, M. J. Lindemann, A. W. Ho, J. H. Hai, J. J. Yu, J. W. Jung, et al. 2009. Th17 cells and IL-17 receptor signaling are essential for mucosal host defense against oral candidiasis. *J. Exp. Med.* 206: 299–311.
 Ishigame, H., S. Kakuta, T. Nagai, M. Kadoki, A. Nambu, Y. Komiyama,
- Ishigame, H., S. Kakuta, T. Nagai, M. Kadoki, A. Nambu, Y. Komiyama, N. Fujikado, Y. Tanahashi, A. Akitsu, H. Kotaki, et al. 2009. Differential roles of interleukin-17A and -17F in host defense against mucoepithelial bacterial infection and allergic responses. *Immunity* 30: 108–119.
- Gaffen, S. L. 2009. Structure and signalling in the IL-17 receptor family. Nat. Rev. Immunol. 9: 556–567.
- Kolls, J. K., P. B. McCray, Jr., and Y. R. Chan. 2008. Cytokine-mediated regulation of antimicrobial proteins. *Nat. Rev. Immunol.* 8: 829–835.
- Korn, T., E. Bettelli, M. Oukka, and V. K. Kuchroo. 2009. IL-17 and Th17 cells. Annu. Rev. Immunol. 27: 485–517.
- 10. Shibata, K., H. Yamada, H. Hara, K. Kishihara, and Y. Yoshikai. 2007. Resident $V\delta 1^+ \gamma \delta$ T cells control early infiltration of neutrophils after *Escherichia coli* infection via IL-17 production. *J. Immunol.* 178: 4466–4472.
- Michel, M. L., A. C. Keller, C. Paget, M. Fujio, F. Trottein, P. B. Savage, C. H. Wong, E. Schneider, M. Dy, and M. C. Leite-de-Moraes. 2007. Identification of an IL-17-producing NK1.1^{neg} iNKT cell population involved in airway neutrophilia. *J. Exp. Med.* 204: 995–1001.
- Takatori, H., Y. Kanno, W. T. Watford, C. M. Tato, G. Weiss, I. I. Ivanov, II, D. R. Littman, and J. J. O'Shea. 2009. Lymphoid tissue inducer-like cells are an innate source of IL-17 and IL-22. J. Exp. Med. 206: 35–41.
- Marchesi, F., A. P. Martin, N. Thirunarayanan, E. Devany, L. Mayer, M. G. Grisotto, G. C. Furtado, and S. A. Lira. 2009. CXCL13 expression in the gut promotes accumulation of IL-22-producing lymphoid tissue-inducer cells, and formation of isolated lymphoid follicles. *Mucosal Immunol.* 2: 486–494.
 Cella, M., A. Fuchs, W. Vermi, F. Facchetti, K. Otero, J. K. Lennerz,
- Cella, M., A. Fuchs, W. Vermi, F. Facchetti, K. Otero, J. K. Lennerz, J. M. Doherty, J. C. Mills, and M. Colonna. 2009. A human natural killer cell subset provides an innate source of IL-22 for mucosal immunity. *Nature* 457: 722–725.

- Sanos, S. L., V. L. Bui, A. Mortha, K. Oberle, C. Heners, C. Johner, and A. Diefenbach. 2009. RORγt and commensal microflora are required for the differentiation of mucosal interleukin 22-producing NKp46⁺ cells. *Nat. Immunol.* 10: 83–91.
- Cupedo, T., N. K. Crellin, N. Papazian, E. J. Rombouts, K. Weijer, J. L. Grogan, W. E. Fibbe, J. J. Cornelissen, and H. Spits. 2009. Human fetal lymphoid tissueinducer cells are interleukin 17-producing precursors to RORC⁺CD127⁺ natural killer-like cells. *Nat. Immunol.* 10: 66–74.
- Satoh-Takayama, N., C. A. Vosshenrich, S. Lesjean-Pottier, S. Sawa, M. Lochner, F. Rattis, J. J. Mention, K. Thiam, N. Cerf-Bensussan, O. Mandelboim, et al. 2008. Microbial flora drives interleukin 22 production in intestinal NKp46⁺ cells that provide innate mucosal immune defense. *Immunity* 29: 958–970.
- Luci, C., A. Reynders, I. I. Ivanov, C. Cognet, L. Chiche, L. Chasson, J. Hardwigsen, E. Anguiano, J. Banchereau, D. Chaussabel, et al. 2009. Influence of the transcription factor RORγt on the development of NKp46+ cell populations in gut and skin. *Nat. Immunol.* 10: 75–82.
- Ramos, H. C., M. Rumbo, and J. C. Sirard. 2004. Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa. *Trends Microbiol.* 12: 509–517.
- Bambou, J. C., A. Giraud, S. Menard, B. Begue, S. Rakotobe, M. Heyman, F. Taddei, N. Cerf-Bensussan, and V. Gaboriau-Routhiau. 2004. In vitro and ex vivo activation of the TLR5 signaling pathway in intestinal epithelial cells by a commensal *Escherichia coli* strain. J. Biol. Chem. 279: 42984–42992.
- Didierlaurent, A., J. Goulding, S. Patel, R. Snelgrove, L. Low, M. Bebien, T. Lawrence, L. S. van Rijt, B. N. Lambrecht, J. C. Sirard, and T. Hussell. 2008. Sustained desensitization to bacterial Toll-like receptor ligands after resolution of respiratory influenza infection. J. Exp. Med. 205: 323–329.
- Hughes, E. A., and J. E. Galán. 2002. Immune response to Salmonella: location, location, location? Immunity 16: 325–328.
- Uematsu, S., K. Fujimoto, M. H. Jang, B. G. Yang, Y. J. Jung, M. Nishiyama, S. Sato, T. Tsujimura, M. Yamamoto, Y. Yokota, et al. 2008. Regulation of humoral and cellular gut immunity by lamina propria dendritic cells expressing Toll-like receptor 5. *Nat. Immunol.* 9: 769–776.
- Uematsu, S., M. H. Jang, N. Chevrier, Z. Guo, Y. Kumagai, M. Yamamoto, H. Kato, N. Sougawa, H. Matsui, H. Kuwata, et al. 2006. Detection of pathogenic intestinal bacteria by Toll-like receptor 5 on intestinal CD11c⁺ lamina propria cells. *Nat. Immunol.* 7: 868–874.
- Didierlaurent, A., I. Ferrero, L. A. Otten, B. Dubois, M. Reinhardt, H. Carlsen, R. Blomhoff, S. Akira, J. P. Kraehenbuhl, and J. C. Sirard. 2004. Flagellin promotes myeloid differentiation factor 88-dependent development of Th2-type response. J. Immunol. 172: 6922–6930.
- Hapfelmeier, S., A. J. Müller, B. Stecher, P. Kaiser, M. Barthel, K. Endt, M. Eberhard, R. Robbiani, C. A. Jacobi, M. Heikenwalder, et al. 2008. Microbe sampling by mucosal dendritic cells is a discrete, MyD88-independent step in DeltainvG S. Typhimurium colitis. J. Exp. Med. 205: 437–450.
- Eilebrecht, S., F. X. Pellay, P. Odenwälder, G. Brysbaert, B. J. Benecke, and A. Benecke. 2008. EBER2 RNA-induced transcriptome changes identify cellular processes likely targeted during Epstein Barr Virus infection. *BMC Res Notes* 1: 100.
- Noth, S., G. Brysbaert, and A. Benecke. 2006. Normalization using weighted negative second order exponential error functions (NeONORM) provides robustness against asymmetries in comparative transcriptome profiles and avoids false calls. *Genomics Proteomics Bioinformatics* 4: 90–109.
- Dumoutier, L., E. Van Roost, D. Colau, and J. C. Renauld. 2000. Human interleukin-10-related T cell-derived inducible factor: molecular cloning and functional characterization as an hepatocyte-stimulating factor. *Proc. Natl. Acad. Sci.* USA 97: 10144–10149.
- Hirota, K., H. Yoshitomi, M. Hashimoto, S. Maeda, S. Teradaira, N. Sugimoto, T. Yamaguchi, T. Nomura, H. Ito, T. Nakamura, et al. 2007. Preferential recruitment of CCR6-expressing Th17 cells to inflamed joints via CCL20 in rheumatoid arthritis and its animal model. J. Exp. Med. 204: 2803–2812.
- Sierro, F., B. Dubois, A. Coste, D. Kaiserlian, J. P. Kraehenbuhl, and J. C. Sirard. 2001. Flagellin stimulation of intestinal epithelial cells triggers CCL20-mediated migration of dendritic cells. *Proc. Natl. Acad. Sci. USA* 98: 13722–13727.
- Caron, G., D. Duluc, I. Frémaux, P. Jeannin, C. David, H. Gascan, and Y. Delneste. 2005. Direct stimulation of human T cells via TLR5 and TLR7/8: flagellin and R-848 up-regulate proliferation and IFN-γ production by memory CD4⁺ T cells. *J. Immunol.* 175: 1551–1557.
 Chalifour, A., P. Jeannin, J. F. Gauchat, A. Blaecke, M. Malissard, T. N'Guyen,
- Chalifour, A., P. Jeannin, J. F. Gauchat, A. Blaecke, M. Malissard, T. N'Guyen, N. Thieblemont, and Y. Delneste. 2004. Direct bacterial protein PAMP recognition by human NK cells involves TLRs and triggers α-defensin production. *Blood* 104: 1778–1783.
- Sanders, C. J., D. A. Moore, III, I. R. Williams, and A. T. Gewirtz. 2008. Both radioresistant and hemopoietic cells promote innate and adaptive immune responses to flagellin. *J. Immunol.* 180: 7184–7192.
- Janot, L., J. C. Sirard, T. Secher, N. Noulin, L. Fick, S. Akira, S. Uematsu, A. Didierlaurent, T. Hussell, B. Ryffel, and F. Erard. 2009. Radioresistant cells expressing TLR5 control the respiratory epithelium's innate immune responses to flagellin. *Eur. J. Immunol.* 39: 1587–1596.
- Caminschi, I., F. Ahmet, K. Heger, J. Brady, S. L. Nutt, D. Vremec, S. Pietersz, M. H. Lahoud, L. Schofield, D. S. Hansen, et al. 2007. Putative IKDCs are functionally and developmentally similar to natural killer cells, but not to dendritic cells. J. Exp. Med. 204: 2579–2590.
- 37. Schmutz, S., N. Bosco, S. Chappaz, O. Boyman, H. Acha-Orbea, R. Ceredig, A. G. Rolink, and D. Finke. 2009. Cutting edge: IL-7 regulates the peripheral pool of adult ROR γ^+ lymphoid tissue inducer cells. J. Immunol. 183: 2217–2221.

- Pickert, G., C. Neufert, M. Leppkes, Y. Zheng, N. Wittkopf, M. Warntjen, H. A. Lehr, S. Hirth, B. Weigmann, S. Wirtz, et al. 2009. STAT3 links IL-22 signaling in intestinal epithelial cells to mucosal wound healing. *J. Exp. Med.* 206: 1465–1472.
- Zenewicz, L. A., G. D. Yancopoulos, D. M. Valenzuela, A. J. Murphy, S. Stevens, and R. A. Flavell. 2008. Innate and adaptive interleukin-22 protects mice from inflammatory bowel disease. *Immunity* 29: 947–957.
 Kinnebrew, M. A., C. Ubeda, L. A. Zenewicz, N. Smith, R. A. Flavell, and
- Kinnebrew, M. A., C. Ubeda, L. A. Zenewicz, N. Smith, R. A. Flavell, and E. G. Pamer. 2010. Bacterial flagellin stimulates Toll-like receptor 5-dependent defense against vancomycin-resistant *Enterococcus* infection. *J. Infect. Dis.* 201: 534–543.
- Veldhoen, M., K. Hirota, J. Christensen, A. O'Garra, and B. Stockinger. 2009. Natural agonists for aryl hydrocarbon receptor in culture medium are essential for optimal differentiation of Th17 T cells. *J. Exp. Med.* 206: 43–49.
 Schraml, B. U., K. Hildner, W. Ise, W. L. Lee, W. A. Smith, B. Solomon,
- Schraml, B. U., K. Hildner, W. Ise, W. L. Lee, W. A. Smith, B. Solomon, G. Sahota, J. Sim, R. Mukasa, S. Cemerski, et al. 2009. The AP-1 transcription factor Batf controls T(H)17 differentiation. *Nature* 460: 405–409.
- Gascoyne, D. M., E. Long, H. Veiga-Fernandes, J. de Boer, O. Williams, B. Seddon, M. Coles, D. Kioussis, and H. J. Brady. 2009. The basic leucine zipper transcription factor E4BP4 is essential for natural killer cell development. *Nat. Immunol.* 10: 1118–1124.
- 44. Satoh-Takayama, N., S. Lesjean-Pottier, P. Vieira, S. Sawa, G. Eberl, C. A. Vosshenrich, and J. P. Di Santo. 2010. IL-7 and IL-15 independently program the differentiation of intestinal CD3⁻NKp46⁺ cell subsets from Id2dependent precursors. *J. Exp. Med.* 207: 273–280.

- Awasthi, A., L. Riol-Blanco, A. Jäger, T. Korn, C. Pot, G. Galileos, E. Bettelli, V. K. Kuchroo, and M. Oukka. 2009. Cutting edge: IL-23 receptor gfp reporter mice reveal distinct populations of IL-17–producing cells. *J. Immunol.* 182: 5904–5908.
- Cerovic, V., C. D. Jenkins, A. G. Barnes, S. W. Milling, G. G. MacPherson, and L. S. Klavinskis. 2009. Hyporesponsiveness of intestinal dendritic cells to TLR stimulation is limited to TLR4. J. Immunol. 182: 2405–2415.
- Burdelya, L. G., V. I. Krivokrysenko, T. C. Tallant, E. Strom, A. S. Gleiberman, D. Gupta, O. V. Kurnasov, F. L. Fort, A. L. Osterman, J. A. Didonato, et al. 2008. An agonist of Toll-like receptor 5 has radioprotective activity in mouse and primate models. *Science* 320: 226–230.
- Vijay-Kumar, M., J. D. Aitken, C. J. Sanders, A. Frias, V. M. Sloane, J. Xu, A. S. Neish, M. Rojas, and A. T. Gewirtz. 2008. Flagellin treatment protects against chemicals, bacteria, viruses, and radiation. *J. Immunol.* 180: 8280–8285.
 Meier, D., C. Bornmann, S. Chappaz, S. Schmutz, L. A. Otten, R. Ceredig,
- Meier, D., C. Bornmann, S. Chappaz, S. Schmutz, L. A. Otten, R. Ceredig, H. Acha-Orbea, and D. Finke. 2007. Ectopic lymphoid-organ development occurs through interleukin 7-mediated enhanced survival of lymphoid-tissueinducer cells. *Immunity* 26: 643–654.
- Bouskra, D., C. Brézillon, M. Bérard, C. Werts, R. Varona, I. G. Boneca, and G. Eberl. 2008. Lymphoid tissue genesis induced by commensals through NOD1 regulates intestinal homeostasis. *Nature* 456: 507–510.
- Tsuji, M., K. Suzuki, H. Kitamura, M. Maruya, K. Kinoshita, I. I. Ivanov, K. Itoh, D. R. Littman, and S. Fagarasan. 2008. Requirement for lymphoid tissue-inducer cells in isolated follicle formation and T cell-independent immunoglobulin A generation in the gut. *Immunity* 29: 261–271.

I.2. La protection contre les infections aigues de Streptococcus pneumoniae sérotype 1 présente une signature immunitaire implicant Th17 et IFN-γ.

Le second papier réalisé dans le cadre de cette étude a été publié dans le journal Immunobiology en novembre 2011 sous le titre « Protection against Streptococcus pneumoniae serotype 1 acute infection shows a signature of Th17- and IFN-γ-mediated immunity. » par Marqués JM, Rial A, Muñoz N, Pellay FX, Van Maele L, <u>Léger H</u>, Camou T, Sirard JC, Benecke A, Chabalgoity JA.

Marqués et al. évaluent, après une infection intranasale avec une dose subléthale de S. pneumoniae serotype1, la réponse transcriptionnelle, humorale et cellulaire d'animaux protégés et non-protégés contre S. pneumoniae sérotype 1. Une augmentation de l'expression d'IL-17A et une régulation spécifique de l'expression des gènes IFN- γ -related corrélée avec une réponse de Th17 semblent être des caractéristiques pertinentes de la réponse immunitaire contre *S. pneumoniae acute pneumonia.*

L'inoculation intranasale d'une dose sublétale de S. *pneumoniae* de sérotype 1 a conféré une protection complète contre une infection ultérieure avec une dose létale de la même souche. L'infection sublétale a suscité une forte réponse anticorps IgM et IgG contre le polysaccharide capsulaire, comme ont pu le constater une semaine plus tard les auteurs de cette étude. Elle a aussi induit un fort afflux de neutrophiles dans les poumons, immédiatement après l'injection de la dose létale. L'analyse transcriptomique à l'aide de puces à ADN pangénomiques des échantillons de poumons a montré l'existence de 149 gènes différentiellement exprimés chez des souris protégées versus des souris non protégées. Parmi ces gènes, on retrouve une régulation positive d'Il17a, IFNg et de plusieurs gènes liés à l'expression d'IL-17A et IFN. Une analyse cinétique a montré des niveaux d'expression élevés de Il17a chez les animaux protégés à tous les points dans le temps alors que IFNg a été augmentée au début de la cinétique chez les souris protégées et plus tardivement chez les animaux non-protégés.

Le marquage des cytokines intracellulaires a montré que les cellules T CD4 + produisent une grande proportion de l'IL-17A des poumons chez les animaux protégés.

Dans l'ensemble, ces résultats ont montré qu'une régulation à la hausse de l'IL-17A et une régulation au cours du temps de l'expression des gènes relatifs à l'IFN, ainsi que la mise en place d'une réponse Th17, sont des caractéristiques d'une immunité protectrice des infections aigues à *S. pneumonia acute pneumonia*.

ARTICLE IN PRESS

Immunobiology xxx (2011) xxx-xxx



Contents lists available at SciVerse ScienceDirect

Immunobiology



journal homepage: www.elsevier.de/imbio

Protection against *Streptococcus pneumoniae* serotype 1 acute infection shows a signature of Th17- and IFN- γ -mediated immunity

Juan M. Marqués^{a,1}, Analía Rial^{a,1}, Natalia Muñoz^a, Francois-Xavier Pellay^b, Laurye Van Maele^{c,d,e}, Hélène Léger^b, Teresa Camou^f, Jean-Claude Sirard^{c,d,e}, Arndt Benecke^b, José A. Chabalgoity^{a,*}

^a Laboratory for Vaccine Research, Departamento de Desarrollo Biotecnológico, Instituto de Higiene, Facultad de Medicina – UdelaR. Avda. A. Navarro 3051,

CP 11600, Montevideo-Uruguay

^b Institut des Hautes Études Scientifiques, Bures-sur-Yvette, France

^c Institut Pasteur de Lille, Centre d'Infection et d'Immunité de Lille, Institut National de la Santé et de la Recherche Médicale, U1019, Lille, France

^d Centre National de la Recherche Scientifique, UMR 8204, Lille, France

^e Univ Lille Nord de France, Lille, France

^f Departamento de Laboratorios del Ministerio de Salud Pública, Montevideo-Uruguay

ARTICLE INFO

Article history: Received 11 July 2011 Received in revised form 18 October 2011 Accepted 21 October 2011

Keywords: IL-17A IFN-γ Pneumonia Protective immunity Streptococcus pneumoniae Th17 Transcriptome

ABSTRACT

Acute pneumonia caused by Streptococcus pneumoniae is a major cause of child mortality. Antibodies are considered the main effectors of protection in this clinical presentation of pneumococcal invasive disease. To get new insights into the mechanisms involved in the protective immunity, we established a murine experimental model of protection against acute pneumococcal pneumonia and then evaluated the transcriptional, humoral and cellular responses in protected and non-protected animals. We found that intranasal inoculation of a sublethal dose of S. pneumoniae serotype 1 conferred complete protection against a subsequent challenge with a lethal dose of the same strain. Sublethal infection elicited a strong IgM and IgG antibody response against the capsular polysaccharide, as assessed one week later, and an exacerbated influx of neutrophils into the lungs immediately after the lethal challenge. Genomewide microarray-based transcriptional analysis of whole lungs showed 149 differentially expressed genes among which we found upregulation of *ll17a*, *lfng* and several IL-17A- and IFN- γ -related genes in protected versus non-protected mice. Kinetics analysis showed higher expression levels of Il17a in protected animals at all time points whereas Ifng was upregulated early in the protected mice and later in the non-protected animals. Intracelluar cytokine staining demonstrated that CD4+ T cells account for a great proportion of the IL-17A produced in the lungs of protected animals. Overall, these results showed that an upregulation of IL-17A- and a timely regulation of IFN-γ-related gene expression, together with development of a Th17 response, are relevant characteristics of the protective immunity against S. pneumoniae acute pneumonia.

© 2011 Elsevier GmbH. All rights reserved.

Introduction

Streptococcus pneumoniae (pneumococcus) is the main etiological agent of human pneumonia and meningitis, otitis media and sinusitis; per year, about 1–2 millions of fatal pneumococcal diseases occur worldwide (van der Poll and Opal 2009). Pneumococcus colonizes asymptomatically the human nasopharynx and in some cases colonization may progress to pneumonia and invasive disease (Kayhty et al. 2006; Salyers 2002). Based on differences in the capsular polysaccharide composition 91 pneumococcus serotypes

* Corresponding author. Tel.: +598 2 487 12 88x1120; fax: +598 2 487 30 73. *E-mail address:* jachabal@higiene.edu.uy (J.A. Chabalgoity).

¹ These authors contributed equally to this work.

have been described (van der Poll and Opal 2009). Some serotypes like serotype 1 are rarely associated with carriage and most frequently cause invasive infections (Hortal et al. 2008; Normark and Normark 2002).

Innate immunity plays a critical role in the resistance to *S. pneu-moniae* respiratory infection, both in human and mouse (Paterson and Mitchell 2006). Innate protection is characterized by an early influx of inflammatory cells, predominantly neutrophils (Kadioglu et al. 2000; Munoz et al. 2010; Rosseau et al. 2007), although lungs' infiltrating CD4⁺ T cells are also involved in the early resistance to pneumococcal infection (Kadioglu et al. 2000, 2004). Another key component of innate immunity is the deposition and activation of complement C3 component on the surface of pneumococcus and it has been shown to be impaired by capsular polysaccharides (Abeyta et al. 2003). Otherwise, activation of the classical

Please cite this article in press as: Marqués, J.M., et al., Protection against *Streptococcus pneumoniae* serotype 1 acute infection shows a signature of Th17- and IFN- γ -mediated immunity. Immunobiology (2011), doi:10.1016/j.imbio.2011.10.012

^{0171-2985/\$ -} see front matter © 2011 Elsevier GmbH. All rights reserved. doi:10.1016/j.imbio.2011.10.012

2

ARTICLE IN PRESS

J.M. Marqués et al. / Immunobiology xxx (2011) xxx-xxx

complement pathway by natural IgM antibodies has been demonstrated to be important in preventing a rapid bacteremia during the infection with pneumococcus (Brown et al. 2002).

