

**THESE DE DOCTORAT DES UNIVERSITÉS BORDEAUX I ET BORDEAUX
SEGALEN**

Ecole Doctorale Sciences de la Vie et de la Santé

Spécialité : Biochimie

Présentée par

Daimel CASTILLO GONZÁLEZ

Pour obtenir le grade de Docteur de l'Université de Bordeaux

Nouveaux ligands de quadruplexes. Approches *in silico* et *in vitro*.

Novel G-quadruplex ligands. *In silico* and *in vitro* approaches.

Directeur de Thèse / Thesis supervisor : Dr. Jean-Louis MERGNY

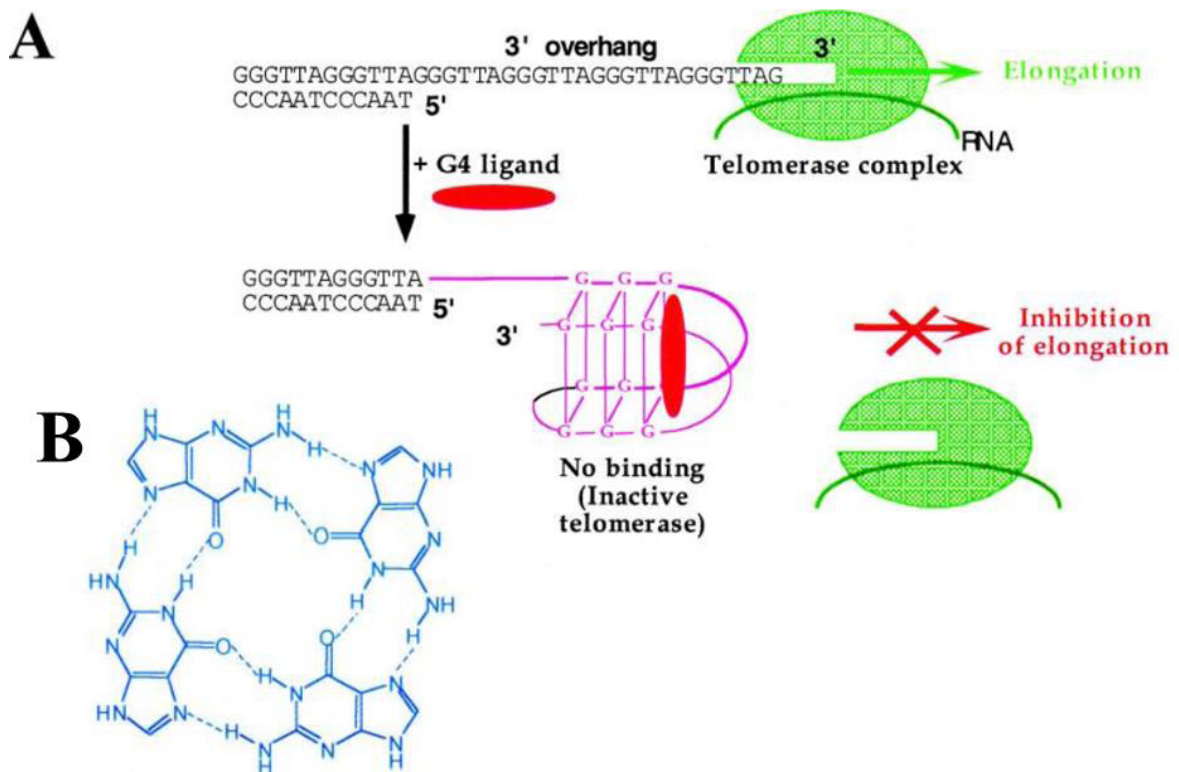
Thèse soutenue le 14 Novembre 2013 à l'IECB, Pessac

Thesis Defense on November 14, 2013 at IECB, Pessac

Résumé étendu en français

Introduction et Objectifs.

Les séquences d'ADN et d'ARN riches en G peuvent adopter des conformations inhabituelles bien différentes de la double-hélice classique. Ces structures sont connues sous le nom de G-quadruplexes ou G4. Les topologies et le repliement de ces structures fascinantes sont très diverses. Les G4 sont stabilisés par la présence de cations monovalents et la formation de 8 liaisons Hydrogène de type Hoogsteen par quartet. Ces structures peuvent être reconnues par de petites molécules, qui stabilisent ces édifices, le plus souvent par des interactions d'empilement $\pi - \pi$. Bien que les G4 soient connus depuis des décennies, l'intérêt pour ce domaine a pris un nouvel essor à la fin des années 90, lorsqu'il a été démontré un effet inhibiteur sur l'activité de la télomérase, une transcriptase inverse impliquée dans l'immortalisation de la plupart des cellules cancéreuses.



