

CONTEXTE

Les dernières années ont vu une accumulation très importante des données bioinformatiques sur l'ARN. De nombreux logiciels et banques de données sont développés régulièrement pour permettre une exploitation efficace de l'ensemble de ces informations. Malheureusement, ces différents outils reposent généralement sur des formats de stockage et de diffusion hétérogènes, rendant ainsi les données stockées difficile à recouper entre elles. La nécessité d'intégration et d'uniformisation des informations est un problème crucial en bioinformatique, et plus particulièrement dans le cas de l'ARN. A la différence des protéines et de par ses caractéristiques intrinsèques, il n'est pas possible d'émettre facilement des hypothèses sur la structure d'un ARN, et donc sur son rôle, en ne possédant que sa séquence pour seule information. Pour les mêmes raisons, en se limitant à la séquence d'un ARN, il est également très difficile d'identifier de nouveaux gènes codant pour ce type de molécule. En conséquence, dans le cas de l'ARN plus qu'ailleurs, il est nécessaire de pouvoir visualiser et manipuler de manière concertée l'ensemble des données d'un ARN provenant des différentes voies d'exploration.

OBJECTIFS

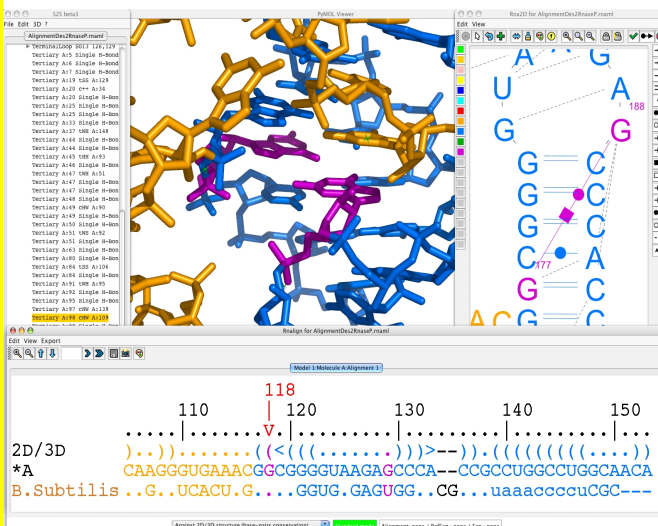
Nous souhaitons mettre en place un environnement permettant d'assouplir l'analyse des molécules d'ARN de la séquence à la structure 3D. Ce projet comprendra dans un premier temps le développement d'interfaces graphiques conviviales permettant à un utilisateur de manipuler de manière interconnectée des données de séquence (une séquence ou bien un alignement de séquences), de structure secondaire (incorporant la connaissance des différents types d'interactions mises en jeu entre les bases) et de structure 3D. Dans une deuxième partie, nous souhaitons associer ces modules aux outils bioinformatiques actuellement disponibles. Nous souhaitons mettre en place une modélisation des procédures d'analyses bioinformatiques disponibles dans le domaine de l'ARN. Cette modélisation devra permettre, via une interface adaptée, de construire et de tester de manière efficace et rapide des protocoles bioinformatiques visant l'analyse d'une molécule d'ARN. Les résultats produits à chaque étape pourront être visualisés et manipulés au moyen des modules développés dans la première étape.

RESULTATS

Nous avons mis en place l'infrastructure S2S (pour « Sequence TO Structure ») permettant de manipuler de manière interconnectée les données de séquences, de structures secondaires et tertiaires de l'ARN. Il propose pour cela trois modules (et présents dans la capture ci-dessous) :

- Rnalign : un éditeur d'alignements de séquences
- Rna2DViewer : un éditeur de structures secondaires
- Rna3DViewer : un visualiseur 3D

S2S est développé en Java et a été testé sous les systèmes LINUX et Mac OSX.



Le module Rnalign permet de réaliser des alignements structuraux en proposant :

- des masques de structures permettant d'afficher, pour chaque séquence dans l'alignement, la conservation des paires de base par rapport à la séquence de référence
- un couplage avec les outils de visualisation 2D et 3D permettant de tenir compte des interactions tertiaires

EN PROJET

- utilisation de l'infrastructure S2S pour l'étude de motifs structuraux d'ARN
- création d'un module permettant la modélisation de structures d'ARN à partir d'un alignement structural. Cette étape nécessitera le couplage à la plateforme S2S des outils ESSA (développé dans le laboratoire de Christine Gaspin) et MANIP (développé dans le laboratoire d'Eric Westhof)
- couplage de la plateforme S2S à différents outils bioinformatiques d'analyse de l'ARN. Ce couplage se fera au moyen de services Web.

CALENDRIER

- Le 19 mai 2005 : release de la version beta 2 du logiciel S2S sur le portail bioinformatique <http://bioinformatics.org/S2S>
- Le 17 mai 2005 : la publication "Sequence to Structure (S2S) : display, manipulate and interconnect RNA data from sequence to structure" est acceptée dans le journal Bioinformatics [2]
- Le 8 mars 2005 : release de la version beta 1 du logiciel S2S sur le portail bioinformatique <http://bioinformatics.org/S2S>
- Le 1er mars 2005 : recrutement en CDD de 3 mois de Thomas Ludwig, titulaire d'un DEA d'Informatique à l'Université Louis Pasteur de Strasbourg. Il travaillera sur le portage et le couplage à la plateforme S2S des outils ESSA et MANIP
- Du 28 au 30 juin 2004 : JOBIM 2004, Montreal, Canada. Présentation orale intitulée : "S2S : un environnement d'annotation, de visualisation et de manipulation des données bioinformatiques de l'ARN »
- Octobre/novembre 2003 : démarrage du développement de S2S à partir du module RnaMLView-RNAVIEW [1]

LABORATOIRES IMPLIQUES

Laboratoire de Bioinformatique, modélisation et simulation des acides nucléiques, Architecture et Réactivité de L'ARN, CNRS - UPR 9002, Institut de Biologie Moléculaire et Cellulaire, Strasbourg

- Eric Westhof, Coordinateur (e.westhof@ibmc.u-strasbg.fr)
- Fabrice Jossinet (f.jossinet@ibmc.u-strasbg.fr)
- Thomas Ludwig (t.ludwig@ibmc.u-strasbg.fr)

Département de Biométrie et Intelligence Artificielle, INRA, Toulouse

- Christine Gaspin (gaspin@toulouse.inra.fr)

MOTS CLES

Bioinformatique structurale, ARN, alignement de séquences

PUBLICATIONS

- [1] Yang H, Jossinet F, Leontis N, Chen L, Westbrook J, Berman H, Westhof E. Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res.* 2003 Jul 1;31(13):3450-60
- [2] Jossinet F, Westhof E Sequence to Structure (S2S) : display, manipulate and interconnect RNA data from sequence to structure, *Bioinformatics, sous presse*