Activation of innate immunity is not only critical as the first line of defense against pneumococcus but also for the generation of appropriate adaptive immunity (Paterson and Mitchell 2006). It has been proposed that acquired immunity against pneumococcal nasopharyngeal colonization is mediated by CD4⁺ T-cells and independent of antibodies (Basset et al. 2007; Malley et al. 2005; Trzcinski et al. 2005, 2008). Recently, it has been also demonstrated that IL-17A, produced by Th17 cells, is a key factor in protective immunity against pneumococcal colonization (Lu et al. 2008), being essential for the recruitment of macrophages and neutrophils to the nasopharynx (Zhang et al. 2009). However, other groups have shown evidences that antibodies and particularly IgA are necessary in protection against pneumococcal colonization (Fukuyama et al. 2010; Richards et al. 2010; Roche et al. 2007; Roche and Weiser 2010). Overall, these data suggest that both antibodies and cellular responses may have an important role in protection against pneumococcal nasopharyngeal colonization (Richards et al. 2010).

Regarding pneumonia and invasive pneumococcal disease there is a consensus that anti-capsular antibodies play an essential role in protection (AlonsoDeVelasco et al. 1995; Lynch and Zhanel 2009; Malley 2010; WHO 2008). In this sense, available pneumococcal vaccines are designed to elicit protective anti-capsular polysaccharide antibodies (Poolman 2004). The binding of specific antibodies to the bacterial surface facilitates its Fc- and complement-mediated opsonophagocytosis by neutrophils and macrophages (AlonsoDeVelasco et al. 1995).

To gain further insights into the mechanisms involved in protection against invasive pneumococcal disease, we established a murine experimental model of protection against acute pneumococcal pneumonia and then evaluated the transcriptional, humoral and cellular responses in protected and non-protected animals. We confirmed the elicitation of a specific humoral response together with an important influx of neutrophils into the lungs of protected mice. We also found upregulation of IL-17A- and IFN- γ -related genes and we confirmed the presence of Th17 cells in protected animals suggesting that these cells and IFN- γ mediated immunity are related to protection against acute pneumonia.

Materials and methods

Ethics statement

All experiments involving animals complied with current national and institutional regulations and ethical guidelines from *"Comisión Honoraria de Experimentación Animal"* (CHEA) – Universidad de la República, Uruguay (approval ID: 08-05-10).

Animal studies

Female C57BL/6J mice, 6–8 weeks old, were obtained from the National Division of Veterinary Laboratories. Animals were maintained in individually ventilated cages and handled in a vertical laminar flow cabinet (ESCO, class II A2). Mice were anaesthetized by intraperitoneal (i.p.) injection of 2.2 mg Ketamine plus 0.11 mg Xylazine. Bacteria were administrated onto mice's nostrils in 50 μ l of physiological saline. Mice survival was recorded daily. For lung sampling, mice were sacrificed by cervical dislocation at different times post-infection. Lungs were collected, homogenized in cell dissociation sieve-tissue grinder kit (Sigma–Aldrich), and serial dilutions were plated onto blood-agar plates to determine the degree of bacterial infection. *S. pneumoniae* colonies were detected by the characteristic green halo caused by its α -hemolytic activity.

Lungs for transcriptional studies were stored in RNAlater[®] (Qiagen).

Bacterial cultures

S. pneumoniae serotype 1, clinical isolate E1586 sequence type ST304 (Zemlickova et al. 2005), was provided by the National Reference Laboratory (Ministry of Health, Uruguay). Working stocks were prepared as previously described (Munoz et al. 2010). Briefly, Todd Hewitt Yeast Broth (THYB) was inoculated with fresh colonies of *S. pneumoniae* grown in blood-agar plates and incubated at 37 °C to reach $OD_{600 \text{ nm}}$ 0.7–0.9. Cultures were stored at $-80 \degree C$ in THYB + glycerol 12% (v/v) up to 3 months. For infections, frozen stocks were thawed, centrifuged 5 min at $2500 \times g$, washed and diluted to the appropriate concentration with sterile physiological saline solution. Sublethal and lethal doses were defined as 4×10^4 CFU and 2×10^7 CFU per mouse, respectively. Number of bacteria in stocks and inoculums was determined by plating serial dilutions onto blood agar plates.

Detection of IgG and IgM antibodies by ELISA

For detection of IgG, IgM and IgA specific for a whole cell pneumococcal type 1 antigen (WCPn1Ag) 96-well microtiter plates (Maxisorp, Nunc) were coated with 1 µg/well of WCPn1Ag and incubated overnight (o/n) at 4 °C. WCPn1Ag was prepared by sonication of a bacterial culture of S. pneumoniae serotype 1 growth to mid-log phase. Briefly, bacteria were grown to an OD_{630 nm} of 0.8-1.0 and then harvested by centrifugation, resuspended in saline and frozen before sonication. After sonication, it was verified by CFU counting that there were not viable bacteria left. The sonicate was then centrifuged, and total protein content was quantified in the supernatant by the method of Pierce. After washing the plates with PBS containing 0.05% Tween-20 (PBS-T), they were blocked with 200 µl/well of 1% bovine seroalbumin in PBS-T (PBS-T-BSA) for 1 h at 37 °C. After washing, 100 µl/well of an appropriate dilution of each serum sample (in PBS-T -0.1%BSA) were added and incubated for 2 h at 37 °C. Plates were washed and 100 μl of diluted peroxidase-conjugated goat anti-mouse IgG or IgA (Southern Biotech, Birmingham, AL) or IgM (SIGMA) were added and plates were incubated o/n at 4 °C. Finally the ELISA was developed by adding 100 µl/well of OPD-H₂O₂ (o-phenylenediamine, SIGMA) and readings were taken at 490 using a microplate reader (MRX II, Dynex Technologies). A standard curve prepared with a positive reference sample (arbitrarily assigned with a value of 1000 Arbitrary Units (AU)/ml) was run in each plate and test samples were interpolated.

For the detection of specific IgG, IgM and IgA antibodies against polysaccharide type 1 (PnPS1) we used a modification of the WHO 3rd generation ELISA used for quantification of anti-polysaccharide IgG antibodies in human serum, using both C-polysaccharide (C-PS) and 22F polysaccharide absorption (Wernette et al. 2003). Briefly, 96-well medium binding microtiter plates (Greiner) were coated with 0.2 µg/well of type 1 specific pneumococcal polysaccharide (American Type Culture Collection [ATCC], Manassas, Va.) for 5 h at 37 °C, in a humidified chamber. Plates were washed with Trisbuffered saline with 0.01% Brij-35 solution (Wash buffer). Test sera were pre-absorbed with $5 \mu g/ml$ cell wall polysaccharide (C-PS, Staten Serum Institute, Copenhagen, Denmark) and $5 \mu g/ml 22F$ capsular polysaccharide (ATCC), and incubated for 30 min at room temperature (RT). Then, 50 µl of each sample were transferred to the coated microtiter plate, and incubated o/n at RT. After washing, 100 µl of diluted peroxidase-conjugated goat anti-mouse IgG or IgA (Southern Biotech, Birmingham, AL) or IgM (SIGMA) were added and plates were incubated o/n at 4°C. Finally the ELISA was

Please cite this article in press as: Marqués, J.M., et al., Protection against *Streptococcus pneumoniae* serotype 1 acute infection shows a signature of Th17- and IFN-γ-mediated immunity. Immunobiology (2011), doi:10.1016/j.imbio.2011.10.012

ARTICLE IN PRESS

J.M. Marqués et al. / Immunobiology xxx (2011) xxx-xxx

developed by adding $100 \,\mu$ l/well of TMB (3,3',5,5'-tetramethylbenzidine, SIGMA) and absorbances at 450 (with reference at 630 nm) were read using a microplate reader (MRX II, Dynex Technologies). A standard curve prepared with a positive reference sample (arbitrarily assigned with a value of 1000 AU/ml) was run in each plate where test samples were interpolated.

Quantitative RT-PCR

RNAlater solution was removed, and lungs were homogenized with an Ultra-Turrax in lysis buffer to then extract total RNA using RNeasy Mini Kit (Qiagen). RNA quality and quantity was assessed by spectrophotometric measurements at 260/280 nm. Prior to cDNA synthesis, 1 µg of total RNA was treated with DNAse-I (Invitrogen) according to manufacturer's instructions. Retrotranscription was done in a final volume of $20 \,\mu l$ in the presence of Random primers (200 ng), dNTPs (0.5 mM), DTT (0.01 M), RNaseOUT (40 U) and SuperScript III Reverse Transcriptase (200 U), Invitrogen. Real time RT-qPCR was conducted by high throughput methods using TaqMan Low Density Array (TLDA) cards (Applied Biosystems) and the 7900HT System (Applied Biosystems). Real-time PCRs for Il17a, Il17f, Il22, Ifng and Cxcl11 mRNA levels, were also performed using QuantiTect[®] SYBR[®] Green PCR Kit (Qiagen) in a Rotor-Gene 6000 (Corbett) or 7900HT System (Applied Biosystems) (primer sequences are available under request). Beta-Actin encoding gene Actb was utilized as house-keeping gene. The relative mRNA amount in each sample was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001) were $\Delta Ct = Ct_{gene of interest} - Ct_{Actb}$, and expressed as relative mRNA levels in the test group compared to the control group.

Microarrays assays

Total RNA was extracted using NucleoSpin[®] RNA L (Macherey-Nagel) and was concentrated in YM30 Microcon centrifugal filter device (Millipore) by centrifugation at 14,000 × g during 60 min at 4 °C. RNA quality was evaluated with an Agilent 2100 Bioanalyser (Agilent Technologies). Poly(A) mRNA from 20 μ g of total RNA was converted into digoxigenin labeled cDNA for hybridization to Mouse Genome Survey Microarrays (Applied Biosystems), using Chemiluminescent RT Labeling Kit for detection in an Applied Biosystems ABI1700 system. The microarray data described in this study is MIAME compliant and they have been deposited in the publicly available database MACE (http://mace.ihes.fr) under accession number: 3034545238.

Data analysis

Microarray data were quality-controlled (Brysbaert et al. 2010). Statistical analysis of the microarray data was carried out as described previously (Noth and Benecke 2005; Noth et al. 2006; Van Maele et al. 2010).

Ontology enrichment analysis

Combining GO, Kegg and PANTHER annotations, we assigned all probes present on the MGS Version 2.0 array to the PANTHER biological processes. We then calculated the relative representation of those probes detected as significantly regulated. *p*-Values for overand under-representation of biological processes were calculated using a binominal distribution.

Quantification of IL-17A and IFN- γ by ELISA

 γ -interferon (IFN- γ) cytokine concentrations in lung homogenates was assayed by enzyme-linked immunoabsorbent

assay (ELISA) using a pair of specific monoclonal antibodies (capture and detection) against each cytokine and dilutions of the recombinant cytokine for the construction of a standard curve (all reagents were supplied by BD Pharmingen) as previously described (Rial et al. 2004). IL-17A concentration in lung homogenates was quantified using the Ready-Set-Go Mouse IL-17A ELISA Reagent Set from eBioscience (San Diego, CA, USA), according to manufacturer's instructions.

Flow cytometry analysis of lungs cell suspensions

At specified time points after challenge mice were sacrificed by cervical dislocation and the pulmonary and systemic circulation was perfused with saline +1 mM EDTA to remove the intravascular pool of cells. Lung cells were isolated after collagenase/DNAse treatment as previously described (Rial et al. 2004), filtered through a 40 μ m cell strainer (BD) and washed and stained for FACS analvsis. Cells were resuspended in FACS EDTA buffer (PBS, 0.1% azide, 1% bovine serum albumin, 5 mM EDTA) and enumerated using a Countess Automated Cell Counter (Invitrogen, USA) previous to the immunophenotypic analysis by flow cytometry. For neutrophils analysis cells were labeled with Ly6G-PE, CD11b-APC (BD Pharmingen, USA) for 30 min on ice, washed, fixed with PFA 4%, and finally resuspended in FACS EDTA. For intracellular cytokine staining, lung cells were stimulated with 50 ng/ml phorbol myristate acetate (PMA, SIGMA) and 500 ng/ml ionomycin (SIGMA) in the presence of 10 µg/ml Brefeldin A (SIGMA), for 5 h at 37 °C, 5% CO₂. Cells were surface stained, for 30 min on ice, with specific antibodies against CD45, CD3 and CD4 (BD), washed with FACS EDTA and then with PBS. After fixing with PFA 4%, cells were permeabilized with 0.5% saponin (SIGMA) in PBS-0.1% BSA and intracellularly stained with antibodies against specific cytokines (PE- IL-17A and Alexa488- or PE- IFN-gamma, all from BD) for 30 min, at RT, and then washed. Cells prepared for flow cytometry were analyzed after gating for CD45⁺ viable cells using a FACSCalibur (BD, San Jose, CA) equipped with argon and red diode lasers used for excitation at 488 and 635 nm. For data acquisition and analysis CellQuest 3.3 software package (BD) was used. Experiments were performed a minimum of three times in an independent manner.

Statistics

Statistical analyses were assessed by non-parametric Mann–Whitney *U* test or Student *T* test and differences were regarded as significant with p < 0.05 (*) and very significant with p < 0.01 (**).

Results

A sublethal infection protects against acute S. pneumoniae serotype 1 pneumonia

We first set up an experimental model of acute pneumonia with a clinical isolate of *S. pneumoniae* serotype 1 (Pn1) based on intranasal (i.n.) administration of 2×10^7 CFU of Pn1 to C57BL/6J mice. The infection was lethal for animals within 4 days and was invasive since bacteria were found at 24 h in the blood. We then established a model of protection against the lethal challenge by prior intranasal administration of a sublethal dose of the same pneumococcus strain. Mice treated with 4×10^4 CFU of Pn1 cleared the infection from the lungs by 24h and all survived a subsequent lethal challenge with 2×10^7 CFU of Pn1 administered either one week (Fig. 1A) or 2 months (results not shown) after treatment, while all saline-treated mice died after the lethal challenge. Both protected and non-protected animals showed similar bacterial loads in the lungs at 4h and 24h post-challenge (Fig. 1B).

Please cite this article in press as: Marqués, J.M., et al., Protection against *Streptococcus pneumoniae* serotype 1 acute infection shows a signature of Th17- and IFN- γ -mediated immunity. Immunobiology (2011), doi:10.1016/j.imbio.2011.10.012

4

ARTICLE IN PRESS

J.M. Marqués et al. / Immunobiology xxx (2011) xxx-xxx



Fig. 1. A sublethal dose of *S. pneumoniae* protects against a subsequent challenge with a lethal dose. On day 0, C57BL/6 mice were treated i.n. with a sublethal dose of *S. pneumoniae* serotype 1 (Pn1, 4×10^4 CFU) or saline. On day 7, mice were challenged i.n. with a Pn1 lethal dose (2×10^7 CFU). (A) Survival rate of animals (n = 6 per group). (B) Box–Whiskers plot of lung bacterial load 4, 24 and 48 h after the challenge of protected and saline-treated control mice. Lungs were harvested at the indicated time point and homogenized to count bacteria in blood agar plate (n = 4 per group at different time point). Results are representative of more than 3 independent experiments.

However, by 48 h protected animals have almost cleared the infection from the lungs while bacterial load was significantly increased in the non-protected group (Fig. 1B). Bacteria were found in the blood of all saline-treated animals 24 h after lethal infection while blood cultures were negative for all protected animals (data not shown). In conclusion, our data showed that pre-treatment with a sublethal dose of Pn1 confers solid and long-lasting protection to invasive pneumococcal infection with the same strain.

Sublethal Pn1 infection elicited early IgM and IgG specific antibodies

Antibodies have been long recognized as the main effectors of protection against pneumococcal invasive diseases. Therefore, we evaluated the humoral response in our model of protection. We analyzed IgG and IgM antibodies specific for whole cell pneumococcal serotype 1 antigen (WCPn1Ag) and capsular pneumococcal polysaccharide type 1 (PnPS1) in serum obtained 7 days after treatments, just before challenge. Mice that received a sublethal dose of Pn1 showed increased levels of specific anti-WCPn1Ag IgM antibodies, while no significant levels of specific IgG (Fig. 2A) nor IgA (data not shown) antibodies were detected. Interestingly, these animals developed significant levels of both IgM and IgG antibodies specific to type 1 capsular polysaccharide (Fig. 2B). Thus, a sublethal i.n. Pn1 infection induced significant IgM and IgG anti-capsular antibodies as early as one week after administration that could be associated with protection against pneumonia.

Transcriptional analysis showed upregulation of Il17a and Ifng in protected mice



To further investigate the immune mechanisms involved in protection, we evaluated whether protection correlated to a specific transcriptional signature in the lungs. For this, we conducted

Fig. 2. Treatment with a sublethal intranasal dose of *S. pneumoniae* induces the production of serum specific antibodies. Mice receiving 4×10^4 CFU were bled 7 days later and specific IgG and IgM antibodies against whole cell pneumococcal type 1 antigen (WcPn1Ag) (A) or purified pneumococcal polysaccharide type 1 (PnPS1) (B) were detected by ELISA. Results are expressed as mean \pm SEM of 8 mice/group. For statistical analysis Mann–Whitney Test was used.

Please cite this article in press as: Marqués, J.M., et al., Protection against *Streptococcus pneumoniae* serotype 1 acute infection shows a signature of Th17- and IFN-γ-mediated immunity. Immunobiology (2011), doi:10.1016/j.imbio.2011.10.012
PROTECTED NON PROTECTED
Pn1 Sublethal
+ Pn1 Lethal
2764
Saline
2724

Fig. 3. Protected and non-protected mice show similar lung transcriptional profile. Diagram showing microarrays analysis and the number of statistically significant (p < 0.01) differentially expressed genes in the comparison between 2 groups (n = 3 per group).

microarray-based genome-wide expression analysis. First, we evaluated the global lung transcriptional response associated with acute pneumonia at 24h. We found that infected mice presented 2724 differentially expressed genes as compared with saline-treated animals: 1174 transcripts were downregulated and 1550 were upregulated (Fig. 3 and Supplementary Table 1). Most altered biological processes were those related to immunity and defense, like interferon-mediated immunity and B-cell and antibody-mediated immunity, as well as those related to signal transduction, cell adhesion, cell communication and cell proliferation (Supplementary Table 2). Our results closely correlated with previously reported observations (Supplementary Fig. 1) (Rosseau et al. 2007). In conclusion, we observed that acute pneumonia induces transcriptional alterations in the lungs that correlate in part with the induction of a strong immune response that is unable to control the infection.

We next compared the lung transcriptional profile of protected and non-protected mice sacrificed 24 h after the lethal challenge. Both protected and non-protected groups presented similar pattern and number of statistically altered genes when compared to saline-treated animals (2764 and 2724 respectively, Fig. 3). Of note, comparison between protected versus non-protected mice showed only 149 genes that were statistically significant (p < 0.01) differentially expressed: 133 genes were up-regulated while only 16 were down-regulated (Fig. 3 and Supplementary Table 3). Within differentially expressed genes there was a clear dominance of those related to immunity (Table 1) and the principal altered biological processes were all related to immunity such as macrophage-, interferon-, B cell- and antibody-, T cell-, cytokine- and chemokine-, granulocyte-mediated immunity processes (Table 2). These data were validated by real time RT-qPCR for a set of relevant genes related to immunity (Supplementary Fig. 2).

Among differentially expressed genes related to immunity, we identified *Ifng* and *Il17a* as relevant genes associated to protection against pneumococcal infection (Table 1). Additionally to *Ifng* there were other altered genes related to IFN- γ - like *Cxcl9*, *Cxcl11*, *Tap1*, *Gbp2*, *Irf1*, *Irf8* and *Socs1*, reinforcing the relevance of IFN- γ -mediated immunity. Similarly, upregulation of genes associated with T-cell mediated immunity like *C2ta* (class II transactivator), *Gzma* (granzyme A), *Ctla4* (cytotoxic T-lymphocyte-associated protein), *Icos* (inducible T-cell co-stimulator) were also a hallmark of protection (Table 1). We found upregulation of several genes related to B-cell- and antibody-mediated immunity like *Igj* (immunoglobulin joining chain), *Igk* (immunoglobulin kappa chain) and *Cd40* (CD40 antigen) that could be associated with the induced humoral response against pneumococcus.

upregulated genes included *Gzmb*, a gene related to granulocyte mediated immunity, and the chemokine ligand 1 *Xcl1* related to ckemokine mediated immunity (Table 1). Overall the pattern of altered genes and altered biological processes suggested an important role of immune response involving granulocytes, B- and T-cells in protection.

IL-17 and IFN- γ responses are accelerated and strengthened in protected animals

We then decided to focus our study on genes related to IL-17A and IFN- γ and we evaluated their kinetic of expression by RT-qPCR in protected and non-protected mice. For IL-17A related genes we extended the analysis to genes like *ll17f* and *ll22* that only emerged as differentially expressed genes in the microarray analysis when considering *p* < 0.05 instead of *p* < 0.01. For IFN- γ -related genes we analyzed *lfng* and *Cxcl11*, and also *Socs1* which is involved in the negative control of IFN- γ signaling pathway.

Early as 8 h after challenge mRNA levels of *ll17a* and *ll17f* were already significantly higher in protected animals (Fig. 4A). The difference between groups peaked by 24 h and was maintained at all time points. By 24 h we found a 7–8 fold increase in *ll17a* mRNA level that correlated with a similar fold increase in protein levels (Fig. 4A and C). Transcript levels of *ll22* were also significantly upregulated at 24 h although they decreased at 48 h reaching similar levels as those in the non-protected group (Fig. 4A). These results showed that upregulation of IL-17-related genes expression is a hallmark of protection.

Regarding IFN- γ -related genes, we found that mRNA levels of *lfng* and *Cxcl11* did also peaked at 24h in protected animals (Fig. 4B); we also confirmed augmented levels of IFN- γ protein at this time point (Fig. 4C). However, and contrary to what it was observed for IL-17-related genes, *lfng* and *Cxcl11* were strongly upregulated by 48h in the non-protected group, and transcript levels were similar or even higher than in protected animals (Fig. 4B). Several other IFN- γ -related genes like *Cxcl9*, *Tap1*, *lrf1*, *lrf8* and *Gbp2* have similar behavior (data not shown). Thus, protected animals displayed an early upregulation of *lfng* and other IFN- γ -related genes that seem to be regulated as suggested by the concurrent upregulation of *Socs1*, whereas in animals that were going to die the IFN- γ mediated inflammatory response was delayed.