(A) Mécanisme d'inhibition de la télomérase par un G-quadruplexe (B) Structure d'une tetrade (ou quartet) de G dans un quadruplex. Image extraite de Riou et al, PNAS, 2002.

En ce qui concerne la télomérase, le cancer et G4, plusieurs groupes ont été impliqués dans la découverte de nouveaux ligands de G4 qui inhiberaient indirectement l'enzyme. La plupart des ligands de quadruplexes ont été identifiés en relation avec cette cible. Des centaines de composés ont été caractérisés au cours de la dernière décennie. En prenant en compte les avantages et la facilité qu'offre l'identification de nouvelles structures à l'aide de techniques de calcul, nous avons construit des modèles mathématiques simples et reproductibles possédant une grande capacité de criblage pour un faible coût de calcul afin de pouvoir identifier de nouveaux ligands G4.

Dans ce travail nous proposons les objectifs spécifiques suivants:

1. Développer et valider expérimentalement de nouveaux modèles de calcul qui rendront possibles le criblage *in silico* d'un grand nombre de composés.
2. Identifier de nouveaux ligands des G-quadruplex.
3. Valider expérimentalement ces prédictions.

Organisation du manuscrit :

Nous allons décrire brièvement la structure de cette thèse.





- Le premier chapitre est une introduction qui présente les télomères, la télomérase et les quadruplexes. L'objectif principal de ce chapitre est de fournir des informations de fond liées à ces structures. La structure et les fonctions des télomères et de la télomérase ainsi que les conditions de formation des quadruplexes de guanine sont discutées. Nous présentons également les relations entre les télomères et le cancer, et comment les premiers ligands développés ont été conçus pour obtenir des propriétés anticancéreuses. D'autres régions du génome non télomériques ayant la possibilité de former G4 sont présentés.
- Le deuxième Chapitre est consacré à la présentation de concepts et paramètres liés aux méthodes QSAR. Ils sont cruciaux pour une meilleure compréhension de la thèse. Les détails concernant la technique employée pour transformer une molécule en un "nombre utile", et comment ce nombre peut être utilisé dans la construction de modèles mathématiques pour prédire une activité. Nous ajoutons des explications sur certains outils statistiques qui permettent de vérifier la qualité des modèles développés.

- Comme l'un des objectifs de cette recherche est l'identification et la prédiction de nouveaux ligands G4 avec nos méthodes de calcul, le chapitre 3 présente une revue de synthèse qui récapitule les stratégies de calcul qui ont été utilisées à cet effet. Une brève description des méthodes de calcul et les ligands identifiés à l'issue de ce récapitulatif est donné. Enfin, certaines caractéristiques importantes des ligands provenant d'études SAR sont discutées.
- Le chapitre 4 est composé de deux articles qui traitent des prédictions sur l'activité de l'enzyme télomérase par le mécanisme de stabilisation G4. Les premiers essais pour vérifier la possibilité d'obtenir des prédictions fiables à l'aide d'une méthodologie QSAR, basée à la fois sur une régression et des méthodes de classification, ont été réalisés. Nous avons ici conçu des modèles appliqués à une famille de composés seulement, des acridines (base de données congénères de composés) qui utilisent des techniques simples mais robuste (LDA et MLR). Cette approche permet de reproduire les valeurs trouvées dans les tests expérimentaux. Dans le même temps, il est possible de relier nos descripteurs moléculaires avec certaines caractéristiques d'activité. En d'autres termes, en utilisant des techniques de calcul très simples, nous pouvons fournir des outils prédictifs. Ils peuvent être utilisés afin d'obtenir de nouvelles informations concernant la SAR à cette famille de composés (interprétation), et nous pouvons reproduire, à l'aide des descripteurs moléculaires, les observations expérimentales rapportées précédemment.
- Dans le chapitre 5, nous avons étendu l'analyse à un ensemble structurellement varié. Nous utilisons également la LDA, mais dans un ensemble de données non congénères de composés. Nous mettons en place une base de données d'inhibiteurs de la télomérase, où les ligands rapportés ont été décrits dans la littérature scientifique comme stabilisateurs de G4. Plusieurs modèles LDA et une stratégie de consensus pour la prédiction des futurs candidats sont proposés. Nous effectuons un criblage virtuel sur une base de molécules comprenant tous les composés approuvés par la FDA. Ce dépistage permet d'identifier de nouveaux ligands G4 avec une efficacité acceptable. De tous les ligands testés choisis pour l'évaluation expérimentale, nous confirmons l'affinité pour les G4 pour six d'entre eux.
- Dans le chapitre 6, nous avons décidé d'appliquer les meilleures pratiques pour l'élaboration de modèles QSAR avec des méthodes basées sur la structure (accueil) dans une stratégie de criblage virtuel intégré non linéaire. Nous utilisons une stratégie non - linéaire pour la modélisation des données. Tout d'abord, nous

“corrigeons” la base de données, avec l'objectif d'éliminer les composés nocifs pour la procédure de modélisation, suivant les recommandations de Tropsha. Par la suite, de nombreux modèles de QSAR sont développés. Nous effectuons un criblage virtuel sur deux bases de données commerciales de composés, suivi par une campagne de “docking” afin de réduire le nombre de candidats. Les ligands sont expérimentalement testés comme stabilisateurs de la séquence télomérique humaine ainsi que d'autres séquences oncogènes et deux d'entre eux présentent des valeurs de stabilisation intéressantes. Le résultat final est l'identification de nouvelles familles, avec des valeurs de stabilisation plus élevées que les molécules décrites dans le chapitre précédent

- Le chapitre 7 est le début d'un travail inachevé. Ce chapitre ouvre des perspectives pour de futures études. Ce chapitre essaie de présenter les propriétés de stabilisation des composés identifiés dans le chapitre 6 dans des séquences considérées comme « caractéristique du cancer ». Ayant obtenu des résultats positifs sur toutes les séquences testées, les prochaines étapes seront d'appliquer ces composés sur des cellules en culture dans le but ultime de découvrir de nouveaux agents antiprolifératifs.

Les travaux effectués lors de ce doctorat ont donné lieu à 4 publications:

-  **FDA-approved drugs selected using virtual screening bind specifically to G-quadruplex DNA.** *Curr Pharm Des*, **19(12)**, 2164-2173. Castillo-Gonzalez, D., Perez-Machado, G., Guedin, A., Mergny, J. L., & Cabrera-Perez, M. A. (2013).
-  **Computational tools in the discovery of new G-quadruplex ligands with potential anticancer activity.** *Curr Top Med Chem*, **12(24)**, 2843-2856. Castillo-Gonzalez, D., Perez-Machado, G., Pallardo, F., Garrigues-Pelufo, T. M., & Cabrera-Perez, M. A. (2012).
-  **Prediction of telomerase inhibitory activity for acridinic derivatives based on chemical structure.** *Eur J Med Chem*, **44(12)**, 4826-4840. Castillo-Gonzalez, D., Cabrera-Perez, M. A., Perez-Gonzalez, M., Morales Helguera, A., & Duran-Martinez, A. (2009).
-  **Telomerase Inhibitory Activity of Acridinic Derivatives: A 3D-QSAR Approach.** *QSAR & Combinatorial Science*, **28(5)**, 526-536. Cabrera-Pérez, M. A., Castillo-González, D., Pérez-González, M., & Durán-Martínez, A. (2009).

Un cinquième article, en rapport avec les chapitres 6 et 7, est en cours de rédaction.

Conclusion :

Avec l'utilisation de la modélisation QSAR, nous pouvons prédire l'IC₅₀ sur la télomérase d'un ensemble de composés congénères. Nous avons également été en mesure de relier les descripteurs moléculaires qui apparaissent dans nos modèles avec certaines caractéristiques structurales préalablement décrites. En outre, nous avons construit des modèles différents utilisant des ensembles non congénères de composés en appliquant une stratégie de consensus et pu identifier six ligands approuvés par la FDA qui stabilisent ces G4. Par la suite, en appliquant des techniques non linéaires et un processus que nous avons proposé pour le traitement des bases de données, nous avons effectué un criblage virtuel de plus de 500 000 ligands à partir d'une base de données commerciale de composés. Nous avons ensuite appliqué un modèle basé sur la structure afin de réduire le nombre de candidats. Nous avons pu identifier de nouveaux ligands avec un potentiel inhibiteur plus importants que les précédents, et qui peuvent également stabiliser d'autres structures G4 impliqués dans d'autres processus liés au cancer. Ces observations ouvrent un spectre large de possibilités à explorer : malgré les limites des techniques de modélisation QSAR explorées tout au long de ce travail, nous considérons qu'elles peuvent être combinées et utilisées avec soin pour répondre à la recherche de nouveaux stabilisateurs G4.