CD4⁺ T cells are an important source of IL-17A but not of IFN- γ

The upregulation of IL17A and cytokines related to Th17-cells like IL-17F and IL-22 in protected mice suggested the presence of Th17-cells in the lungs of these mice. Using intracellular cytokine detection we found a significant increase in the percentage of IL17A expressing cells in protected mice compared to non-protected mice sacrificed 24h after Pn1 lethal challenge (Fig. 5A). Therefore, the increase in IL-17A mRNA and protein levels correlated with the increase in IL17A-expressing cells. Regarding IFN- γ -expressing cells no statistically significant difference was observed between protected and non-protected mice but both groups showed increased numbers compared to naïve animals (Fig. 5A). However IFN- γ positive cells from protected mice showed higher IFN- γ mean fluorescence intensity than those from nonprotected mice (data not shown). Almost 40% of IL17A-producing cells in the lungs of protected mice were identified as CD4⁺ T cells (Fig. 5B). Nevertheless, T CD4⁺ cells represented a minor IFN- γ producer population in protected mice as well as in the other groups (Fig. 5B). Overall, these data indicated that protection against Pn1 acute pneumonia is associated with the presence of Th17 cells but not Th1 cells in the lungs.

J.M. Marqués et al. / Immunobiology xxx (2011) xxx-xx

Table 1

6

Genes related to immune response in lung microarray comparison of protected *versus* non-protected animals. On day 0, C57BL/6 mice were treated i.n. with a sublethal dose of *S. pneumoniae* serotype 1 (Pn1, 4×10^4 CFU) or saline. On day 7, mice were challenged i.n. with a Pn1 lethal dose (2×10^7 CFU). The table shows statistically significant (p < 0.01) differentially expressed genes related to immunity 24 h after the administration of a lethal dose in the lungs of protected (sublethal + lethal Pn1) *versus* non-protected mice (saline + lethal Pn1), 3 animals per group. LogQ is the logarithmic base 2 fold change. PANTHER annotation is shown.

Gene	Gene_Name	Celera ID	Entrez ID	LOG_Q
Cxcl9	chemokine (C-X-C motif) ligand 9	mCG12528.2	17329	3.04
Ifng	interferon gamma	mCG1237.1	15978	2.97
Igk-V21	immunoglobulin kappa chain variable 21 (V21)	mCG142161.1	16098	2.93
Gzmb	granzyme B	mCG130827	14939	2.55
U29423	cDNA sequence U29423	mCG1050270	243461	2.47
Igj	immunoglobulin joining chain	mCG21585.1	16069	2.13
Alox12e	arachidonate lipoxygenase, epidermal	mCG21175.1	11685	1.99
Il17a	interleukin 17a	mCG7914.1	16171	1.97
C2ta	class II transactivator	mCG126563.1	12265	1.97
Igk	immunoglobulin kappa chain complex	mCG142162.1	626145 545854	1.87
Gbp6	guanylate binding protein family, member 6	mCG3631.2	634650	1.85
Mgl1	macrophage galactose specific lectin 1	mCG21504.2	17312	1.83
Igk-v8-21	immunoglobulin kappa variable 8–21	null	620400	1.81
Gbp2	guanylate nucleotide binding protein 2	mCG122229	14469	1.79
Irf8	interferon regulatory factor 8	mCG21199.1	15900	1.69
Gbp6	guanylate binding protein 6	mCG142489 mCG142653	236573	1.66
Gbp8	guanylate-binding protein 8	mCG20328.2	76074	1.65
Clec4n	C-type lectin domain family 4, member n	mCG130574.2	56620	1.63
Gbp1	guanylate nucleotide binding protein 1	mCG122231	14468	1.60
Xcl1	chemokine (C motif) ligand 1	mCG8362.2	16963	1.51
Icos	inducible T-cell co-stimulator	mCG113578.1	54167	1.44
Lair1	leukocyte-associated Ig-like receptor 1	mCG59003.2	52855	1.43
Irf1	interferon regulatory factor 1	mCG13768.1	16362	1.41
Il18bp	interleukin 18 binding protein	mCG6766.2	16068	1.39
Ccr5	chemokine (C–C motif) receptor 5	mCG16378.1	12774	1.39
Gbp4	guanylate nucleotide binding protein 4	mCG21119.2	55932	1.37
Gzma	granzyme A	mCG4516.2	14938	1.34
Aif1	allograft inflammatory factor 1	mCG15923.1	11629	1.31
Lect2	leukocyte cell-derived chemotaxin 2	mCG5202.1	16841	1.30
Sell	selectin, lymphocyte	mCG9377.2	20343	1.26
Ifi47	interferon gamma inducible protein 47	mCG14415.3	15953	1.25
Cd209c	CD209c antigen	mCG1026173.1 mCG1028293.1	628890 170776	1.24
Cd244	CD244 natural killer cell receptor 2B4	mCG10012.1	18106	1.11
Tap1	transporter 1, ATP-binding cassette	mCG23012.2	21354	1.10
Igk-V8	immunoglobulin kappa chain-V8	mCG131871.2	384422 545854 16071	1.10
Prf1	perforin 1 (pore forming protein)	mCG15751.1	18646	1.10
Tap2	transporter 2, ATP-binding cassette	mCG23013.2	21355	1.06
Ccl7	chemokine (C-C motif) ligand 7	mCG52384.1	20306	1.06
Clec7a	C-type lectin domain family 7	mCG130109.2	56644	1.06
Cd40	CD40 antigen	mCG17530.2	21939	1.05
Cxcl13	chemokine (C-X-C motif) ligand 13	mCG8493.1	55985	-1.09
Plat	plasminogen activator, tissue	mCG13091.1	18791	-1.18

Protected animals showed higher recruitment of PMNs

Flow cytometry analysis of lung single cell suspensions showed that lethal Pn1 challenge induced a high recruitment of neutrophils (PMNs) both in sublethal Pn1 and saline treated mice compared to naïve mice, when evaluated 24 h after challenge (Fig. 6). More

Table 2

Microarrays analysis shows altered biological processes related to immune response. Ontology enrichment analysis for biological processes based on the 149 differentially expressed genes (3 between protected and non-protected animals 24 h after the lethal challenge.

Biological process	p-Value
Immunity and defense	9.31E – 10
Macrophage-mediated immunity	8.65E - 07
Interferon-mediated immunity	1.40E - 06
B-cell- and antibody-mediated immunity	1.77E - 04
Cytokine and chemokine mediated signaling pathway	2.19E - 04
T-cell mediated immunity	1.02E - 02
Granulocyte-mediated immunity	1.57E - 02
Cytokine/chemokine mediated immunity	1.79E - 02
Proteolysis	2.23E - 02
Signal transduction	3.47E - 02
Nucleoside, nucleotide and nucleic acid transport	3.47E - 02

importantly, protected mice showed higher amounts of PMNs than non-protected mice. PMNs have been considered as one of the principal effectors cells in the clearance of pneumococcus, most likely working in conjunction with specific anti-PS antibodies. In this sense, both generated antibodies and recruited PMNs seem to be critical for the observed protection.

Discussion

It is largely assumed that systemic antibodies to the capsular polysaccharide primarily determine the immunity against pneumococcal invasive disease (Malley 2010). To gain further insights into the molecular and cellular basis of this protection we established and characterized a murine model of protection against acute pneumonia using a clinical isolate of *S. pneumoniae* serotype 1. Then, we evaluated several aspects of transcriptional, humoral and cellular responses in protected and non-protected animals.

We showed that a sublethal Pn1 infection elicited specific antibodies against both pneumococcal polysaccharide type 1 (PnPS1) and whole cell antigen (WcPn1Ag), and resulted in 100% survival after a subsequent homologous lethal challenge. Anti-whole cell antigen antibodies were exclusively of IgM isotype, whereas the elicited anti-capsular antibodies were of both IgM and IgG



Fig. 4. In protected animals there is a strongly IL-17A- and an early IFN- γ -related genes response. On day 0, C57BL/6 mice were treated i.n. with a sublethal dose of *S. pneumoniae* serotype 1 (Pn1, 4 × 10⁴ CFU) or saline. On day 7, mice were challenged i.n. with a Pn1 lethal dose (2 × 10⁷ CFU). Lungs were harvested at 1.5, 4, 8, 24 and 48 h after the lethal dose (*n* = 7 at 24 h and *n* = 3 for other time points). Kinetics of relative mRNA levels referred to saline control group is shown for the Th17 related cytokines *ll17a*, *ll17f* and *ll22* (A), *lfng*, *Cxcl11* and *Socs1* for IFN- γ related genes (B). Quantity of IL-17A and IFN- γ protein per lung 24 h after challenge (C). Statistics are shown for Pn1sublethal + Pn1lethal compared to saline + Pn1 lethal groups in kinetic point utilizing Mann–Whitney Test when possible or Student's *T* test.

isotypes. A previous work reported that early anti-capsular antibodies raised during a *S. pneumoniae* infection were mainly of IgM isotype (Richards et al. 2010), whereas in our work we found high levels of both IgM and IgG anti-PnPS1. Unlike other pneumococcal serotypes, capsular polysaccharide type 1 has the particularity of presenting a zwitterionic motif within each repeating structural unit (Cobb and Kasper 2005). It has been shown that zwitterionic polysaccharides, and particularly PnPS1, can activate CD4⁺ T cells both *in vivo* and *in vitro* trough a MHCII-dependent mechanism (Velez et al. 2009), conferring T-dependent properties to this antigen (Cobb and Kasper 2005; Chung et al. 2003; Groneck et al. 2009). Therefore, we suggest that the early presence of high levels of anti-PnPS1 IgG observed in our model may be due to the zwitterionic characteristic of type 1 pneumococcal polysaccharide.

The results obtained from the microarray assays showed that i.n. challenge with a high dose of pneumococcus induced a strong immune response with 2274 differentially expressed genes that is nevertheless unable to control the infection. Instead, comparative analysis of lungs from protected *versus* non-protected animals showed a comparatively small amount of genes

8

ARTICLE IN PRESS

J.M. Marqués et al. / Immunobiology xxx (2011) xxx-xxx



Fig. 5. Protected mice showed increased IL-17A⁺ CD4⁺ T-cells. On day 0, C57BL/6 mice were treated i.n. with a sublethal dose of *S. pneumoniae* serotype 1 (Pn1, 4×10^4 CFU) or saline. On day 7, mice were challenged i.n. with a Pn1 lethal dose (2×10^7 CFU) or saline. Lungs were harvested 24 h after the lethal dose (n = 3) and processed for flow cytometry analysis. (A) Results are expressed as mean \pm SEM of the percentage of IL17A⁺ or IFN γ^+ positive cells in the total of CD45⁺ viable cells acquired for each group. (B) Mean \pm SEM of the percentage of CD3⁺CD4⁺IL17A⁺ or CD3⁺CD4⁺IFN- γ cells in the total of IL17A- or IFN γ -positive cells, respectively for each group. For statistical analysis Student's *T* test was used.

differentially expressed (149 genes), with an overrepresentation of genes belonging to biological processes related to immunity and defense. Among these genes, we found a clear signature of upregulated IL17- as well as IFN- γ -related genes in protected animals. We also found marked differences in the dynamic of Il17a- and Ifngrelated genes expression between protected and non-protected animals. In non-protected animals, transcript levels of *ll17*-related genes remained low at all time points, whereas IFN- γ and IFN- γ related genes were highly upregulated by 48 h even surpassing the level of expression found in protected animals. This could be in line with previous reports showing that IFN- γ do not have a protective role in pneumococcal pneumonia (Rijneveld et al. 2002; van der Poll and Opal 2009). However, in protected animals we also found consistent evidence that a strong IFN- γ related response is present earlier in time. This response seems to be also regulated as seen by the upregulation of the negative factor Socs1. These results are consistent with other reports that postulated a beneficial role of IFN- γ as a phagocyte-activating cytokine in the

clearance of *S. pneumoniae* (Rubins and Pomeroy 1997; Yamamoto et al. 2004). Altogether, these results suggest the importance of a timely regulated IFN- γ response; an early response may be associated with protection, whereas a delayed response may contribute to the uncontrolled proinflammatory response that is a hallmark of pneumococcal pneumonia.

The T helper type 17 (Th17) cells produce several effectors cytokines including IL-17A, IL-17F and IL-22 and all have been shown to be critical for mucosal immunity against extracellular pathogens (Aujla et al. 2007; Curtis and Way 2009; Kolls and Khader 2010; Ye et al. 2001) and particularly in prevention of pneumococcal colonization (Lu et al. 2008; Moffitt et al. 2011). Moreover, it has been shown that zwitterionic polysaccharides like PnPS1 and Cell wall pneumococcal polysaccharide are able to induce IL-17A production and Th17 responses (Groneck et al. 2009; Lu et al. 2008). Accordingly, our results confirmed the presence of Th17 cells in lungs of protected animals. Furthermore, we showed that CD4⁺ T cells are one of the main IL-17A producing cell populations, but the

J.M. Marqués et al. / Immunobiology xxx (2011) xxx-xxx



Fig. 6. Protected mice showed increased numbers of neutrophils. On day 0, C57BL/6 mice were treated i.n. with a sublethal dose of *S. pneumoniae* serotype 1 (Pn1, 4×10^4 CFU) or saline. On day 7, mice were challenged i.n. with a Pn1 lethal dose (2×10^7 CFU) or saline. Lungs were harvested 24 h after the lethal dose (n = 5) and processed for flow cytometry analysis. Neutrophils (PMNs) were identified as Ly6G⁺ CD11b⁺ cells. Results are expressed as mean ± SEM for each group. Mann–Whitney Test was used for statistical analysis.

results also suggested that there are other cell populations involved in IL-17A production in the lungs. Other populations that might produce IL-17A include $\gamma\delta$ T cells (Ley et al. 2006), NKT cells (Michel et al. 2007), NKT-like cells (Ley et al. 2006), and CD8⁺ T cells (Happel et al. 2003). In particular, it has been recently shown that infection by pneumococcus induces the generation of IL-17A by $\gamma\delta$ T cell (McNeela et al. 2010).

Contrary to our findings regarding IL-17A, we found that CD4⁺ T cells were not the main producers of IFN- γ in protected animals. A recent report identified NK cells as the main source of IFN- γ in the lungs (McNeela et al. 2010). Instead, preliminary experiments conducted by us identified CD8⁺ T cells as a major producer of IFN- γ in the lungs (results not shown), but so far we cannot discard that lung NK cells are also a main producer in our experimental setup. Work aimed at evaluating this is currently undergoing.

It is well known that IL-17A mediates direct recruitment, differential activation and increased survival of neutrophils and macrophages (Kolls and Linden 2004; Lu et al. 2008; Miyamoto et al. 2003), and also plays an important role in activating epithelial innate immune responses (Ishigame et al. 2009). Otherwise, IL-17A has also been shown as an activating T cell cytokine in antigenspecific T-cell-mediated immune responses (Ishigame et al. 2009; Nakae et al. 2002). More recently, a single work using myelin oligodendrocyte glycoprotein (MOG)-specific T-cell receptor transgenic mice (2D2) has proposed that Th17 cells may serve as effective Bcell helpers (Mitsdoerffer et al. 2010). Here we show that protected mice had increased numbers of neutrophils, specific antibodies and a Th17 response. Overall, our results suggest that in protected animals Th17 cells may collaborate with B-cells in the production of specific antibodies to the zwitterionic PnPS1 that work together with recruited neutrophils for clearance of the infection.

In summary, our results support the critical role that antibodies and PMNs play in the protection against acute pneumococcal pneumonia, and also suggest that upregulation of IL-17- and a timely regulated IFN-γ-related gene expression response, together with development of a Th17 cell response, are relevant characteristics of the protective immunity against *S. pneumoniae* acute pneumonia. This information may be relevant for the development of new vaccines against *S. pneumoniae* infection.

Acknowledgments

This work was supported by a grant of the European Community (STREP grant SavinMucoPath INCO-CT-2006-032296) and ECOS-Sud (U08S02) as well as a grant from the Genopole Evry (to A.B.). F.X.P. was a recipient of a PhD fellowship from the Nausica combat sa Leucémie Association.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.imbio.2011.10.012.

References

- Abeyta, M., Hardy, G.G., Yother, J., 2003. Genetic alteration of capsule type but not PspA type affects accessibility of surface-bound complement and surface antigens of *Streptococcus pneumoniae*. Infect. Immun. 71, 218.
- AlonsoDeVelasco, E., Verheul, A.F., Verhoef, J., Snippe, H., 1995. Streptococcus pneumoniae: virulence factors, pathogenesis, and vaccines. Microbiol. Rev. 59, 591.
- Aujla, S.J., Dubin, P.J., Kolls, J.K., 2007. Interleukin-17 in pulmonary host defense. Exp. Lung Res. 33, 507.
- Basset, A., Thompson, C.M., Hollingshead, S.K., Briles, D.E., Ades, E.W., Lipsitch, M., Malley, R., 2007. Antibody-independent, CD4⁺ T-cell-dependent protection against pneumococcal colonization elicited by intranasal immunization with purified pneumococcal proteins. Infect. Immun. 75, 5460.
- Brown, J.S., Hussell, T., Gilliland, S.M., Holden, D.W., Paton, J.C., Ehrenstein, M.R., Walport, M.J., Botto, M., 2002. The classical pathway is the dominant complement pathway required for innate immunity to *Streptococcus pneumoniae* infection in mice. Proc. Natl. Acad. Sci. U.S.A. 99, 16969.
- Brysbaert, G., Pellay, F.X., Noth, S., Benecke, A., 2010. Quality assessment of transcriptome data using intrinsic statistical properties. Genomics Proteomics Bioinformatics 8, 57.
- Cobb, B.A., Kasper, D.L., 2005. Zwitterionic capsular polysaccharides: the new MHCIIdependent antigens. Cell. Microbiol. 7, 1398.
- Curtis, M.M., Way, S.S., 2009. Interleukin-17 in host defence against bacterial, mycobacterial and fungal pathogens. Immunology 126, 177.
- Chung, D.R., Kasper, D.L., Panzo, R.J., Chitnis, T., Grusby, M.J., Sayegh, M.H., Tzianabos, A.O., 2003. CD4⁺ T cells mediate abscess formation in intra-abdominal sepsis by an IL-17-dependent mechanism. J. Immunol. 170, 1958.
- Fukuyama, Y., King, J.D., Kataoka, K., Kobayashi, R., Gilbert, R.S., Oishi, K., Hollingshead, S.K., Briles, D.E., Fujihashi, K., 2010. Secretory-IgA antibodies play an important role in the immunity to *Streptococcus pneumoniae*. J. Immunol. 185, 1755.
- Groneck, L., Schrama, D., Fabri, M., Stephen, T.L., Harms, F., Meemboor, S., Hafke, H., Bessler, M., Becker, J.C., Kalka-Moll, W.M., 2009. Oligoclonal CD4⁺ T cells promote host memory immune responses to Zwitterionic polysaccharide of *Streptococcus pneumoniae*. Infect. Immun. 77, 3705.
 Happel, K.I., Zheng, M., Young, E., Quinton, L.J., Lockhart, E., Ramsay, A.J., Shellito, J.E.,
- Happel, K.I., Zheng, M., Young, E., Quinton, L.J., Lockhart, E., Ramsay, A.J., Shellito, J.E., Schurr, J.R., Bagby, G.J., Nelson, S., Kolls, J.K., 2003. Cutting edge: roles of Tolllike receptor 4 and IL-23 in IL-17 expression in response to *Klebsiella pneumoniae* infection. J. Immunol. 170, 4432.
- Hortal, M., Sehabiague, G., Camou, T., Iraola, I., Estevan, M., Pujadas, M., 2008. Pneumococcal pneumonia in hospitalized Uruguayan children and potential prevention with different vaccine formulations. J. Pediatr. 152, 850. Ishigame, H., Kakuta, S., Nagai, T., Kadoki, M., Nambu, A., Komiyama, Y., Fujikado, N.,
- Ishigame, H., Kakuta, S., Nagai, T., Kadoki, M., Nambu, A., Komiyama, Y., Fujikado, N., Tanahashi, Y., Akitsu, A., Kotaki, H., Sudo, K., Nakae, S., Sasakawa, C., Iwakura, Y., 2009. Differential roles of interleukin-17A and -17F in host defense against mucoepithelial bacterial infection and allergic responses. Immunity 30, 108.
- Kadioglu, A., Coward, W., Colston, M.J., Hewitt, C.R., Andrew, P.W., 2004. CD4-Tlymphocyte interactions with pneumolysin and pneumococci suggest a crucial protective role in the host response to pneumococcal infection. Infect. Immun. 72, 2689.
- Kadioglu, A., Gingles, N.A., Grattan, K., Kerr, A., Mitchell, T.J., Andrew, P.W., 2000. Host cellular immune response to pneumococcal lung infection in mice. Infect. Immun. 68, 492.
- Kayhty, H., Auranen, K., Nohynek, H., Dagan, R., Makela, H., 2006. Nasopharyngeal colonization: a target for pneumococcal vaccination. Expert Rev. Vaccines 5, 651.
- Kolls, J.K., Khader, S.A., 2010. The role of Th17 cytokines in primary mucosal immunity. Cytokine Growth Factor Rev. 21, 443.
- Kolls, J.K., Linden, A., 2004. Interleukin-17 family members and inflammation. Immunity 21, 467.
- Ley, K., Smith, E., Stark, M.A., 2006. IL-17A-producing neutrophil-regulatory Tn lymphocytes. Immunol. Res. 34, 229.

G Model

10

J.M. Marqués et al. / Immunobiology xxx (2011) xxx-xxx

- Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods 25, 402.
- Lu, Y.J., Gross, J., Bogaert, D., Finn, A., Bagrade, L., Zhang, Q., Kolls, J.K., Srivastava, A., Lundgren, A., Forte, S., Thompson, C.M., Harney, K.F., Anderson, P.W., Lipsitch, M., Malley, R., 2008. Interleukin-17A mediates acquired immunity to pneumococcal colonization, PLoS Pathog, 4, e1000159. Lynch 3rd, J.P., Zhanel, G.G., 2009. *Streptococcus pneumoniae*: epidemiology, risk
- factors, and strategies for prevention. Semin. Respir. Crit. Care Med. 30, 189.
- Malley, R., 2010. Antibody and cell-mediated immunity to Streptococcus pneumoniae: implications for vaccine development. J. Mol. Med. 88, 135.
- Malley, R., Trzcinski, K., Srivastava, A., Thompson, C.M., Anderson, P.W., Lipsitch, M., 2005. CD4⁺ T cells mediate antibody-independent acquired immunity to pneumococcal colonization. Proc. Natl. Acad. Sci. U.S.A. 102, 4848.
- McNeela, E.A., Burke, A., Neill, D.R., Baxter, C., Fernandes, V.E., Ferreira, D., Smeaton, S., El-Rachkidy, R., McLoughlin, R.M., Mori, A., Moran, B., Fitzgerald, K.A., Tschopp, J., Petrilli, V., Andrew, P.W., Kadioglu, A., Lavelle, E.C., 2010. Pneumolysin activates the NLRP3 inflammasome and promotes proinflammatory cytokines independently of TLR4. PLoS Pathog. 6, e1001191.
- Michel, M.L., Keller, A.C., Paget, C., Fujio, M., Trottein, F., Savage, P.B., Wong, C.H., Schneider, E., Dy, M., Leite-de-Moraes, M.C., 2007. Identification of an IL-17producing NK1.1(neg) iNKT cell population involved in airway neutrophilia. J. Exp. Med. 204, 995.
- Mitsdoerffer, M., Lee, Y., Jager, A., Kim, H.J., Korn, T., Kolls, J.K., Cantor, H., Bettelli, E., Kuchroo, V.K., 2010. Proinflammatory T helper type 17 cells are effective B-cell helpers. Proc. Natl. Acad. Sci. U.S.A. 107, 14292.
- Miyamoto, M., Prause, O., Sjostrand, M., Laan, M., Lotvall, J., Linden, A., 2003. Endogenous IL-17 as a mediator of neutrophil recruitment caused by endotoxin
- exposure in mouse airways. J. Immunol. 170, 4665. Moffitt, K.L., Gierahn, T.M., Lu, Y.J., Gouveia, P., Alderson, M., Flechtner, J.B., Higgins, D.E., Malley, R., 2011. T(H)17-based vaccine design for prevention of *Streptococ*cus pneumoniae colonization. Cell Host Microbe 9, 158.
- Munoz, N., V.A.N.M., L., Marques, J.M., Rial, A., Sirard, J.C., Chabalgoity, J.A., 2010. Mucosal administration of flagellin protects mice from Streptococcus pneumoniae lung infection. Infect. Immun.
- Nakae, S., Komiyama, Y., Nambu, A., Sudo, K., Iwase, M., Homma, I., Sekikawa, K., Asano, M., Iwakura, Y., 2002. Antigen-specific T cell sensitization is impaired in IL-17-deficient mice, causing suppression of allergic cellular and humoral responses. Immunity 17, 375.
- Normark, B.H., Normark, S., 2002. Evolution and spread of antibiotic resistance. J. Intern. Med. 252, 91.
- Noth, S., Benecke, A., 2005. Avoiding inconsistencies over time and tracking difficulties in Applied Biosystems AB1700/Panther probe-to-gene annotations. BMC Bioinformatics 6, 307.
- Noth, S., Brysbaert, G., Benecke, A., 2006. Normalization using weighted negative second order exponential error functions (NeONORM) provides robustness against asymmetries in comparative transcriptome profiles and avoids false calls. Genomics Proteomics Bioinformatics 4, 90.
- Paterson, G.K., Mitchell, T.J., 2006. Innate immunity and the pneumococcus. Microbiology 152, 285.
- Poolman, J.T., 2004. Pneumococcal vaccine development. Expert Rev. Vaccines 3, 597
- Rial, A., Lens, D., Betancor, L., Benkiel, H., Silva, J.S., Chabalgoity, J.A., 2004. Intranasal immunization with a colloid-formulated bacterial extract induces an acute inflammatory response in the lungs and elicits specific immune responses. Infect. Immun. 72, 2679
- Richards, L., Ferreira, D.M., Miyaji, E.N., Andrew, P.W., Kadioglu, A., 2010. The immunising effect of pneumococcal nasopharyngeal colonisation; protection against future colonisation and fatal invasive disease. Immunobiology 215, 251.

- Rijneveld, A.W., Lauw, F.N., Schultz, M.J., Florquin, S., Te Velde, A.A., Speelman, P., Van Deventer, S.J., Van Der Poll, T., 2002. The role of interferon-gamma in murine pneumococcal pneumonia. J. Infect. Dis. 185, 91.
- Roche, A.M., King, S.J., Weiser, J.N., 2007. Live attenuated Streptococcus pneumoniae strains induce serotype-independent mucosal and systemic protection in mice. Infect. Immun. 75, 2469.
- Roche, A.M., Weiser, J.N., 2010. Identification of the targets of cross-reactive antibodies induced by Streptococcus pneumoniae colonization. Infect. Immun. 78, 2231.
- Rosseau, S., Hocke, A., Mollenkopf, H., Schmeck, B., Suttorp, N., Kaufmann, S.H., Zerrahn, J., 2007. Comparative transcriptional profiling of the lung reveals shared and distinct features of Streptococcus pneumoniae and influenza A virus infection. Immunology 120, 380.
- Rubins, J.B., Pomeroy, C., 1997. Role of gamma interferon in the pathogenesis of bacteremic pneumococcal pneumonia. Infect. Immun. 65, 2975
- Salyers, A.W.D., 2002. Bacterial Pathogenesis. A Molecular Approach, 2nd edition. ASM Press.
- Trzcinski, K., Thompson, C., Malley, R., Lipsitch, M., 2005. Antibodies to conserved pneumococcal antigens correlate with, but are not required for, protection against pneumococcal colonization induced by prior exposure in a mouse model. Infect. Immun. 73, 7043.
- Trzcinski, K., Thompson, C.M., Srivastava, A., Basset, A., Malley, R., Lipsitch, M., 2008. Protection against nasopharyngeal colonization by Streptococcus pneumoniae is mediated by antigen-specific CD4 $^{+}$ T cells. Infect. Immun. 76.2678
- van der Poll, T., Opal, S.M., 2009. Pathogenesis, treatment, and prevention of pneumococcal pneumonia. Lancet 374, 1543.
- Van Maele, L., Carnoy, C., Cayet, D., Songhet, P., Dumoutier, L., Ferrero, I., Janot, L., Erard, F., Bertout, J., Leger, H., Sebbane, F., Benecke, A., Renauld, J.C., Hardt, W.D., Ryffel, B., Sirard, J.C., 2010. TLR5 signaling stimulates the innate produc-tion of IL-17 and IL-22 by CD3(neg)CD127⁺ immune cells in spleen and mucosa. J. Immunol. 185, 1177.
- Velez, C.D., Lewis, C.J., Kasper, D.L., Cobb, B.A., 2009. Type I Streptococcus pneumoniae carbohydrate utilizes a nitric oxide and MHC II-dependent pathway for antigen presentation. Immunology 127, 73.
- Wernette, C.M., Frasch, C.E., Madore, D., Carlone, G., Goldblatt, D., Plikaytis, B., Benjamin, W., Quataert, S.A., Hildreth, S., Sikkema, D.J., Kayhty, H., Jonsdottir, I., Nahm, M.H., 2003. Enzyme-linked immunosorbent assay for quantitation of human antibodies to pneumococcal polysaccharides. Clin. Diagn. Lab. Immunol. 10, 514
- WHO, 2008. 23-Valent pneumococcal polysaccharide vaccine. WHO position paper. Wkly. Epidemiol. Rec. 83, 373.
- Yamamoto, N., Kawakami, K., Kinjo, Y., Miyagi, K., Kinjo, T., Uezu, K., Nakasone, C., Nakamatsu, M., Saito, A., 2004. Essential role for the p40 subunit of interleukin-12 in neutrophil-mediated early host defense against pulmonary infection with Streptococcus pneumoniae: involvement of interferon-gamma. Microbes Infect. 6,1241.
- Ye, P., Garvey, P.B., Zhang, P., Nelson, S., Bagby, G., Summer, W.R., Schwarzen-berger, P., Shellito, J.E., Kolls, J.K., 2001. Interleukin-17 and lung host defense against Klebsiella pneumoniae infection. Am. J. Respir, Cell Mol. Biol. 25, 335.
- Zemlickova, H., Crisostomo, M.I., Brandileone, M.C., Camou, T., Castaneda, E., Corso, A., Echaniz-Aviles, G., Pasztor, M., Tomasz, A., 2005. Serotypes and clonal types of penicillin-susceptible Streptococcus pneumoniae causing invasive disease in children in five Latin American countries. Microb. Drug Resist. 11, 195.
- Zhang, Z., Clarke, T.B., Weiser, J.N., 2009. Cellular effectors mediating Th17dependent clearance of pneumococcal colonization in mice. J. Clin. Invest. 119, 1899

I.3 .Rôle clé de l'IFN- γ et IL-17 dans les cellules dendritiques des voies respiratoires en réponse à un pré-traitement à l' α -galactosylcéramide lors de la protection contre la pneumonie aiguë due à *Streptococcus pneumoniae*.

Le troisième papier réalisé dans le cadre de cette étude est en cours de publication dans The Journal of Infectious Diseases sous le titre « Key role for respiratory dendritic cells, IFN-γ and IL-17 in protection against acute Streptococcus pneumoniae pneumonia in response to α-galactosylceramide. » par Ivanov S, Fontaine J, Paget C, Macho-Fernandez E, Van Maele L, Renesson J, Maillet I, Muñoz N, Wolf, Rial A, <u>Léger H</u>, Ryffel B, Chabalgoity J. A, Sirard J-C, Benecke A, Faveeuw C and Trottein F. A.

L'administration d' α -galactosylcéramide (α -GalCer), un glycosphingolipide, permet d'activer les cellules invariant Natural Killer T (iNKT), une population de lymphocytes T α β lipid-reactive. Cette activation exogène est une approche prometteuse pour contrôler les infections bactériennes respiratoires résistantes aux antibiotiques. Cette étude a permis la caractérisation des mécanismes de protection contre les infections, létales, par *S. pneumoniae (sérotype 1)* grâce à l' α -GalCer. *S. pneumoniae (sérotype 1)* est induit des infections aiguës à des voies respiratoires.

Le glycosphingolipide α -GalCer a été administré aux souris par voie intranasale avant l'infection par *S. pneumoniae*. Cette étude a montré que les cellules dendritiques respiratoires CD103 + (DC) jouent un rôle majeur dans l'activation (libération d'IFN- γ et d'IL-17) des cellules pulmonaires iNKT tandis que les cellules de poumon CD11b^{high} DC et les macrophages alvéolaires et interstitiels, jouent une fonction peu importante. Après l'infection, *S. pneumoniae* est rapidement éliminé (2-4 h) des espaces alvéolaires, un phénomène qui dépend des DC et des neutrophiles des voies respiratoires. Les macrophages ne semblent pas jouer de rôle dans cette activation ni sur l'induction précoce de la production d'IFN- γ et d'IL-17. Le pré-traitement à l' α -galactosylceramide (α -GalCer) entraine donc une protection totale contre une infection léthale avec *S. pneumoniae* serotype 1.

L'analyse transcriptomique révèle que cette protection est associée avec l'activation de nombreux gènes stimulés par les interferons ou associés à IL-17.

En conclusion, ces données éclairent une nouvelle fonction des cellules CD103 + DC pulmonaire dans l'activation des cellules iNKT. Cette étude permet de mieux comprendre comment les infections à bactéries extracellulaires encapsulées sont contrôlées par les organismes, permettant ainsi de concevoir de nouvelles approches thérapeutiques.

Projet annexe n°2 _ Exploration des gènes régulés par HMGA1 à l'échelle génomique en collaboration avec le Dr. Eilebrecht

Le petit ARN nucléaire 7SK (7SK snRNA) contrôle négativement la transcription en inactivant le facteur de transcription positif b (P-TEFb). Il est aussi un composant des complexes d'initation de la transcription HIV dépendante et indépendant de la protéine Tat.

Des études récentes ont montré que l'ARN 7SK contrôle également de façon directe l'activité transcriptionnelle de la protéine HMGA1 en interagissant directement via sa boucle 2 (7SK L2) avec le premier motif A/T-hook DNA binding de HMGA1. HMGA1 est un facteur maitre de régulation de l'expression génique ; sa dérégulation est associée à presque tous les types de cancer chez l'homme. En effet, le degré de surexpression HMGA est corrélé à la malignité de la tumeur et au potentiel métastatique de la tumeur.

Les auteurs ont développé plusieurs ARNs 7SK L2 chimères avec l'ARN 2 exprimé du virus d'Epstein- (EBER2) pour cibler les fonctions de régulation de la transcription de HMGA1. L'efficacité des chimères 7SK L2-EBER2 à interférer avec l'activité de transcription de HMGA1 par les grandes dépasse l'efficacité des ARN 7SK de type sauvage en raison de la forte activité du promoteur EBER2. En outre, la chimère 7SK L2-EBER2 n'interfère pas avec la transcription du facteur de transcription positif b (P-TEFb) ou avec la formation du 7SK sn / hnRNP.

La comparaison des effets de l'ARN 7SK de type sauvage ARN sur la dynamique du transcriptome avec les effets induits par les deux mutants L2 7SK ainsi que les changements dans l'expression des gènes après l'inhibition de HMGA1, permettent l'identification et la caractérisation des effets dépendants et indépendants de HMGA1 induits par 7SK snRNA. Par ailleurs, cette étude a mis en évidence une activté régulatrice de 7SK ARN L2 indépendante de P-TEFb et HMGA1.

HMGA1-dependent and independent 7SK RNA gene regulatory activity

Sebastian Eilebrecht,¹ Christophe Bécavin,¹ Hélène Léger,¹ Bernd-Joachim Benecke² and Arndt Benecke^{1,*}

¹Institut des Hautes Études Scientifiques & CNRS USR3078; Bures sur Yvette; France; ²Department of Biochemistry; Ruhr University Bochum; Bochum, Germany

Key words: HMGA1, 7SK RNA, gene expression regulation, RNA chimeras, transcriptome profiling, chromatin, long non-coding RNA

The small nuclear 7SK RNA negatively controls transcription by inactivating positive transcription elongation factor b (P-TEFb) and is an integral component of Tat-dependent and independent HIV-1 transcription initiation complexes. 7SK RNA has recently been shown to also directly control HMGA1 transcription activity. HMGA1 is a master regulator of gene expression and its deregulation is associated with virtually any type of human cancer. The degree of HMGA1 overexpression thereby correlates with tumor malignancy and metastatic potential. 7SK snRNA directly interacts through its loop2 (7SK L2) with the first A/T-hook DNA binding motif of HMGA1. We have developed several 7SK L2 RNA chimera with the Epstein Barr Virus expressed RNA 2 (EBER2) to target HMGA1 function in transcription regulation. The efficiency of interfering with HMGA1 transcription activity by the chimeric 7SK L2—EBER2 fusions by large exceeds the efficiency of 7SK wild-type RNA due to the stronger EBER2 promoter activity. Furthermore, the 7SK L2—EBER2 chimera do not interfere with P-TEFb controlled transcription elongation or the formation of 7SK sn/hnRNPs. The comparison of the effects of wild-type 7SK RNA on cellular transcriptome dynamics with those induced by the two 7SK L2 mutants as well as the changes in gene expression following inhibition of HMGA1 allow the identification and characterization of HMGA1-dependent and independent effects of 7SK snRNA. We furthermore also present evidence for P-TEFb and HMGA1-independent 7SK RNA L2 regulatory activity.



The small nuclear (sn) 7SK RNA belongs to the highly abundant non-coding regulatory RNAs in eukaryotic cells and is synthesized by RNA polymerase III.¹⁻⁴ It has been shown to be a negative regulator of the transcription elongation reaction of RNA polymerase II by inhibiting the positive transcription elongation factor b (P-TEFb),^{5,6} and thus to act as a global repressor for the expression of genes transcribed by RNA polymerase II. Beyond that, 7SK RNA is involved in the transcription regulation of diverse viral pathogens via its interaction with P-TEFb. During early HIV-1 transcription, 7SK RNA is released by the Tat protein and the TAR RNA from the inactive P-TEFb complex, resulting in efficient transcription of viral genes.^{7,8} Recent studies have shown that the amount of inactive 7SK/P-TEFb snRNP can be controlled by the human T-Lymphotropic Virus Type 1 (HTLV-1) Tax protein during HTLV-1 infection.⁹

We recently discovered a second major role of this small nuclear RNA in the negative regulation of the function of the high mobility group protein HMGA1 by directly competing with DNA for binding to the first A/T-hook motif of the protein.¹⁰ The interaction of 7SK RNA with HMGA1 is mediated by the second stem-loop structure of the RNA (L2), which is not involved in the formation of the P-TEFb snRNP or 7SK

hnRNP complexes.¹⁰⁻¹² The main function of HMGA1 is the regulation of gene expression by altering chromatin-structure, changing the DNA binding-affinity of transcription factors or modifying nucleosome positions.¹³⁻¹⁶ Depending on the position of the target sequence within the corresponding gene, HMGA1 activates or inactivates gene expression. The majority of genes affected in their expression by the interaction between 7SK RNA and HMGA1 have been shown to be repressed by HMGA1.¹⁰ HMGA1 overexpression has been detected in virtually every type of cancer and moreover, HMGA1 has been ascribed a role as a major factor during cancerogenesis, repressing the expression of tumor repressors on the one hand and enhancing the expression of oncogenes on the other.¹⁷⁻²¹ Given these key functions during malignant transformation, HMGA1 has been considered as a promising drug target for cancer therapy.²² Due to the negative regulatory role of 7SK RNA on HMGA1 function, the overexpression of 7SK RNA, or rather of the HMGA1 binding 7SK L2 substructure, has to be considered as a potential novel therapeutic approach to control HMGA1. Here we report on two important observations regarding the control of gene expression by 7SK snRNA. First, we have investigated the transcriptome dynamics induced by the overexpression of 7SK wild-type RNA and compared those to our previous review in reference 10 and additional transcriptome profiles obtained with different 7SK L2-EBER2

^{*}Correspondence to: Arndt Benecke; Email: arndt@ihes.fr Submitted: 10/18/10; Revised: 11/18/10; Accepted: 11/22/10 DOI: 10.4161/rna.8.1.14261

chimera in order to dissect 7SK RNA activity on HMGA1 from its activity on P-TEFb regulation. Using signaling network inference methodology we reconstruct and compare 7SK RNA, HMGA1 and 7SK L2 dependent changes on the inferred signalosome activity. Second, by comparing the transcription regulatory properties of two 7SK L2 mutants with both wild-type L2 and the entire 7SK RNA we identify a yet unknown activity of the 7SK L2 substructure in transcription regulation, which seems also HMGA1-independent.

Results

7SK snRNA induced transcriptome dynamics. We had previously focused our attention to the characterization of the specific role of 7SK RNA in regulating HMGA1-dependent chromatin dynamics and transcription regulation.¹⁰ In this study we aimed at understanding the relative contribution of HMGA1 to the regulatory activity of 7SK snRNA.

EMSAs with 7SK loop2 (L2) RNA and extracts from HMGA1a-FLAG (the major and longest alternative spliced HMGA1 isoform C-terminal FLAG fusion¹⁰) expressing HEK293 cells reveal a specific 7SK L2 binding activity in comparison to non-transfected cells (Fig. 1A). In the presence of HMGA1a-FLAG approximately half of the amount of labeled 7SK L2 RNA forms a specific complex with HMGA1a-FLAG, which can be super-shifted by the addition of anti-FLAG antibody, but not by anti-GST antibody (Fig. 1A and lanes 3-5). However, 7SK L2 complexes of high molecular size, that are independent from HMGA1a-FLAG overexpression can be detected with both, extracts from non-transfected as well as from HMGA1a-FLAG expressing HEK293 cells (Fig. 1A and lanes 2-5, arrows). Though our previous results point at a major role of the 7SK L2 substructure in the regulation of HMGA1 function these findings may indicate additional cellular components with a 7SK L2 binding activity.

We thus set out to globally identify genes whose expression is impacted by overexpression of full length human 7SK RNA in HEK293 cells. We therefore expressed full length human 7SK RNA controlled by the wild-type 7SK promoter in these cells and compared the downstream transcriptome with the gene expression profile of mock-transfected (empty pUC18) control cells. RT-qPCR analyses revealed a significant (p < 0.01)

overexpression of 7SK RNA (about 190%), when compared to the mock-transfected control (Fig. 1B). Transcriptome analyses of these cells, using three independent biological replicates for either condition, revealed a total of 844 genes regulated by 7SK RNA in a statistically significant manner (p < 0.01). Of these genes, 194 (or 23%) showed an enhanced expression after 7SK overexpression, whereas 650 genes (or 77%) were repressed in their expression (Fig. 1C). Notably, the genes, whose expression was upregulated upon overexpression of 7SK RNA were significantly enriched with HMGA1 target genes.¹⁰ Among the global set of genes regulated by 7SK RNA we were able to classify certain subgroups with regard to the cellular functions of the corresponding gene products. Genes products having known activities in mRNA transcription regulation (Fig. 1D), genes related to oncogenesis (Fig. 1E), genes related to Interleukins (IL) (Fig. 1F), genes coding for different histones (Fig. 1G) and genes, whose products are involved in cell growth, differentiation and apoptosis (Fig. 1H) are regulated by full-length 7SK RNA. Notably, the interleukin receptor 2α (IL2-R α) gene, which is found to be significantly downregulated upon 7SK RNA overexpression (Fig. 1F), is also known to be positively regulated by HMGA1,^{10,16,23-29} agreeing with the idea of 7SK RNA as a negative regulator of HMGA1 function. These observations are also in perfect agreement with gene ontology enrichment analyses. We identify as sole biological process enriched in the 7SK RNA target genes 'mRNA transcription' (p < 2.53 x 10E⁻³, binomial distribution). Furthermore, the enrichment analysis for molecular function identifies DNA binding proteins including histones and transcription factors as statistically significantly enriched (Fig. 1I).

The 7SK gene promoter belongs to the RNA polymerase III type 3 promoters and is located entirely upstream of the transcription start site. It consists of an Octamer motif (Oct-1),³⁰⁻³² a CACCC-Box,³³ a Sph Postoctamer Homology (SPH) domain^{34,35} as well as a Proximal Sequence Element (PSE),³⁶ and a TATA-Box (Fig. 2A). This promoter efficiency results in about 2 x 10⁵ copies of 7SK RNA per HeLa cell,¹ making 7SK RNA one of the most abundant small nuclear RNAs in eukary-otic cells. Hence, it is no surprise that the overexpression of 7SK RNA controlled by its own promoter, as done here, results only in an about two-fold increase of expression of this RNA (Fig. 1B), limiting the effectiveness of the approach. Nevertheless,

Figure 1 (See opposite page). Changes in gene expression upon overexpression of full length 75K snRNA. (A) 200 ng of FAM6-labeled 75K L2 RNA were incubated with cellular extract from non-transfected (n.t.) in comparison to HMGA1a-FLAG transfected (HMGA1a-FLAG t.) HEK293 cells. The specificity of the resulting HMGA1a-FLAG/7SK L2 RNA complex was verified by addition of anti-FLAG antibody. Anti-GST antibody was used for control. The complexes were analyzed in EMSAs compared to 7SK L2 RNA alone (control). The HMGA1a-FLAG/7SK L2 RNA complex is assigned. Not yet identified 7SK L2 RNA containing complexes are marked by arrows. (B) HEK293 cells were transfected with p7SK coding for full length human 7SK RNA controlled by the full 75K gene promoter. The amount of 75K RNA was quantified by RT-qPCR analyzes in comparison to mock transfected control cells. (C) Heatmap of the subset of statistically significant (p < 0.01) changes in gene expression between HEK293 cells overexpressing 7SK snRNA (p7SK) compared to mock transfected cells (control). L indicates the log,-fold change in expression, S indicates the signal, P indicates the p-value and the roman numbers indicate independent biological replicates. (D) Genes related to transcription, which show a significantly increased (red) or decreased (blue) expression upon overexpression of full length 7SK RNA compared to the control. (E) Genes related to cancer, which show a significant differential expression upon overexpression of full length 7SK RNA compared to the control. (F) Interleukin (IL) related genes, which show a significant differential expression upon overexpression of full length 7SK RNA compared to the control. (G) Histones, which show a significant differential expression upon overexpression of full length 7SK RNA compared to the control. (H) Genes related to cell growth, differentiation or apoptosis, which are significantly differentially expressed upon overexpression of 7SK RNA compared to the control. (I) Gene ontology enrichment analysis for 'molecular functions'. The graph represents the fraction of genes corresponding to one of the statistically significantly (binomial distribution) enriched ontologies compared to the entire set of 7SK RNA target genes. In grey is shown the fraction of genes not included in any of the over-represented ontologies.



7SK RNA intra-cellular concentrations are apparently tightly controlled, and the overexpression of exogenous 7SK RNA sufficiently impacts on the dynamic equilibrium between bound and free 7SK RNA, as we are able to identify several hundreds of 7SK RNA target genes in the transcriptome profiling (Fig. 1C).

As 7SK RNA acts as a global repressor of RNA polymerase II transcription elongation mediated by the stem-loop structures L1, L3 and L4,^{11,12} the expression of full length 7SK RNA will target both, HMGA1 function and P-TEFb function. In order to better characterize and dissect the relative impact on these two distinct 7SK RNA target activities, we next set out to compare



Figure 2. Comparison of the viral EBER2 gene promoter with the 7SK gene promoter. (A) Schematic representation of the 7SK gene promoter: The entirely gene external 7SK promoter contains an Octamer motif (Oct-1), a CACCC-Box, a Sph Post Octamer Homology (SPH) domain as well as a Proximal Sequence Element (PSE) and a TATA-Box. The positions of each element with respect to the transcription start (+1; arrow) is annotated. The coding sequence for 7SK RNA is highlighted in grey. (B) Schematic representation of the viral EBER2 gene promoter: The gene external promoter elements consist of a Sp1 transcription factor-(SP1) and an Activating Transcription Factor (ATF)-binding site as well as an EBER TATA box (ETAB). The gene internal elements are Box A and Box B regulatory sequences. The positions of each element with respect to the transcription start (+1; arrow) is given. The coding sequence for EBER2 RNA is highlighted in grey. (C) The transcription rate of 7SK RNA was compared to that of EBER2 RNA in in vitro transcription analyses using HeLa nuclear extract. (D) The signal intensities in (C) were calculated using ImageJ software. The signal of 7SK RNA was arbitrarily set to 1. (E) The signal intensities shown in (D) were converted to relative expression rates considering the numbers of ³²P-labeled U-residues in each of the RNAs (76 for 7SK RNA and 37 for EBER2 RNA).

the wild-type 7SK RNA induced transcriptome dynamics to the previously reported effects as well as newly generated 7SK substructure chimera-induced transcriptome dynamics.

7SK—EBER2 chimera and EBER2 background activity. One of the strongest promoters known among the genes transcribed by RNA polymerase III is the type 2 promoter of the EBER2 gene, which codes for the 172 nt long EBER2 RNA of the Epstein Barr Virus.³⁷ It contains gene external as well as gene internal promoter elements, both being essential for efficient transcription by RNA polymerase III. To the former elements belong a binding site for the Sp1 transcription factor, an Activating Transcription Factor (ATF) binding site³⁸ as well as the EBER TATA-Box (ETAB) (**Fig. 2B**). The gene internal promoter elements comprise A- and B-Box regulatory sequences^{39,40} (**Fig. 2B**). The efficient transcription of the EBER2 gene results in more than 10⁷ copies of EBER2 RNA per infected cell.³⁷

We compared the transcription efficiency of the EBER2 gene with that of the 7SK gene by in vitro transcription analyses with HeLa nuclear extract (Fig. 2C). The quantified and normalized signals (Fig. 2D) reveal a more than 75-fold higher transcription rate of EBER2 RNA compared to 7SK RNA (Fig. 2E), mirroring roughly the cellular copy numbers of the two transcripts

Figure 3 (See opposite page). Secondary structures of EBER2 RNA, truncated control RNAs and EBER2 7SK L2 RNA chimera. The secondary structure of each RNA was calculated using RNAstructure 4.4 software.⁵¹ Gene-internal box (A) and box (B) elements of the strong viral EBER2 promoter are highlighted in grey. (A) The secondary structure of wild type EBER2 RNA consists of a basal stem and two stem-loop structures (La and Lb). (B) In the EBER2-Lb RNA nt79 to nt141 of wild type EBER2 RNA have been deleted. (C) A shortened version of the EBER2-Lb RNA is lacking nt 79 to nt 148 of wild type EBER2 RNA. (D) In the EBER2 7SK L2 RNA chimera the Lb (nt 80 to nt 140) was exchanged by the loop 2 structure (nt 113 to nt 154) of wild type 7SK RNA. The 7SK sequence is marked in green. (E) A mutated version of the EBER2 7SK L2 RNA chimera (EBER2 7SK L2 RNA. The 7SK sequence is marked in green. (E) A mutated version of the EBER2 7SK L2 RNA chimera (EBER2 7SK L2 RNA. The mutated 7SK sequence is marked in red. (F) A second mutant of the EBER2 7SK L2 RNA chimera (EBER2 7SK L2 m137) contains a three nucleotides mutation (GGC to CCG) at position nt 89 to nt 91 in addition to the mutation in the EBER2 7SK L2 m137 RNA chimera in order to restore the predicted secondary structure of the 7SK L2. The mutated 7SK sequence is marked in red here as well. (G) 2 ng of ³²P-labeled EBER2 7SK L2 m137 and EBER2 7SK L2 RNA were incubated with increasing amounts of immuno-purified HMGA1a-FLAG fusion protein (approx. 5 ng and 10 ng). The resulting complexes were analyzed in EMSAs compared to the corresponding RNAs alone (-). (H) The HMGA1-binding activity of EBER2 7SK L2 RNA and EBER2 7SK L2 m137 RNA from (G) was calculated using ImageJ. The activity of EBER2 7SK L2 m137 RNA was set arbitrarily to 1.



Figure 3. For figure legend, see page 146.

determined in earlier studies.^{1,37} The high expression rate, as well as the composite secondary structure of EBER2 RNA make it a potentially very potent tool to overexpress small hairpin-RNA structures. The EBER2 RNA mainly consists of a basal stem and the two stem-loop structures La and Lb (**Fig. 3A**). The gene internal promoter elements of the EBER2 gene are located exclusively in the basal stem as well as the stem-loop structure La (Fig. 3A), predestining the stem-loop structure Lb to be exchanged by the hairpin-RNA structure of interest.

We have previously constructed a chimeric RNA, in which the sequence from +81 nt to +140 nt of wild type EBER2 RNA was exchanged by the sequence from +113 nt to +154 nt (L2) of



Figure 4. Comparison of the effects of EBER2-Lb and EBER2-Lb-short on the cellular transcriptome of HEK293 cells. (A) Venn diagram of EBER2-Lb versus wild-type EBER2 and EBER2-Lb-short versus wild-type EBER2 target genes at p < 0.01. (B) Principal component analysis in correlation space (PCC) of the weighted averages of the transcriptome profiles from the indicated biological conditions. The inertia of the two first principal components are depicted along with their relative contributions. The arrows illustrate the effects of EBER2-Lb, EBER2-Lb-short and the added effect of the 7SK L2 RNA expression. (C) Euclidean distance-based hierarchical complete linkage clustering of the variance-weighted averages of the indicated transcriptome profiles.

wild type 7SK RNA (EBER2 7SK L2, Fig. 3D).10 We were able to show that this chimera maintains the HMGA1 binding activity of wild type 78K L2 RNA as well as the promoter activity of the wild type EBER2 gene.¹⁰ Transcriptome analyses following overexpression of this RNA chimera compared to a mutated 7SK L2 chimeric construct with a reduced HMGA1 binding activity (EBER2 7SK L2 m137, Fig. 3E) revealed the regulation of a wide variety of HMGA1-controlled genes.10 A second mutated RNA chimera discussed here (EBER2 7SK L2 m122/m137, Fig. 3F) presumably restores the 7SK L2 secondary structure, but does not restore HMGA1 binding activity when compared to EBER2 7SK L2 m137 or EBER2-7SK L2.10 This finding indicates that HMGA1 binding is contingent on the actual sequence of the L2 of 7SK RNA rather than only the (predicted) secondary structure since otherwise the combined EBER2 7SK L2 m122/m137 mutant would restore binding to wild-type L2 levels (Fig. 3D and F). In order to better quantify the effect of the 7SK L2 single mutant on HMGA1 binding, we performed additional experiments to detect the HMGA1 binding activity of EBER2 7SK L2 RNA in comparison to EBER2 7SK L2 m137 RNA in EMSAs with increasing amounts of immuno-purified HMGA1a-FLAG fusion protein (Fig. 3G). The relative affinity of each RNA to HMGA1a was then quantified, revealing, that the EBER2 7SK L2 RNA chimera exhibits a 3.5-fold higher affinity to HMGA1 than the mutated EBER2 7SK L2 m137 RNA chimera (Fig. 3H).

As we have previously reported, EBER2 as well as EBER2-Lb have profound effects by themselves on the cellular transcriptome,⁴¹ calling for very careful analysis and correction of the transcriptome activity of EBER2-based chimera. In the hope to construct EBER2-Lb backbones with even further reduced background activity, we designed an additional control RNA lacking the EBER2 Lb substructure (EBER2-Lb-short, **Fig. 3C**, compare to **B**). This Lb deleted EBER2 mutant, however, also contains all gene internal promoter elements. In comparison to the

wild type EBER2 RNA the longer control RNA (EBER2-Lb; Fig. 3B) has been used previously to distinguish effects of the basal stem and the stem-loop structure La from those of the stem-loop structure Lb of EBER2 RNA, both having drastic effects on gene expression.⁴¹ We here have thus investigated and compared the effects of both EBER2-Lb and EBER2-Lb-short on the cellular transcriptome using the effects of EBER2 wildtype RNA as reference. As reported previously in reference 41, EBER2-Lb has strong effects on the cellular transcriptome and acts quite differently than EBER2 alone, resulting in the statistically significant differential expression of some 3,500 genes (Fig. 4A). The mechanism by which EBER2 and EBER2-Lb interfere with cellular transcription or the measurement itself are at this point completely obscure. When comparing the effects of the new EBER2-Lb-short construct to EBER2 wild-type in five independent biological replications to the overlap of the effects with the corresponding comparison of EBER2-Lb and EBER2, two observations can be made. First, EBER2-Lb-short also has significant effects on the measured transcriptome when compared to EBER2 wt. Second, these effects are quite distinct from those induced by EBER2-Lb as the overlap is minimal (Fig. 4A). Taken together, the EBER2-Lb-short construct confirms the not understood but highly reducible effects of EBER2 expression on the determination of the transcriptome, and furthermore, also confirms that different EBER2 deletion mutants can have significantly different effects.

In order to decide, which of the two EBER2-Lb back-bone constructs was best suited for the characterization of 7SK RNA substructure chimera, we directly compared their gene transcription activity to the ones previously reported for EBER2-7SK L2, the EBER2-7SK L2 m137 (both generated in the context of the EBER2-Lb) including control transfected cells (pUC18). The variance-weighted average transcriptome profiles calculated from the three to five biological replicates of each biological condition

were visualized using principal component analysis in correlation space (PCC, Fig. 4B). Such type of analysis allows a geometric representation and visualization of high-dimensional objects in low dimensional space. In this case, the relationship between the different indicated biological conditions to one-another is represented. Note that the distance between any two biological conditions is defined as the averaged correlation calculated from the weighted averages of the gene expression profiles of those conditions. The orientation of biological conditions to each other is a resulting observable of the analysis. The arrows superposed onto the PCC results indicate the independent contributions of the EBER2 backbones and the 7SK L2 effects. In case of the EBER2-Lb, the effect of the 7SK L2 m137 mutant is intermediary to the 7SK L2 wild-type versus EBER2-Lb control, which is in accordance with our previous observations of the 7SK L2 m137 mutant being still able to target HMGA1 albeit with lowered affinity.¹⁰ Furthermore, Euclidean distance-based hierarchical clustering using a complete linkage method (Fig. 4C), identifies the EBER2-Lb-short transcriptome effects to be much closer related to the EBER2 wild-type than EBER2-Lb and wild-type 7SK RNA. It is for these two observations that we have decided that the EBER2-Lb-short backbone is less well suited than the previously constructed EBER2-Lb. Consequently, for all following analyses, the EBER2-Lb has been used as control or as backbone for the different 7SK substructure chimera. The investigation of the different effects of EBER2 and EBER2 internal deletion constructs on the cellular transcriptome and its determination, however, will need to be thoroughly investigated further, especially in any analysis of EBER2-based chimeric RNAs.

Comparison of the 7SK RNA and HMGA1 target genes. We next compared the transcriptome dynamics induced by wild-type 7SK RNA overexpression to those previously reported for HMGA1 knock-down and EBER2-7SK L2/EBER2-7SK L2 m137 overexpression on the gene level (Fig. 5A and B). As reported review in reference 10, the expression of the EBER2 7SK L2 RNA chimera resulted in the significantly (p < 0.01) differential expression of 2715 genes (Fig. 5A and green), whereas the expression of EBER2 7SK L2 m137 RNA led to a significant (p < 0.01) change in expression of 761 genes (Fig. 5B and red). Noteworthy, the number of genes regulated by each RNA roughly mirrors the ability of the corresponding RNA to bind HMGA1 and correlates with the relative binding affinities determined here (Fig. 3G and H). To check the effects of the overexpressed RNA chimera for HMGA1 targets, we compared each set of differentially expressed genes with the gene expression profiles recorded after shRNA-mediated knockdown of endogenous HMGA1.¹⁰ We have previously shown that a total of 1,649 (or 61%) of the genes affected by EBER2 7SK L2 RNA expression were also found to be HMGA1 target genes.10 More interestingly, even though the total number of differentially expressed genes is much lower downstream of EBER2 7SK L2 m137 RNA expression in comparison to EBER2 7SK L2 RNA, the common subset with HMGA1 targets contains 649 genes (85%) of the total number of genes expressed differentially in this condition (Fig. 5B and red). Among the 844 significantly (p < 0.01) regulated full length 7SK RNA targets, we observe a common subset

with HMGA1-controlled genes of 53 (6%) (Fig. 5B and blue). With 20 genes out of 194 genes (10%) having increased expression levels upon 7SK overexpression, the common subset with HMGA1 targets is significantly elevated in comparison to the 33 genes out of 650 genes (5%), whose expression is decreased upon 7SK overexpression (Fig. 5B and blue). To further substantiate and illustrate the close relationship between 7SK L2 and HMGA1 target gene sets, and thereby indirectly lend further support to the analyses presented with respect to the 'rescue' mutant (Fig. 6B) as well as the conclusions drawn here, we further investigated the overlap between the EBER2 7SK L2 and the EBER2-7SK L2 m137 target genes. 723 genes (or 95%, Fig. 5C, yellow) of those genes whose expression is affected in a statistically significant manner after overexpression of the EBER2 7SK L2 m137 RNA chimera (Fig. 5C and red) are also found to be among the EBER2 7SK L2 RNA target genes (Fig. 5C and green). Almost all genes statistically significantly (p < 0.01) differentially expressed by either of these conditions are also affected by the corresponding other condition (Fig. 5D). The Pearson correlation coefficients of either set (Fig. 5D and $R^2 = 96$, green and $R^2 = 97$, red) are comparably high with respect to the joint subset (Fig. 5D and $R^2 = 97$, yellow), indicating, that both RNA chimera target the same set of genes, solely differing in the efficacy of gene expression regulation which appears to be a direct function of HMGA1 binding activity (Fig. 3G). Therefore, the EBER2-7SK L2 m137 targets are a (complete) subset of the EBER2-7SK L2 targets.

Among the genes with statistically significant (p < 0.05) elevated expression in all of the biological conditions compared here, we find e.g., the NF κ B activator TANK, the melanoma antigen MAGED4, the histone H2AFY, the differentiation regulator METRNL and the transcription factor SOX4 (Fig. 5H). All 20 genes in the common subset of genes, whose expression is increased upon 7SK RNA overexpression, and which are known HMGA1 target genes, also show an increased expression upon shRNA-mediated knockdown of HMGA1 (Fig. 5F). Finally, all 33 genes in the common subset between 7SK RNA and HMGA1 target genes, are overexpressed after shRNA-mediated knockdown of HMGA1 (Fig. 5G).

Taken together, these comparative analyses establish two important facts. First, as expected, wild-type 7SK RNA and HMGA1 target genes are only partially overlapping. While, based on the comparison with the much higher expressed EBER2-7SK L2 chimera, we believe that most, if not all, HMGA1 target genes can also be affected by 7SK RNA, the reverse is not true. Some 7SK RNA targets are either independent of HMGA1 or regulated in the opposite manner by 7SK RNA and HMGA1 (Fig. 5G). This likely is explained by opposing roles of HMGA1 and P-TEFb in the expression of these genes, and both activities being simultaneously targeted by full-length 7SK RNA. Second, 7SK RNA also has positive effects on gene expression which are independent of HMGA1 and P-TEFb inhibition, as not all of the genes induced in their expression levels are also HMGA1 target genes (Fig. 5B) and can not be P-TEFb target genes as no inhibitory activity of free (not 7SK-bound) P-TEFb has been identified. Those genes, including the ones identified through



Figure 5. For figure legend, see page 151.

the comparison with the EBER2-7SK L2 m122/m137 'rescue' mutant (Fig. 6B), might indicate that other, yet to be identified, roles of 7SK RNA in transcription regulation might exist. Note that since the target cells analyzed here do neither contain the HIV1 promoter nor express TAT protein, this (these) additional regulatory functions of 7SK RNA would indeed be novel.

Comparison of the inferred gene encoded networks downstream of 7SK RNA, HMGA1 and EBER2-7SK L2. Large-scale gene expression information can be used to infer the effective networks of the presumably expressed gene products under the arguably important hypothesis that the gene expression changes translate linearly to changes in cellular protein levels. We have used the available gene product annotation information of the Kegg, PANTHER and GO databases to reconstruct the HEK293 protein network and infer its activity based on the changes in the transcriptome dynamics following expression of wild-type 7SK RNA, as described here, as well as the previously reported HMGA1 and EBER2-7SK L2 dependent activities (**Fig. 6A**).¹⁰ The inferred networks have then been visualized using Cytoscape software⁴² and colored using a linear gradient from blue (reduced activity) to grey (unchanged activity) to red (induced activity). Note that only the largest spanning and Figure 5 (See opposite page). EBER2 7SK L2 RNA chimera are a potent tool to affect HMGA1-dependent gene expression. (A and B) The genes expressed differentially in a significant manner (p < 0.01) after overexpression of the EBER2 7SK L2 and EBER2 7SK L2 m137 RNA chimera in comparison to the overexpression of EBER2-Lb RNA, which are shown in green and red, respectively, were compared to the genes with a significant (p < 0.01) change in expression after shRNA-mediated knockdown of HMGA1 (grey). The fraction of genes significantly differentially expressed in both conditions is shown in dark green and dark red, respectively. The genes significantly (p < 0.01) differentially expressed after full length 7SK RNA overexpression (blue) were compared to the genes affected by shRNA-mediated knockdown of HMGA1 (grey). The common subset between these conditions is shown in dark blue. The genes differentially expressed after full length 75K RNA overexpression are divided into negatively and positively regulated genes and compared to the genes affected by shRNA-mediated knockdown of HMGA1. The common subsets of all comparisons are shown as percentage of the genes affected by each condition in total as well as by the shRNA-mediated knockdown of HMGA1 (grey). (C) Comparison of the genes differentially expressed in a significant manner (p < 0.01) after EBER2 75K L2 RNA expression compared to expression of EBER2-Lb RNA (green) with the genes significantly (p < 0.01) differentially expressed after EBER2 7SK L2 m137 RNA expression compared to EBER2-Lb RNA expression (red). The common subset is marked in yellow. (D) The log₂-fold change of gene expression after overexpression of EBER2 7SK L2 RNA (x-axis) is compared to that after overexpression of EBER2 7SK L2 m137 RNA (y-axis). Genes significantly regulated by both conditions are marked in yellow. Those, which are solely regulated significantly by the overexpression of EBER2 75K L2 RNA are marked in green and those solely regulated by the overexpression of EBER2 75K L2 m137 RNA are marked in red. The error bars show the positive sd. (E) The common subset of genes expressed differentially after overexpression of EBER2 7SK L2 RNA compared to the control (EBER2-Lb) was ordered by the log,-fold changes (2,715 genes in (C); green). Those genes in the same time expressed differentially after EBER2 7SK L2 m137 RNA expression in comparison to the expression of EBER2-Lb RNA are plotted in yellow (723 genes in (C); yellow). The ratio of the number of genes differentially expressed in the overlap (yellow) per every 50 genes differentially expressed after EBER2 7SK L2 RNA overexpression in comparison to the control (green) is plotted against the log,-fold change in gene expression. (F) Genes, whose expression is increased upon both, 7SK RNA overexpression and HMGA1 knockdown. (G) Genes, whose expression is decreased upon 7SK RNA overexpression and increased after HMGA1 knockdown. (H) Different genes, regulated by all four conditions shown in (A and B) The error bars show the sd.

fully connected tree is depicted. As becomes immediately obvious from the reconstructed network and the inferred activities, the HMGA1 and EBER2-7SK L2 signalosomes are virtually identical in both the identity of the regulated nodes as well as their degree of inferred activity. This is in complete agreement with our previous proposition of all HMGA1 target genes being also 7SK RNA targets.¹⁰ The comparison of those two networks with the one inferred from wild-type 7SK RNA activity, however, reveals significant differences with respect to both the identity of regulated components as well as their level or even sign of regulation (Fig. 6A). This is expected, as expression of fulllength 7SK RNA also should display activity on P-TEFb target genes^{5,6} in addition to HMGA1 target genes. The overexpression of wild-type 7SK RNA under control of its own promoter being far less efficient than the expression of the EBER2-7SK L2 chimera, dampen the overall effects of 7SK RNA when compared to the chimera. Note that we have also reconstructed the EBER2-7SK L2 m137 inferred signalosome activity. As expected, and in complete agreement with our previous gene-based analyses,¹⁰ it resembles closely the EBER2-7SK L2 and HMGA1 networks (data not shown, see also Fig. 6). Given the fact that the EBER2-7SK L2 m137 mutant does not abrogate the effects of the 7SK L2 substructure on HMGA1-dependent transcription control, we decided to also determine the transcriptome dynamics induced by the expression of the EBER2-7SK L2 m122/m137 double mutant (Fig. 3). This mutant, as discussed above, preserves the predicted secondary structure of the 7SK L2 substructure, while changing the nucleotide sequence of the L2 on two blocks of three consecutive positions. The HEK293 transcriptome has thus been recorded in four independent biological replications following the overexpression of this double mutant (see Materials and Methods).

Evidence of HMGA1-independent 7SK L2 activities on target genes. When comparing the gene expression profiles obtained using the EBER2-7SK L2 m122/m137 double mutant with those of the other EBER2 chimera, wild-type 7SK RNA, and HMGA1 shRNA-mediated knock-downs, we made the

intriguing observation of several genes being statistically significantly (p < 0.05) regulated by wild-type 7SK RNA, the chimeric EBER2-7SK L2, and the double mutant m122/m137, but not HMGA1 and the single mutant EBER2-7SK L2 m137 (Fig. 6B). Despite the small number of genes displaying this regulatory phenotype-a dozen, of which half are of unknown function-this finding is of particular importance as it suggests that there are other, yet unidentified, factors being targeted by the 7SK L2 in addition to HMGA1 in HEK293 cells. Those must be either of low abundance, or have very restricted in terms of amplitude, or specific in terms of identity, effects on gene expression to result in such low penetration in the present analyses. Furthermore, this finding suggests that the interaction of these unknown factors (directly or indirectly) with 7SK L2 RNA depends on different properties of the L2 substructure. HMGA1, as our previous and present experiments show, displays sequence specificity towards 7SK RNA in that the nucleotides mutated in the 7SK L2 m137 mutant reduce significantly binding to HMGA1.10 The double mutant 7SK L2 m122/m137 has similarly reduced HMGA1 binding activity despite having a 'rescued' secondary structure.¹⁰ Therefore, the predicted secondary structure of 7SK L2-at least at this position-does not impact HMGA1 recognition. The by these comparisons predicted unknown factor, however, displays a distinct binding preference. While its activity can be altered by the wild-type 7SK L2 structure as well as the entire 7SK RNA, it seems no longer targeted by the 7SK L2 m137 mutant (data not shown). The 'rescue' double mutant 7SK L2 m122/m137, however, again seems to affect this yet to be identified activity, as the effects of EBER2-7SK L2 m122/m137 overexpression are similar if not stronger than those of the wild-type 7SK L2 (Fig. 6B), suggesting that in this case the secondary structure of the 7SK L2 is a critical component of interaction.

These observations are further supported when average Euclidean distances are computed for the HMGA1 and 7SK target genes and used to hierarchically cluster the variance-weighted, averaged transcriptome profiles of the wild-type 7SK RNA, the EBER2-7SK L2 chimera and the HMGA1 transcriptomes



profiles using available PANTHER, Kegg and GO annotation were depicted for wild-type 7SK RNA, HMGA1 and 7SK L2 using Cytoscape software (www.cytoscape.org/). Predicted activity of gene-products in the three networks is illustrated by a blue (diminished activity) to grey (unchanged activity) to red (increased activity) linear color gradient. Only connected components are drawn. (B) Examples of genes upregulated by EBER2-7SK L2, EBER2-7SK L2 m122/m137, wild-type 7SK RNA, but not HMGA1 are shown. Log₂-fold changes are given. (C) Euclidean distance-based hierarchical complete linkage clustering of the variance-weighted averages of the indicated biological conditions. The Euclidean distances were computed only for those genes identified to be regulated by HMGA1 and 7SK RNA. (D) Pearson correlation analysis of the variance-weighted averages of the indicated biologic conditions with respect to the 7SK RNA and HMGA1 average transcriptome profiles. Only the +0.50 to +1.00 space is shown and the diagonal indicated.

(Fig. 6C). The resulting cluster structure reveals a closer relationship of the double/rescue mutant with 7SK RNA when compared to the single mutant. Furthermore, using correlation analysis based on the Pearson correlation coefficients calculated for the same sets of target genes it can be shown that the EBER2-7SK L2 wild-type overexpression results in gene expression changes equally well correlated with the effects of HMGA1 knock-downs and wild-type 7SK RNA overexpression, whereas the 'rescue' mutant is slightly closer related to the wild-type 7SK RNA effects than to the HMGA1 effects (Fig. 6D). Please note also that these observations are corroborating the evidence for 7SK L2 highmolecular weight complexes forming in extracts from HEK293 cells (Fig. 1A and arrows).

7SK L2 effects on transcriptome dynamics are evolutionary conserved. To further substantiate our results obtained in the human HEK293 cell line and especially further investigate the significance of the HMGA1-independent effects of 7SK L2, we thought to analyse the evolutionary conservation of HMGA1-dependent and independent regulatory effects. To this end we decided to overexpress the EBER2 7SK L2 RNA chimera in COS-7 cells, a kidney cell line derived from African green monkeys (AGM, Chlorocebus sabaeus). AGM's are an important model for AIDS research as they are natural hosts for SIV infection. We have previously established how transcriptome analyses can be achieved in AGMs using human microarrays,57 and successfully used this technique to investigate the differences in gene expression in CD4+ Tcells from AIDS nonprogressor versus progressor primate species.58 Transcriptomes from EBER2 7SK L2 RNA overexpressing COS-7 cells were compared to COS-7 cells expressing the EBER2-Lb RNA and revealed a total of 203 genes being significantly (p < 0.05) upregulated in their expression (Fig. 7A). 151 genes (or 74%) out of these are found to be overexpressed upon EBER2 7SK L2 RNA expression in HEK293 cells as well. Among these genes we found the ribosomal protein RPS14, the nephronophthisis associated protein NPHP4, the transcription factor SOX4 and



Figure 7. EBER2 7SK L2 RNA overexpression affects comparable sets of genes in COS-7 and HEK293 cells. (A) Heat-map of the subset of genes, whose expression is significantly (p < 0.05) upregulated in COS-7 cells after EBER2 7SK L2 RNA expression compared to EBER2-Lb expression. The expression of the same set of genes in HEK293 cells at identical conditions is also displayed for comparison. L indicates the log₂-fold change in expression, S indicates the signal, P indicates the p-value and the roman numbers indicate independent biological replicates. (B) Genes, whose expression is significantly increased upon EBER2 7SK L2 RNA overexpression in COS-7 (light green) as well as in HEK293 cells (dark green) and whose expression is increased as well upon 7SK RNA full length overexpression (blue) as well as HMGA1 knockdown (grey) in HEK293 cells. The error bars show the sd. (C) Differential expression of the genes shown in Figure 6B after EBER2 7SK L2 RNA expression in COS-7 cells compared to EBER2-Lb expression. The error bars show the sd. (D) Different genes, regulated differentially in a significant manner after EBER2 7SK L2 RNA overexpression in comparison to the overexpression of EBER2-Lb RNA in both, HEK293 cells (dark green) and COS-7 cells (light green). The error bars show the sd.

the cytochrome b5 domain containing protein FLJ32499, whose expression is significantly enhanced upon EBER2 7SK L2 RNA expression, full length 7SK RNA overexpression and shRNA-mediated knockdown of HMGA1 in HEK293 cells (Fig. 7B). Four out of the six genes identified as HMGA1-independent 7SK L2 target genes (Fig. 6B), are also found to be upregulated upon EBER2 7SK L2 RNA expression in COS-7 cells (Fig. 7C). Please note, that due to non-exclusive mappings between the AGM and human gene sets, as well as sequence divergence between both

species, it is not clear whether the absence of regulation of SNX9 and EIF3S5 is indeed reflecting a biological difference between humans and AGMs, or rather a technical detection problem. Among the EBER2 7SK L2 RNA-affected genes in COS-7 cells shown in **Figure 7A** there are 33, which are also affected in a significant (p < 0.01) manner after EBER2 7SK L2 RNA expression in HEK293 cells (**Fig. 7D**). Therefore, the regulatory properties of 7SK RNA on HMGA1-dependent and independent gene regulation are at least partially conserved across primates.



Figure 8. Schematic drawing of 75K RNA and its different cellular interaction partners. 75K snRNA has been shown to be stabilized by both, a γ -monomethylephosphate-GTP cap on its 5'-end, which is synthesized by the methylephosphate capping enzyme (MePCE), and by the interaction with the La-related protein LARP7. One function of 75K RNA is the inactivation of the positive transcription elongation factor b (T-TEFb) by its interaction through L1, L3 and L4 with HEXIM1 and P-TEFb (CdK9 and CycT1/T2), which results in an inhibition of the RNA polymerase II transcription elongation reaction. 75K RNA, which is released from P-TEFb, has also been shown to be captured in a complex with diverse heterogenous nuclear ribonuclear particles (hnRNPs), among those RHA, hnRNP A1, hnRNP Q, hnRNP R and hnRNP A2/B1. In both of these interactions, with P-TEFb and with the different hnRNPs, only the stem-loop structures L1, L3 and L4 are involved. Recently we were able to demonstrate an interaction of 75K RNA with the chromatin master regulator and oncogenic marker HMGA1, a member of the high mobility group proteins. This interaction takes place via the stem-loop 2 (L2) substructure of 75K RNA and has profound effects on HMGA1-dependent gene expression. The EBER2 75K L2 RNA chimera discussed in this study carry solely the 75K L2 substructure and thus are a potent tool to directly target HMGA1 function, avoiding off-target effects on P-TEFb function, that would be expected for full length 75K RNA overexpression. Furthermore, the two different L2 mutants affecting either the sequence (m122/m137) or the predicted secondary structure in addition to the sequence (m137) display different activities with respect to HMGA1 and lead to the postulation of a yet to be identified activity ('X') associated with 75K RNA L2.

Discussion

With the recent discovery of HMGA1 as being a major cellular target of the 7SK snRNA10 in addition to the positive transcription elongation factor P-TEFb,5,6 it has become feasible to understand the transcription regulatory effects of this highly abundant and essential polymerase III transcript. Notably, the fact that P-TEFb is a non-essential component of the basal transcription machinery, as well as P-TEFb-independent effects of 7SK RNA depletion,43 have for long strongly argued for other activities of 7SK RNA in transcription regulation. While having previously focused our attention exclusively on the effects of the 7SK L2 substructure on HMGA1,10 we have here investigated the effects of wild-type 7SK RNA on the cellular transcriptome in the same model system and systematically compared the regulatory effects of full-length 7SK RNA and the effects of the isolated 7SK L2 substructure on HMGA1 target gene expression. These comparisons have allowed to establish HMGA1-dependent and HMGA1-independent effects of 7SK RNA on HMGA1 target genes (Fig. 5) as well as genes not affected by HMGA1 at all (Fig. 6B). These findings have been substantiated by the investigation of the effects of two 7SK L2 mutants on the cellular transcriptome dynamics (Figs. 4-6). As suggested by the

previous analyses,10 the 7SK L2 m137 mutant targets HMGA1 in a manner identical to wild-type 7SK L2 (Fig. 5D), however, displays lowered binding affinity and concomitantly lowered inhibitory activity on the positive and negative regulatory actions of HMGA1. In contrast, the double 7SK L2 m122/m137 'rescue' mutant with restored (predicted) secondary structure, and previously not investigated for its effects on transcriptome dynamics for displaying similar binding affinity to HMGA1 as the single mutant,¹⁰ reveals HMGA1-independent 7SK L2 substructure activity on a few target genes. This finding does not only potentially explain the existence of 7SK L2-dependent gene regulatory phenomena which are not detectable by HMGA1 knock-downs, but importantly also leads to the hypothesis of yet unidentified factors being targeted by 7SK RNA and in particular the L2 substructure (Fig. 8). Importantly, the comparative investigation of both mutants in HMGA1 binding and their transcriptome activities, has lead to the observation that while HMGA1 binding strength is affected by the nucleotide sequence of the 7SK L2 substructure but not necessarily by the secondary structure of this precise part of 7SK snRNA. This seems to be different for the unknown activity targeted by 7SK RNA, as different behavior is observed whether or not the predicted secondary structure is destroyed (L2 m137) or not (L2 m122/m137, Fig. 8). Finally, on a more technical note, we have extended previous observations on the various background activities of different EBER2 internal deletion mutants,^{10,41} and demonstrated the need for carefully controlling the activity of chimera constructed using EBER2 as an expression vehicle for non-translated RNA substructures. These latter findings, albeit awaiting a molecular explanation, should further aid in the dissection of specific and non-specific effects of EBER2 chimera which turn out to be a very valuable tool for instance in the characterization of HMGA1 activity as reported on here and review in reference 10.

The further characterization of HMGA1-dependent and independent effects of 7SK snRNA on cellular gene expression provide a means to better understand the functional roles of both 7SK RNA and HMGA1 in physiology. HMGA1, being a 'hub' of nuclear function and through its activity tightly linked to oncogenesis and metastasis,^{13,17-22} might be targeted by the expression of the 7SK L2 substructure in a therapeutically mean-ingful manner. In this respect, our demonstration of at least partial evolutionary conservation of 7SK RNA mediated effects on HMGA1-dependent and independent gene expression using an African green monkey derived cell-line, further illuminates the significance of 7SK RNA in gene regulation and also sets the grounds for possibly investigating therapeutic strategies in a non-human primate model system.

Apart from a major role during oncogenesis HMGA1 has been shown to be involved in HIV-1 transcription. The 5'-long terminal repeat (LTR) of HIV-1 contains multiple high affinity HMGA1 binding sites, which in the majority of cases overlap with important binding sites for transcription factors.⁴⁴ HMGA1 therein enhances the affinity of the activating transcription factor-3 (ATF-3) for an activating protein-1 (AP-1) binding site located downstream of the transcriptional start site.⁴⁵ ATF-3 subsequently recruits the hSWI/SNF chromatin remodeling complex to the HIV-1 promoter resulting in transcription activation.⁴⁵

7SK RNA has been recently found to play important roles in HIV-1 transcription as well. During the first rounds of inefficient viral transcription the synthesized HIV-1 transcriptional transactivator Tat binds to a recently discovered site in the distal part of the 5'-stem-loop substructure of 7SK RNA (7SK loop1).⁴⁶ Thereby Tat removes HEXIM from its preferred position at two highly defined binding sites within the 7SK loop1 substructure.^{7,46} The transcribed transactivator RNA (TAR) finally replaces the 7SK/HEXIM snRNP from P-TEFb, resulting in the activation of P-TEFb.⁷ Both 7SK substructures, which are involved in the interaction with P-TEFb, the 5'-terminal stem-loop structure as well as the 3'-terminal stem loop structure,¹¹ have been studied more in detail.^{46,47} The latter has been found to exhibit structural similarities with HIV-1 TAR, supporting the image of TAR replacing 7SK RNA from P-TEFb.⁴⁷

The direct interaction of HMGA1 with 7SK RNA thus may depict plasticity between HIV-1 transcription activation by HMGA1 via ATF-3 and transcription elongation after Tat/TARdependent replacement of 7SK RNA from inactive P-TEFb. Imaginable, direct interactions of HMGA1 with Tat or with HIV-1 TAR have to be further investigated, even though a direct binding to TAR seems to be unlikely due to structural dissimilarities between TAR and the 7SK L2 substructure.

It remains to be seen whether or not 7SK snRNA also is capable of directly affecting other, yet unidentified, cellular factors as the analyses reported here suggest.

Materials and Methods

Cell culture. HEK293, COS-7 and HeLa cells were cultured as described previously in reference 41. Transfections were either performed by the calciumphosphate coprecipitation method⁴⁸ or by FugeneHD (Roche) according to the manufacturers instructions.

Molecular biology. Extract preparations. The preparation of nuclear extract for in vitro transcription analyses was performed as described previously in reference 49. To prepare extracts for EMSAs or immunopurifications, transfected cells were reconstituted in isotonic IP-lysis buffer (50 mM Tris/HCl, pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% TRITONTM X-100, protease inhibitor) and homogenized in a dounce-homogenizer before cellular debris was removed by high speed centrifugation (16,000x g). The extract was subsequently dialyzed against IP-buffer (20 mM HEPES/KOH, pH 7.9, 100 mM KCl, 0.2 mM EDTA, 20% glycerol).

ImmunopurificationofHMGA1a-FLAG. Immunoprecipitations of FLAG fusion proteins were performed with M2 anti-FLAG agarose (Sigma) as recommended by the manufacturer. Briefly, anti-FLAG agarose was incubated with extracts containing FLAG fusion proteins in IP-buffer (20 mM HEPES/KOH, pH 7.9, 100 mM KCl, 0.2 mM EDTA, 20% glycerol) for 2 hours at 4°C. The agarose was pelleted by centrifugation and the supernatant was removed. After washing three times with the 20-fold volume of IP-washing buffer (20 mM HEPES/KOH, pH 7.9, 100 mM KCl, 0.2 mM EDTA, 0,025% Tween 20), the HMGA1a-FLAG was eluted with the 5-fold volume of elution buffer (100 mM Glycine/HCl, pH 3.5). The elution fraction was subsequently dialzed against IP-buffer before use in EMSAs.

Electrophoretic mobility shift assays. For electrophoretic mobility shift assays (EMSAs), ³²P-labeled RNA was denatured for 5 minutes at 75°C and renatured by cooling slowly down to room temperature. The nucleic acid was incubated either with cellular extract or with immuno-purified protein in EMSA-buffer (10 mM HEPES/KOH, pH 7.9, 100 mM KCl, 5 mM MgCl₂, 0.25 mM DTT, 0.1 mM EDTA, 10% glycerol and 2 μ g yeast tRNA per lane) for 10 minutes at 37°C and subsequently for 10 minutes on ice. The resulting complexes were separated by 10% native PAGE and visualized by autoradiography. Signal intensities were calculated using ImageJ software.

Total RNA preparation. Total RNA from transfected cells was either prepared using the RNeasy Midi Kit (Qiagen) as recommended by the manufacturer or as described previously in reference 50.

In vitro transcription. In vitro transcription analyses to detect promoter activity were performed using nuclear extract from HeLa cells.⁴⁹ RNA for EMSAs was prepared using T7 RNA Polymerase (NEB) on a linearized template DNA containing the

coding sequence of the RNA of interest controlled by a T7 promoter, as recommended by the manufacturer. To label RNA with ³²P, the reaction mixture contained a 10-fold excess of ³²P-labeled α -UTP. Signal intensities were calculated using ImageJ software.

RNA structure prediction. The secondary structure of RNA was calculated by RNAstructure (Version 4.4).⁵¹

Quantitative RT-PCR analyses. qRT-PCR analyses were performed using SuperScriptTM 2 Reverse Transcriptase (Invitrogen), Lightcycler (Roche) and Fast Start DNA MasterPLUS SYBR Green I reaction mix (Roche) as recommended by the manufacturer. Primers used to quantify the housekeeping gene Ubiquitin were 5'-GTT GAG ACT TCG TGG TGG TG-3' (sense) and 5'-TCT CGA CGA AGG CGA CTA AT-3' (antisense). Primers to detect endogenous 7SK RNA were 5'-CAT CCC CGA TAG AGG AGG AC-3' (sense) and 5'-GCC TCA TTT GGA TGT GTC TG-3' (antisense).

Transcriptome analyses. For microarray analyses, RNA amplification, labeling, hybridization and detection were performed following the protocols supplied by Applied Biosystems using the corresponding kits (Applied Biosystems, ProdNo: 4339628 and 4336875). The data obtained were analyzed as described previously in reference 10, 41, 52–55. Transcriptome data, annotated to MIAME I+II standards, were deposited in the public database MACE (mace.ihes.fr):

(1) 7SK wild-type RNA overexpression: MACE Acc. No.: 2655275478

Reviewer login: 2655275478 password: ppEjkv9IJ_ (2) EBER2-Lb-short & EBER2-7SK L2 m122/m137: MACE

Acc. No.: 2147641814

Reviewer login: 2147641814 password: 9vBB0Nut4l (3) EBER2 7SK L2 & EBER2-Lb in COS-7: MACE Acc. No.: 2804044630

Reviewer login: 2804044630 password: ZZ-yrBstvA
In addition we used previously published transcriptome data:
(3) EBER2 and EBER2-Lb: MACE Acc. No.: 3034287118
(4) EBER2-7SK L2 and EBER2-7SK L2 m137: MACE Acc.

No.: 2979879726

(5) HMGA1 knock-down: MACE Acc. No.: 2970966830 Gene ontology enrichment analyses were also described previously in reference 41 and 55. The signalosome network inference

References

- Gurney T Jr, Eliceiri GL. Intracellular distribution of low molecular weight RNA species in HeLa cells. J Cell Biol 1980; 87:398-403.
- Matera AG, Ward DC. Nucleoplasmic organization of small nuclear ribonucleoproteins in cultured human cells. J Cell Biol 1993; 121:715-27.
- Zieve G, Benecke BJ, Penman S. Synthesis of two classes of small RNA species in vivo and in vitro. Biochemistry 1977; 16:4520-5.
- Zieve G, Penman S. Small RNA species of the HeLa cell: metabolism and subcellular localization. Cell 1976; 8:19-31.
- Nguyen VT, Kiss T, Michels AA, Bensaude O. 75K small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. Nature 2001; 414:322-5.
- Yang Z, Zhu Q, Luo K, Zhou Q. The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. Nature 2001; 414:317-22.

 Barboric M, Lenasi T. Kick-sTARting HIV-1 transcription elongation by 7SK snRNP deporTATion. Nat Struct Mol Biol 17:928-30.

- D'Orso I, Frankel AD. RNA-mediated displacement of an inhibitory snRNP complex activates transcription elongation. Nat Struct Mol Biol 17:815-21.
- Cho WK, Jang MK, Huang K, Pise-Masison CA, Brady JN. Human T-lymphotropic virus type 1 Tax protein complexes with P-TEFb and competes for Brd4 and 7SK snRNP/HEXIM1 binding. J Virol 2010; 84:12801-9
- Eilebrecht S, Brysbaert G, Wegert T, Urlaub H, Benecke BJ, Benecke A. 7SK small nuclear RNA directly affects HMGA1 function in transcription regulation. Nucleic Acids Research 2011; 39:2057-72.
- Egloff S, Van Herreweghe E, Kiss T. Regulation of polymerase II transcription by 7SK snRNA: two distinct RNA elements direct P-TEFb and HEXIM1 binding. Mol Cell Biol 2006; 26:630-42.

will be described in detail elsewhere (Bécavin et al. in preparation) and the network representation tool used was Cytoscape.⁴²

Constructs used in this study. *EBER2 7SK L2 fusion transcripts.* The EBER2 gene from -171 nt to +226 nt containing the complete viral promoter sequence was cloned into the EcoRI and HindIII restriction sites of puC18 vector (Fermentas). This clone was used to construct all EBER2 7SK L2 RNA chimera. To obtain the EBER2 7SK L2 construct, the sequence from +81 nt to +140 nt of wild-type EBER2 RNA was exchanged by the sequence from +113 nt to +154 nt of wild type 7SK RNA. To obtain the EBER2 7SK L2 m137 construct nucleotides +104 to +106 (CGG) of the EBER2 7SK L2 m122/m137 in addition the nucleotides +89 to +91 (CCG) were mutated to GGC. To obtain the EBER2-Lb (EBER2-Lb-short) construct, nucleotides +80 to +145 (+78 to + 148) of wild type EBER2 RNA were deleted.

HMGA1a-FLAG. To obtain a vector for the expression of FLAG fusion protein, the hybridized oligonucleotides 5'-CTA GAG GGC GAC TAC AAA GAC GAT GAC GAC AAA GGA TGA A-3' (sense) and 3'-TC CCG CTG ATG TTT CTG CTA CTG CTG TTT CCT ACT TGG CC-5' (antisense) were inserted into the XbaI and AgeI restriction sites of the vector peDNA4/TO/mye-his B (Invitrogen). The coding sequence of wild type HMGA1a from +1 nt to +321 nt was inserted into the PstI and XbaI restriction sites of the vector mentioned above.

P7SK. To overexpress full length 7SK RNA, the complete human 7SK gene including the full length promoter was cloned into the puC18 vector (Fermentas).⁵⁶

Acknowledgements

We are grateful to Brendan Bell for his helpful comments on the manuscript. Nicolas Tchitchek is thanked for instructions on the use of the microarray and relevant statistical analysis methodologies. This work was supported by the Centre National de la Recherche Scientifique (CNRS) and the Genopole Evry (to A.B.) as well as the Deutsche Forschungsgemeinschaft (BE 531/19-3, to: B.J.B). C.B. is a recipient of a pre-doctoral fellowship from the Agence Nationale de recherches sur le SIDA et les hépatites virales (ANRS). The authors declare to have no competing interests.

- Van Herreweghe E, Egloff S, Goiffon I, Jady BE, Froment C, Monsarrat B, Kiss T. Dynamic remodelling of human 7SK snRNP controls the nuclear level of active P-TEFb. EMBO J 2007; 26:3570-80.
- 13. Reeves R. Molecular biology of HMGA proteins: hubs of nuclear function. Gene 2001; 277:63-81.
- 14. Reeves R. HMGA proteins: flexibility finds a nuclear niche? Biochem Cell Biol 2003; 81:185-95.
- Reeves R. HMGA proteins: isolation, biochemical modifications and nucleosome interactions. Methods Enzymol 2004; 375:297-322.
- Reeves R, Beckerbauer L. HMGI/Y proteins: flexible regulators of transcription and chromatin structure. Biochim Biophys Acta 2001; 1519:13-29.
- Chiappetta G, Avantaggiato V, Visconti R, Fedele M, Battista S, Trapasso F, et al. High level expression of the HMGI (Y) gene during embryonic development. Oncogene 1996; 13:2439-46.

- Chiappetta G, Bandiera A, Berlingieri MT, Visconti R, Manfioletti G, Battista S, et al. The expression of the high mobility group HMGI (Y) proteins correlates with the malignant phenotype of human thyroid neoplasias. Oncogene 1995; 10:1307-14.
- Cleynen I, Van de Ven WJ. The HMGA proteins: a myriad of functions (Review). Int J Oncol 2008; 32:289-305.
- 20. Fusco A, Fedele M. Roles of HMGA proteins in cancer. Nat Rev Cancer 2007; 7:899-910.
- Hess JL. Chromosomal translocations in benign tumors: the HMGI proteins. Am J Clin Pathol 1998; 109:251-61.
- Reeves R, Beckerbauer LM. HMGA proteins as therapeutic drug targets. Prog Cell Cycle Res 2003; 5:279-86.
- Himes SR, Reeves R, Attema J, Nissen M, Li Y, Shannon MF. The role of high-mobility group I(Y) proteins in expression of IL-2 and T cell proliferation. J Immunol 2000; 164:3157-68.
- 24. John S, Reeves RB, Lin JX, Child R, Leiden JM, Thompson CB, Leonard WJ. Regulation of cell-typespecific interleukin-2 receptor alpha-chain gene expression: potential role of physical interactions between Elf-1, HMG-I(Y) and NFkappaB family proteins. Mol Cell Biol 1995; 15:1786-96.
- John S, Robbins CM, Leonard WJ. An IL-2 response element in the human IL-2 receptor alpha chain promoter is a composite element that binds Stat5, Elf-1, HMG-I(Y) and a GATA family protein. EMBO J 1996; 15:5627-35.
- Lehn DA, Elton TS, Johnson KR, Reeves R. A conformational study of the sequence specific binding of HMG-I (Y) with the bovine interleukin-2 cDNA. Biochem Int 1988; 16:963-71.
- Magnuson NS, Spies AG, Nissen MS, Buck CD, Weinberg AD, Barr PJ, Magnuson JA, Reeves R. Bovine interleukin 2: regulatory mechanisms. Vet Immunol Immunopathol 1987; 17:183-92.
- Reeves R, Elton TS, Nissen MS, Lehn D, Johnson KR. Posttranscriptional gene regulation and specific binding of the nonhistone protein HMG-1 by the 3' untranslated region of bovine interleukin 2 cDNA. Proc Natl Acad Sci USA 1987; 84:6531-5.
- Reeves R, Leonard WJ, Nissen MS. Binding of HMG-I(Y) imparts architectural specificity to a positioned nucleosome on the promoter of the human interleukin-2 receptor alpha gene. Mol Cell Biol 2000; 20:4666-79.
- Danzeiser DA, Urso O, Kunkel GR. Functional characterization of elements in a human U6 small nuclear RNA gene distal control region. Mol Cell Biol 1993; 13:4670-8.
- Murphy S, Pierani A, Scheidereit C, Melli M, Roeder RG. Purified octamer binding transcription factors stimulate RNA polymerase III—mediated transcription of the 75K RNA gene. Cell 1989; 59:1071-80.
- Sturm RA, Das G, Herr W. The ubiquitous octamerbinding protein Oct-1 contains a POU domain with a homeo box subdomain. Genes Dev 1988; 2:1582-99.

- Kleinert H, Bredow S, Benecke BJ. Expression of a human 7S K RNA gene in vivo requires a novel pol III upstream element. EMBO J 1990; 9:711-8.
- 34. Kunkel GR, Cheung TC, Miyake JH, Urso O, McNamara-Schroeder KJ, Stumph WE. Identification of a SPH element in the distal region of a human U6 small nuclear RNA gene promoter and characterization of the SPH binding factor in HeLa cell extracts. Gene Expr 1996; 6:59-72.
- Schaub M, Myslinski E, Schuster C, Krol A, Carbon P. Staf, a promiscuous activator for enhanced transcription by RNA polymerases II and III. EMBO J 1997; 16:173-81.
- Dahlberg JE, Blattner FR. Sequence of the promoteroperator proximal region of the major leftward RNA of bacteriophage lambda. Nucleic Acids Res 1975; 2:1441-58.
- Lerner MR, Andrews NC, Miller G, Steitz JA. Two small RNAs encoded by Epstein-Barr virus and complexed with protein are precipitated by antibodies from patients with systemic lupus erythematosus. Proc Natl Acad Sci USA 1981; 78:805-9.
- Howe JG, Shu MD. Epstein-Barr virus small RNA (EBER) genes: unique transcription units that combine RNA polymerase II and III promoter elements. Cell 1989; 57:825-34.
- Geiduschek EP, Tocchini-Valentini GP. Transcription by RNA polymerase III. Annu Rev Biochem 1988; 57:873-914.
- Rosa MD, Gottlieb E, Lerner MR, Steitz JA. Striking similarities are exhibited by two small Epstein-Barr virus-encoded ribonucleic acids and the adenovirusassociated ribonucleic acids VAI and VAII, Mol Cell Biol 1981; 1:785-96.
- Eilebrecht S, Pellay FX, Odenwalder P, Brysbaert G, Benecke BJ, Benecke A. EBER2 RNA-induced transcriptome changes identify cellular processes likely targeted during Epstein Barr Virus infection. BMC Res Notes 2008; 1:100.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003; 13:2498-504.
- Haaland RE, Herrmann CH, Rice AP. siRNA depletion of 7SK snRNA induces apoptosis but does not affect expression of the HIV-1 LTR or P-TEFb-dependent cellular genes. J Cell Physiol 2005; 205:463-70.
- 44. Henderson A, Bunce M, Siddon N, Reeves R, Tremethick DJ. High-mobility-group protein I can modulate binding of transcription factors to the U5 region of the human immunodeficiency virus type 1 proviral promoter. J Virol 2000; 74:10523-34.
- Henderson A, Holloway A, Reeves R, Tremethick DJ. Recruitment of SWI/SNF to the human immunodeficiency virus type 1 promoter. Mol Cell Biol 2004; 24:389-97.

- Muniz L, Egloff S, Ughy B, Jady BE, Kiss T. Controlling cellular P-TEFb activity by the HIV-1 transcriptional transactivator Tat. PLoS Pathog 6:1001152.
- Durney MA, D'Souza VM. Preformed protein-binding motifs in 7SK snRNA: structural and thermodynamic comparisons with retroviral TAR. J Mol Biol.
- Graham FL, van der Eb AJ. Transformation of rat cells by DNA of human adenovirus 5. Virology 1973; 54:536-9.
- Dignam JD, Martin PL, Shastry BS, Roeder RG. Eukaryotic gene transcription with purified components. Methods Enzymol 1983; 101:582-98.
- Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenolchloroform extraction. Anal Biochem 1987; 162:156-9.
 Zuker M. Computer prediction of RNA structure.
- Methods Enzymol 1989; 180:262-88.
 52. Noth S, Benecke A. Avoiding inconsistencies over time and tracking difficulties in Applied Biosystems AB1700/Panther probe-to-gene annotations. BMC Bioinformatics 2005; 6:307.
- Noth S, Brysbaert G, Benecke A. Normalization using weighted negative second order exponential error functions (NeONORM) provides robustness against asymmetries in comparative transcriptome profiles and avoids false calls. Genomics Proteomics Bioinformatics 2006: 4:90-109.
- Noth S, Brysbaert G, Pellay FX, Benecke A. Highsensitivity transcriptome data structure and implications for analysis and biologic interpretation. Genomics Proteomics Bioinformatics 2006; 4:212-29.
- 55. Wilhelm E, Kornete M, Targat B, Vigneault-Edwards J, Frontini M, Tora L, et al. TAF6delta orchestrates an apoptotic transcriptome profile and interacts function-
- ally with p53. BMC Mol Biol 11:10. 56. Surig D, Bredow S, Benecke BJ. The seemingly identical 75K and U6 core promoters depend on different transcription factor complexes. Gene Expr 1993;
- 3:175-85. 7. Jacquelin B, Mayau V, Brysbaert G, Regnault B, Diop
- OM, Arenzana-Seisdedos F, et al. Long oligonucleotide microarrays for African Green Monkey gene expression profile analysis. FASEB J 2007; 21:3262-71.
- Jacquelin B, Mayau V, Targat B, Liovat AS, Kunkel D, Petitjean G, et al. Strong but Rapidly Controlled Interferon Type I Response in Nonpathogenic SIV Infection of African Green Monkeys J. Clinical Invest 2009; 119:3544-55.

Projet annexe n°3 _ Exploration des gènes régulés par TAF6delta à l'échelle génomique en collaboration avec le professeur Bell de l'Université de Sherbrooke (Qc, Canada).

Le variant d'épissage alternatif TAF6, est une sous unité de facteur de transcription de la polymérase II, TFIID. TAF6 a été identifié et caractérisé par Brendan Bell et son équipe (Bell et al, 2001). Par ailleurs, le rôle de cette protéine est encore méconnu.

Une première étude réalisée en 2008 a permis de mettre en évidence l'implication de TAF6δ dans les mécanismes de l'apoptose sans la nécessité de l'implication du suppresseur de tumeur p53. Par ailleurs, il a été montré que TAF 6δ joue aussi un rôle pivot dans la régulation de la transcription, rôle qui a été abordé lors de cette étude.

Projet annexe n°4 _ Transmission transgénérationnelle du phénotype Prenatal restraint stress (PRS) en collaboration avec le Dr. Sara Morley-Fletcher du laboratoire de Glycobiologie Structurale et Fonctionnelle.

Des études épidémiologiques précédentes ont mise en évidence le fait qu'un traumatisme précoce pouvait être à l'origine de troubles anxieux à l'âge adulte. Le modèle d'étude du stress prénatal (PRS, prenatal restraint stress) chez le rat consiste à soumettre des rattes Sprague- gestantes à un stress de contention repeté 3 fois/jour pendant 45 minutes durant la dernière semaine de gestation. Leur progéniture mâle représente un modèle précieux pour l'étude de l'interaction entre facteurs environnementaux et le développement neurologique dans l'étude de la pathogenèse des comportements anxieux/dépressifs.

Lors de cette étude, nous avons évalué la possible transmission transgenerationnelle du stress prénatal via la ligne maternelle. Le poids et le comportement des mères stressés et des petits ont été évaluées par des tests classiques de même que l'anxiété après le sevrage. Le stress prénatal modifie l'expression des principaux marqueurs de la neuroplasticité neuronale de l'hippocampe tels que récepteurs létabotropiques au glutammate, BDNF ; cette l'altération neuroplasticité persistant à travers les générations. Les rats PRS présentent une hyperactivité de l'axe corticotrope qui persiste à travers les générations. Enfin, des études transcriptomiques ont été conduites pour cibler les gènes modulés de façon transgénérationnelle par le stress prénatal.

Ces données offrent une démonstration claire le PRS déclenche une programmation de certains gènes impliqués dans le façonnage du phénotype vulnérable qui caractérise ce modèle animal et que la programmation de ces gènes est transmise aux cours des générations.

Bibliographie

- Albert T J, Jason N, Ott M, Richmond T, Nuwaysir K, Nuwaysir E F, Stengele K P, and Green R D. 2003. "Light-directed 5'-->3' Synthesis of Complex Oligonucleotide Microarrays." *Nucleic Acids Research* Vol 31, Issue 7, Pp. e35.
- Allenby G., Bocquel M T, Saunders M, Kazmer S, Speck J, Rosenberger M, and Lovey A. 1993. "Retinoic Acid Receptors and Retinoid X Receptors: Interactions with Endogenous Retinoic Acids." *Proceedings of the National Academy of Sciences of the United States of America* Vol. 90, pp. 30-34
- Anon."p052_55_Ochsenbein.pdf."http://www.cea.fr/content/download/5423/35426/file/p0 52 55 Ochsenbein.pdf.
- Arany Z, Sellers W R, Livingston D M, and Eckner R. 1994. "E1A-associated P300 and CREB-associated CBP Belong to a Conserved Family of Coactivators." *Cell* (June): 77 (6) 799–800.
- Auger C J, Bentley G E, Auger A P, Ramamurthy M, and Ball G F. 2002. "Expression of cAMP Response Element Binding Protein-binding Protein in the Song Control System and Hypothalamus of Adult European Starlings (Sturnus Vulgaris)." *Journal* of Neuroendocrinology (October) 14 (10) : 805-13.
- Baba D, Maita N, Jee JG, Uchimura Y, Saitoh H, Sugasawa K, Hanaoka F, Tochio H, Hiroaki H and Shirakawa M. 2005. "Crystal Structure of Thymine DNA Glycosylase Conjugated to SUMO-1." *Nature* (June) 16; 435 (7044): 979-82.
- Bammler T, Beyer RP, Bhattacharya S, Boorman GA, Boyles A, Bradford BU, Bumgarner RE, Bushel PR, Chaturvedi K, Choi D, Cunningham ML, Deng S, Dressman HK, Fannin RD, Farin FM, Freedman JH, Fry RC, Harper A, Humble MC, Hurban P, Kavanagh TJ, Kaufmann WK, Kerr KF, Jing L, Lapidus JA, Lasarev MR, Li J, Li YJ, Lobenhofer EK, Lu X, Malek RL, Milton S, Nagalla SR, O'malley JP, Palmer VS, Pattee P, Paules RS, Perou CM, Phillips K, Qin LX, Qiu Y, Quigley SD, Rodland M, Rusyn I, Samson LD, Schwartz DA, Shi Y, Shin JL, Sieber SO, Slifer S, Speer MC, Spencer PS, Sproles DI, Swenberg JA, Suk WA, Sullivan RC, Tian R, Tennant RW, Todd SA, Tucker CJ, Van Houten B, Weis BK, Xuan S and Zarbl H. 2005. "Standardizing Global Gene Expression Analysis Between Laboratories and Across Platforms." *Nature Methods* (April): 2 (5): 351 356
- Bannister A J. and Kouzarides T. 2005. "Reversing Histone Methylation." *Nature* (July) 436 (7054): 1103–1106.
- Bartlett J, Blagojevic J, Carter D, Eskiw C, Fromaget M, Job C, Shamsher M, Trindade IF, Xu M and Cook PR. 2006. "Specialized Transcription Factories." *Biochemical Society Symposium* (73): 67–75.
- Bécavin C, Tchitchek N, Mintsa-Eya C, Lesne A, and Benecke A. 2011. "Improving the Efficiency of Multidimensional Scaling in the Analysis of High-dimensional Data Using Singular Value Decomposition." *Bioinformatics (Oxford, England)* (May) 27 (10): 1413–1421.
- Benecke, A. 2006. "Chromatin Code, Local Non-equilibrium Dynamics, and the

Emergence of Transcription Regulatory Programs." *The European Physical Journal. E, Soft Matter* (March) 19 (3): 353–366.

- Benecke, A. 2003. "Genomic Plasticity and Information Processing by Transcription Coregulators." *Complexus* 1 (2): 65–76.
- Bennett MT, Rodgers MT, Hebert AS, Ruslander LE, Eisele L, and Drohat AC. 2006. "Specificity of Human Thymine DNA Glycosylase Depends on N-Glycosidic Bond Stability." *Journal of the American Chemical Society* (September) 128 (38): 12510– 12519.
- Borrow J, Goddard AD, Sheer D and Solomon E. 1990. "Molecular Analysis of Acute Promyelocytic Leukemia Breakpoint Cluster Region on Chromosome 17." *Science* (*New York, N.Y.*) (September) 249 (4976): 1577–1580.
- Bour G, Lalevée S, and Rochette-Egly C. 2007. "Protein Kinases and the Proteasome Join in the Combinatorial Control of Transcription by Nuclear Retinoic Acid Receptors." *Trends in Cell Biology* (June) 17 (6): 302–309.
- Bridgham JT, Eick GN, Larroux C, Deshpande K, Harms MJ, Gauthier ME, Ortlund EA, Degnan BM, Thornton JW. 2010. "Protein Evolution by Molecular Tinkering: Diversification of the Nuclear Receptor Superfamily from a Ligand-dependent Ancestor." *PLoS Biology* 8 (10).
- http://www.ncbi.nlm.nih.gov/pubmed/20957188.
- Brysbaert G, Noth S, and Benecke A. 2007. "Generation of Synthetic Transcriptome Data with Defined Statistical Properties for the Development and Testing of New Analysis Methods." *Genomics, Proteomics & Bioinformatics / Beijing Genomics Institute* 5 (1) (February): 45–52.
- Brysbaert, Guillaume, François-Xavier Pellay, Sebastian Noth, and Arndt Benecke. 2010. "Quality Assessment of Transcriptome Data Using Intrinsic Statistical Properties." *Genomics, Proteomics & Bioinformatics / Beijing Genomics Institute* 8 (1) (March): 57–71.
- Chalkiadaki A, and Talianidis I. 2005. "SUMO-dependent Compartmentalization in Promyelocytic Leukemia Protein Nuclear Bodies Prevents the Access of LRH-1 to Chromatin." *Molecular and Cellular Biology* 25 (12) (June): 5095–5105.
- Chen D, Lucey MJ, Phoenix F, Lopez-Garcia J, Hart SM, Losson R, Buluwela L, Coombes RC, Chambon P, Schär P, Ali S.. 2003. "T:G Mismatch-specific thymine-DNA Glycosylase Potentiates Transcription of Estrogen-regulated Genes Through Direct Interaction with Estrogen Receptor Alpha." *The Journal of Biological Chemistry* 278 (40) (October 3): 38586–38592.
- Chen RZ, Pettersson U, Beard C, Jackson-Grusby L, Jaenisch R. 1998. "DNA Hypomethylation Leads to Elevated Mutation Rates." *Nature* 395 (6697) (September 3): 89–93.
- Chiurazzi P, Pomponi MG, Pietrobono R, Bakker CE, Neri G, Oostra BA. 1999. "Synergistic Effect of Histone Hyperacetylation and DNA Demethylation in the Reactivation of the FMR1 Gene." *Human Molecular Genetics* 8 (12) (November): 2317–2323.
- Choi I, Lee S, Kyoung Chung H, Suk Lee Y, Eui Kim K, Choi D, Park EK, Yang D,

Ecoiffier T, Monahan J, Chen W, Aguilar B, Lee HN, Yoo J, Koh CJ, Chen L, Wong AK, Hong YK. 2012. "9-cis Retinoic Acid Promotes Lymphangiogenesis and Enhances Lymphatic Vessel Regeneration: Therapeutic Implications of 9-cis Retinoic Acid for Secondary Lymphedema." *Circulation* 125 (7) (February 21): 872–882. *Chrivia JC, Kwok RP, Lamb N, Hagiwara M, Montminy MR, Goodman RH. 1993. "Phosphorylated CREB Binds Specifically to the Nuclear Protein CBP." *Nature* 365 (6449) (October 28): 855–859.

- Comings, D E. 1967. "Histones of Genetically Active and Inactive Chromatin." *The Journal of Cell Biology* 35 (3) (November): 699–708.
- Conticello S G. 2008. "The AID/APOBEC Family of Nucleic Acid Mutators." *Genome Biology* 9 (6): 229.
- Cortázar D, Kunz C, Saito Y, Steinacher R and Schär P. 2007. "The Enigmatic Thymine DNA Glycosylase." *DNA Repair* 6 (4): 489–504.
- Cortázar D, Kunz C, Selfridge J, Teresa Lettieri, Saito Y, MacDougall E, Wirz A, Schuermann D, Jacobs A L, Siegrist F, Steinacher R, Jiricny J, Bird A and Schär P. 2011. "Embryonic Lethal Phenotype Reveals a Function of TDG in Maintaining Epigenetic Stability." *Nature* 470 (7334) (February 17): 419–423.
- Coupry I, Roudaut C, Stef M, Delrue MA, Marche M, Burgelin I, Taine L, Cruaud C, Lacombe D, and Arveiler B. 2002. "Molecular Analysis of the CBP Gene in 60 Patients with Rubinstein-Taybi Syndrome." *Journal of Medical Genetics* 39 (6) (June): 415–421.
- Doucas V, Tini M, Egan DA, and Evans RM. 1999. "Modulation of CREB Binding Protein Function by the Promyelocytic (PML) Oncoprotein Suggests a Role for Nuclear Bodies in Hormone Signaling." *Proceedings of the National Academy of Sciences* 96 (6) (March 16): 2627–2632.
- Drapkin R, Reardon J T, Ansari A, Huang J-C, Zawel L, Ahn K, Sancar A, and Reinberg D. 1994. "Dual Role of TFIIH in DNA Excision Repair and in Transcription by RNA Polymerase II.", *Published Online: 21 April 1994;* | *Doi:10.1038/368769a0* 368 (6473) (April 21): 769–772.
- Eftedal, I, Guddal P H, Slupphaug G, Volden G, and Krokan H E. 1993. "Consensus Sequences for Good and Poor Removal of Uracil from Double Stranded DNA by uracil-DNA Glycosylase." *Nucleic Acids Research* 21 (9) (May 11): 2095–2101.
- Eilebrecht S, Smet-Nocca C, Wieruszeski J-M, and Benecke A. 2010. "SUMO-1 Possesses DNA Binding Activity." *BMC Research Notes* 3 (1): 146.
- Esteller, M. 2006. "CpG Island Methylation and Histone Modifications: Biology and Clinical Significance." *Ernst Schering Research Foundation Workshop* (57): 115– 126.
- Fodor S P, Read J-L, Pirrung M C, Stryer L, Lu A T, and Solas D. 1991. "Light-directed, Spatially Addressable Parallel Chemical Synthesis." *Science (New York, N.Y.)* 251 (4995) (February 15): 767–773.
- Fukushige, S, Kondo E, and Horii A. 2009. "Methyl-CpG Targeted Recruitment of P300 Reactivates Tumor Suppressor Genes in Human Cancer Cells." *Biochemical and Biophysical Research Communications* 379 (4) (January 20): 1021–1026.

- Gallinari, P, and Jiricny J. 1996. "A New Class of uracil-DNA Glycosylases Related to Human thymine-DNA Glycosylase." *Nature* 383 (6602) (October 24): 735–738.
- Germain P, Chambon P Eichele G, Evans R M, Lazar M A, Leid M, De Lera A R, Lotan R, Mangelsdorf D J, and Gronemeyer H. 2006. "International Union of Pharmacology. LX. Retinoic Acid Receptors." *Pharmacological Reviews* 58 (4) (December 1): 712–725.
- Gilbert N, and Bickmore W A. 2006. "The Relationship Between Higher-order Chromatin Structure and Transcription." *Biochemical Society Symposium* (73): 59–66.
- Glass C K and Rosenfeld M G. 2000. "The Coregulator Exchange in Transcriptional Functions of Nuclear Receptors." *Genes & Development* 14 (2) (January 15): 121–141.
- Goodman R H, and Smolik S. 2000. "CBP/p300 in Cell Growth, Transformation, and Development." *Genes & Development* 14 (13) (July 1): 1553–1577.
- Gu W, and Roeder R G. 1997. "Activation of P53 Sequence-specific DNA Binding by Acetylation of the P53 C-terminal Domain." *Cell* 90 (4) (August 22): 595–606.
- Hanahan D, and Weinberg R A. 2000. "The Hallmarks of Cancer." *Cell* 100 (1) (7): 57–70.
- Hanawalt P C, and Spivak G. 2008. "Transcription-coupled DNA Repair: Two Decades of Progress and Surprises." *Nature Reviews. Molecular Cell Biology* 9 (12) (December): 958–970.
- Hardeland U, Bentele M, Jiricny J, and Schär P. 2000. "Separating Substrate Recognition from Base Hydrolysis in Human Thymine DNA Glycosylase by Mutational Analysis." *Journal of Biological Chemistry* 275 (43) (October 27): 33449 –33456.
- Hardeland U, Bentele M, Jiricny J, and Schär P. 2003. "The Versatile Thymine DNAglycosylase: a Comparative Characterization of the Human, Drosophila and Fission Yeast Orthologs." *Nucleic Acids Research* 31 (9) (May 1): 2261–2271.
- Hardeland U, Steinacher R, Jiricny J, and Schär P. 2002. "Modification of the Human thymine-DNA Glycosylase by Ubiquitin-like Proteins Facilitates Enzymatic Turnover." *The EMBO Journal* 21 (6) (March 15): 1456–1464.
- Harp J M., Leif Hanson B, Timm D E, and Bunick G J. 2000. "Asymmetries in the Nucleosome Core Particle at 2.5 Å Resolution." Acta Crystallographica Section D Biological Crystallography 56 (12) (November 1): 1513–1534.
- He Z, Wu L, Fields MW, and Zhou J.. 2005. "Use of Microarrays with Different Probe Sizes for Monitoring Gene Expression." *Applied and Environmental Microbiology* 71 (9) (August): 5154–5162.
- Hennekam R C. 2006. "Rubinstein-Taybi syndrome." *European Journal of Human Genetics* 14 (9) (July 26): 981–985.
- Huang T T, and D'Andrea A D. 2006. "Regulation of DNA Repair by Ubiquitylation." *Nature Reviews. Molecular Cell Biology* 7 (5) (May): 323–334.
- Huang Y-S, Chang C-C, Huang T-C, Hsieh Y-L, and Shih H-M. 2012. "Daxx Interacts with and Modulates the Activity of CREB." *Cell Cycle (Georgetown, Tex.)* 11 (1) (January 1): 99–108.
- Iyer NG, Ozdag H, and Caldas C. 2004. "p300/CBP and Cancer." Oncogene 23 (24) (May

24): 4225-4231.

- Kamitani T, Nguyen H P, Kito K, Fukuda-Kamitani T, and Yeh E T H. 1998. "Covalent Modification of PML by the Sentrin Family of Ubiquitin-like Proteins." *Journal of Biological Chemistry* 273 (6) (February 6): 3117–3120.
- Kane M D, Jatkoe T A, Stumpf C R, Lu J, Thomas J D, and Madore S J. 2000. "Assessment of the Sensitivity and Specificity of Oligonucleotide (50mer) Microarrays." *Nucleic Acids Research* 28 (22) (October 15): 4552–4557.
- Kohli RM, Maul RW, Guminski AF, McClure RL, Gajula KS, Saribasak H, McMahon MA, Siliciano RF, Gearhart PJ, and Stivers JT. 2010. "Local Sequence Targeting in the AID/APOBEC Family Differentially Impacts Retroviral Restriction and Antibody Diversification." *The Journal of Biological Chemistry* 285 (52) (December 24): 40956–40964.
- Kubicek S, and Jenuwein T. 2004. "A Crack in Histone Lysine Methylation." *Cell* 119 (7) (November 29): 903–906.
- Kunz C, Focke F, Saito Y, Schuermann D, Lettieri T, Selfridge J, and Schär P. 2009. "Base Excision by Thymine DNA Glycosylase Mediates DNA-Directed Cytotoxicity of 5-Fluorouracil." Ed. James E. Haber. *PLoS Biology* 7 (4):
- Kwok, R P, Lundblad J R, Chrivia J C, Richards J P, Bächinger H P, Brennan R G, Roberts S G, Green M R, and Goodman R H. 1994. "Nuclear Protein CBP Is a Coactivator for the Transcription Factor CREB." *Nature* 370 (6486) (July 21): 223– 226.
- Lavau, C, Du C, Thirman M, and Zeleznik-Le N. 2000. "Chromatin-related Properties of CBP Fused to MLL Generate a Myelodysplastic-like Syndrome That Evolves into Myeloid Leukemia." *The EMBO Journal* 19 (17) (September 1): 4655–4664.
- Lavelle C, and Benecke A. 2006. "Chromatin Physics: Replacing Multiple, Representation-centered Descriptions at Discrete Scales by a Continuous, Functiondependent Self-scaled Model." *The European Physical Journal. E, Soft Matter* 19 (3) (March): 379–384.
- Le X-F, Yang P, and Chang K-S. 1996. "Analysis of the Growth and Transformation Suppressor Domains of Promyelocytic Leukemia Gene, PML." *Journal of Biological Chemistry* 271 (1) (January 5): 130–135.
- Lesne, A. 2006. "The Chromatin Regulatory Code: Beyond a Histone Code." *The European Physical Journal. E, Soft Matter* 19 (3) (March): 375–377.
- Lin R J, Sternsdorf T, Tini M, and Evans R M. 2001. "Transcriptional Regulation in Acute Promyelocytic Leukemia." *Oncogene* 20 (49) (October 29): 7204–7215.
- Lindahl T, Ljungquist S, Siegert W, Nyberg B, and Sperens B. 1977. "DNA Nglycosidases: Properties of uracil-DNA Glycosidase from Escherichia Coli." *The Journal of Biological Chemistry* 252 (10) (May 25): 3286–3294.
- Linney E, Donerly S, Mackey L, and Dobbs-McAuliffe B. 2011. "The Negative Side of Retinoic Acid Receptors." *Neurotoxicology and Teratology* 33 (6) (November): 631– 640.
- Liu X, Wang L, Zhao K, Thompson P R, Hwang Y, Marmorstein R, and Cole P A. 2008. "The Structural Basis of Protein Acetylation by the p300/CBP Transcriptional

Coactivator." Nature 451 (7180) (February 14): 846-850.

- Lucey MJ, Chen D, Lopez-Garcia J, Hart SM, Phoenix F, Al-Jehani R, Alao JP, White R, Kindle KB, Losson R, Chambon P, Parker MG, Schär P, Heery DM, Buluwela L, and Ali S. 2005. "T: G Mismatch-specific thymine-DNA Glycosylase (TDG) as a Coregulator of Transcription Interacts with SRC1 Family Members Through a Novel Tyrosine Repeat Motif." *Nucleic Acids Research* 33 (19): 6393–6404.
- Lund A H, and van Lohuizen M. 2004. "Epigenetics and Cancer." *Genes & Development* 18 (19) (September 1): 2315–2335.
- Maden M. 2007. "Retinoic Acid in the Development, Regeneration and Maintenance of the Nervous System." *Nature Reviews Neuroscience* 8 (10) (October 1): 755–765.
- Matsuzaki K, Minami T, Tojo M, Honda Y, Saitoh N, Nagahiro S, Saya H, and Nakao M. 2003. "PML-nuclear Bodies Are Involved in Cellular Serum Response." *Genes to Cells* 8 (3) (March 1): 275–286.
- Matunis M J, Coutavas E, and Blobel G. 1996. "A Novel Ubiquitin-like Modification Modulates the Partitioning of the Ran-GTPase-activating Protein RanGAP1 Between the Cytosol and the Nuclear Pore Complex." *The Journal of Cell Biology* 135 (6) (December 15): 1457–1470.
- Mohan R D, Rao A Gagliardi J, and Tini M. 2007. "SUMO-1-dependent Allosteric Regulation of Thymine DNA Glycosylase Alters Subnuclear Localization and CBP/p300 Recruitment." *Molecular and Cellular Biology* 27 (1) (January): 229–243.
- Morikawa T, Kuchiba A, Qian ZR, Mino-Kenudson M, Hornick JL, Yamauchi M, Imamura Y, Liao X, Nishihara R, Meyerhardt JA, Fuchs CS, and Ogino S. 2011. "Prognostic Significance and Molecular Associations of Tumor Growth Pattern in Colorectal Cancer." *Annals of Surgical Oncology* (December 22).
- Mund C, Musch T, Strödicke M, Assmann B, Li E, and Lyko F. 2004. "Comparative Analysis of DNA Methylation Patterns in Transgenic Drosophila Overexpressing Mouse DNA Methyltransferases." *The Biochemical Journal* 378 (Pt 3) (March 15): 763–768.
- Naegeli H, and Sugasawa K. 2011. "The Xeroderma Pigmentosum Pathway: Decision Tree Analysis of DNA Quality." *DNA Repair* 10 (7) (July 15): 673–683.
- Neddermann P and Jiricny J. 1993. "The Purification of a Mismatch-specific thymine-DNA Glycosylase from HeLa Cells." *Journal of Biological Chemistry* 268 (28) (October 5): 21218–21224.
- Neddermann P, Gallinari P, Lettieri T, Schmid D, Truong O, Hsuan JJ, Wiebauer K, and Jiricny J. 1996. "Cloning and Expression of Human G/T Mismatch-specific Thymine-DNA Glycosylase." *Journal of Biological Chemistry* 271 (22) (May 31): 12767–12774.
- Noth S, and Benecke A. 2005. "Avoiding Inconsistencies over Time and Tracking Difficulties in Applied Biosystems AB1700/Panther Probe-to-gene Annotations." *BMC Bioinformatics* 6: 307.
- Noth S, Brysbaert G, and Benecke A. 2006. "Normalization Using Weighted Negative Second Order Exponential Error Functions (NeONORM) Provides Robustness Against Asymmetries in Comparative Transcriptome Profiles and Avoids False

Calls." *Genomics, Proteomics & Bioinformatics / Beijing Genomics Institute* 4 (2) (May): 90–109.

- Park J, Seo T, Kim H, and Choe J. 2005. "Sumoylation of the Novel Protein hRIP {beta} Is Involved in Replication Protein A Deposition in PML Nuclear Bodies." *Molecular and Cellular Biology* 25 (18) (September): 8202–8214.
- Pease A C, Solas D, Sullivan E J, Cronin M T, Holmes C P and Fodor S P. 1994. "Lightgenerated Oligonucleotide Arrays for Rapid DNA Sequence Analysis." *Proceedings* of the National Academy of Sciences of the United States of America 91 (11) (April 24): 5022–5026.
- Puttagunta R, and Di Giovanni S. 2011. "Retinoic Acid Signaling in Axonal Regeneration." *Frontiers in Molecular Neuroscience* 4: 59.
- Reitzel, A M, Pang K, Ryan J F, Mullikin J C, Martindale M Q, Baxevanis A D, and Tarrant A M. 2011. "Nuclear Receptors from the Ctenophore Mnemiopsis Leidyi Lack a Zinc-finger DNA-binding Domain: Lineage-specific Loss or Ancestral Condition in the Emergence of the Nuclear Receptor Superfamily?" *EvoDevo* 2 (1): 3.
- Rhinn M, and Dollé P. 2012. "Retinoic Acid Signalling During Development." *Development* 139 (5) (March 1): 843–858.
- Riedl T, Hanaoka F, and Egly J-M. 2003. "The Comings and Goings of Nucleotide Excision Repair Factors on Damaged DNA." *The EMBO Journal* 22 (19) (October 1): 5293–5303.
- Roelfsema JH, and Peters DJ.. 2007. "Rubinstein–Taybi Syndrome: Clinical and Molecular Overview." *Expert Reviews in Molecular Medicine* 9 (23): 1–16.
- Rubinstein J H, and Taybi H. 1963. "Broad Thumbs and Toes and Facial Abnormalities. A Possible Mental Retardation Syndrome." *American Journal of Diseases of Children* (1960) 105 (June): 588–608.
- Schena M, Shalon D, Davis RW, and Brown PO. 1995. "Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray." *Science (New York, N.Y.)* 270 (5235) (September 20): 467–470.
- Schumacher B. 2009. "Transcription-blocking DNA Damage in Aging: a Mechanism for Hormesis." *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 31 (12) (December): 1347–1356.
- Segers A, Muyldermans S, and Wyns L. 1991. "The Interaction of Histone H5 and Its Globular Domain with Core Particles, Depleted Chromatosomes, Polynucleosomes, and a DNA Decamer." *The Journal of Biological Chemistry* 266 (3) (25): 1502– 1508.
- Simpson, R T. 1978. "Structure of the Chromatosome, a Chromatin Particle Containing 160 Base Pairs of DNA and All the Histones." *Biochemistry* 17 (25) (November 12): 5524–5531.
- Slack A, Cervoni N, Pinard M, and Szyf M. 1999. "Feedback Regulation of DNA Methyltransferase Gene Expression by Methylation." *European Journal of Biochemistry / FEBS* 264 (1) (August): 191–199.
- Sluder A E, and Maina C V. 2001. "Nuclear Receptors in Nematodes: Themes and

Variations." Trends in Genetics: TIG 17 (4) (April): 206-213.

- Smet-Nocca C, Wieruszeski J-M, Chaar V, Leroy A, and Benecke A. 2008. "The thymine-DNA Glycosylase Regulatory Domain: Residual Structure and DNA Binding." *Biochemistry* 47 (25) (June 24): 6519–6530.
- Smet-Nocca C, Wieruszeski J-M, Léger H, Eilebrecht S, and Benecke A. 2011. "SUMO-1 Regulates the Conformational Dynamics of thymine-DNA Glycosylase Regulatory Domain and Competes with Its DNA Binding Activity." *BMC Biochemistry* 12: 4.
- Spriet C. 2006. Instrumentation biophotonique pour la mesure d'interactions moléculaires dans la cellule.
- Stengele K-P, Bühler J, Bühler S, Kvassiouk E, Green R, Prykota T, and Pfleiderer W. 2005. "Recent Highlights on Photolithic Oligonucleotide Array in Situ Synthesis." *Nucleosides, Nucleotides & Nucleic Acids* 24 (5-7): 891–896.
- Su B, Chen X, Zhong C, Guo N, He J, and Fan Y.2012. "All-trans Retinoic Acid Inhibits Mesangial Cell Proliferation by Up-regulating p21Waf1/Cip1 and p27Kip1 and Down-regulating Skp2." *Journal of Nephrology* (February 15): 0.
- Svejstrup J Q. 2002. "Mechanisms of Transcription-coupled DNA Repair." *Nature Reviews. Molecular Cell Biology* 3 (1) (January): 21–29.
- Takahashi H, Hatakeyama S, Saitoh H, and Nakayama KI. 2005. "Noncovalent SUMO-1 Binding Activity of Thymine DNA Glycosylase (TDG) Is Required for Its SUMO-1 Modification and Colocalization with the Promyelocytic Leukemia Protein." *The Journal of Biological Chemistry* 280 (7) (February 18): 5611–5621.
- Taybi, H, and Rubinstein J H. 1965. "Broad Thumbs and Toes, and Unusual Facial Features; A Probable Mental Retardation Syndrome." *The American Journal of Roentgenology, Radium Therapy, and Nuclear Medicine* 93 (February): 362–366.
- Tini M, Benecke A, Um S-J, Torchia J, Evans R M, and Chambon P. 2002. "Association of CBP/p300 Acetylase and Thymine DNA Glycosylase Links DNA Repair and Transcription." *Molecular Cell* 9 (2) (February): 265–277.
- Todaro LB, Veloso MJ, Campodónico PB, Puricelli LI, Farías EF, and Bal de Kier Joffé ED. 2012. "A Clinically Relevant Bi-cellular Murine Mammary Tumor Model as a Useful Tool for Evaluating the Effect of Retinoic Acid Signaling on Tumor Progression." *Breast Cancer (Tokyo, Japan)* (February 29).
- Um S, Harbers M, Benecke A, Pierrat B, Losson R, and Chambon P. 1998. "Retinoic Acid Receptors Interact Physically and Functionally with the T:G Mismatch-specific thymine-DNA Glycosylase." *The Journal of Biological Chemistry* 273 (33) (August 14): 20728–20736.
- Varghese, A. J. 1971. "Photochemical Reactions of Cytosine Nucleosides in Frozen Aqueous Solution and in Deoxyribonucleic Acid." *Biochemistry* 10 (12): 2194–2199.
- Vendel A C, and Lumb K J. 2003. "Molecular Recognition of the Human Coactivator CBP by the HIV-1 Transcriptional Activator Tat." *Biochemistry* 42 (4) (February 4): 910– 916.
- Vendel A C, and Lumb K J. 2004. "NMR Mapping of the HIV-1 Tat Interaction Surface of the KIX Domain of the Human Coactivator CBP." *Biochemistry* 43 (4) (February 3): 904

- Vértessy B G, and Tóth J. 2009. "Keeping Uracil Out of DNA: Physiological Role, Structure and Catalytic Mechanism of dUTPases." Accounts of Chemical Research 42 (1) (January 20): 97–106.
- Walsh C P, and Xu G L. 2006. "Cytosine Methylation and DNA Repair." *Current Topics in Microbiology and Immunology* 301: 283–315.
- Wiebauer K, and Jiricny J. 1989. "In Vitro Correction of G o T Mispairs to G o C Pairs in Nuclear Extracts from Human Cells." Nature. May 18;339(6221): 234-6.
- Winkley, M W, and Robins R K. 1970. "Direct Glycosylation of 1,3,5-triazinones. A New Approach to the Synthesis of the Nucleoside Antibiotic 5-azacytidine (4-amino-1beta-D-ribofuranosyl-1,3,5-triazin-2-one) and Related Derivatives." *The Journal of Organic Chemistry* 35 (2) (February): 491–495.
- Wydner K L, Bhattacharya S, Eckner R, Lawrence J B, and Livingston D M. 1995."Localization of Human CREB-binding Protein Gene (CREBBP) to 16p13.2-p13.3 by Fluorescence in Situ Hybridization." *Genomics* 30 (2) (November 20): 395–396.
- Zhang Z, Burch PE, Cooney AJ, Lanz RB, Pereira FA, Wu J, Gibbs RA, Weinstock G, and Wheeler DA. 2004. "Genomic Analysis of the Nuclear Receptor Family: New Insights into Structure, Regulation, and Evolution from the Rat Genome." *Genome Research* 14 (4) (April): 580–590.
- Zhong S, Delva L, Rachez C, Cenciarelli C, Gandini D, Zhang H, Kalantry S, Freedman LP, and Pandolfi PP. 1999. "A RA-dependent, Tumour-growth Suppressive Transcription Complex Is the Target of the PML-RARalpha and T18 Oncoproteins." *Nature Genetics* 23 (3) (November): 287–295.
- Zhou Q, Agoston AT, Atadja P, Nelson WG, and Davidson NE.2008. "Inhibition of Histone Deacetylases Promotes Ubiquitin-dependent Proteasomal Degradation of DNA Methyltransferase 1 in Human Breast Cancer Cells." *Molecular Cancer Research: MCR* 6 (5) (May): 873–883.
- Zhou X, Meeker AK, Makambi KH, Kosti O, Kallakury BV, Sidawy MK, Loffredo CA, and Zheng YL. 2012. "Telomere Length Variation in Normal Epithelial Cells Adjacent to Tumor: Potential Biomarker for Breast Cancer Local Recurrence." *Carcinogenesis* 33 (1): 113–118.