Université du Québec
Institut National de la Recherche Scientifique
Centre Armand-Frappier Santé et Biotechnologie

ÉVOLUTION DES PHÉNOMÈNES D'ÉCHANGES CONFORMATIONNELS ENTRE SOUS-FAMILLES ET AU SEIN D'UNE SOUS-FAMILLE DE RIBONUCLÉASES

Par

David Bernard

Thèse présentée pour l'obtention du grade de
Philosophiæ doctor (Ph.D.) en biologie

Jury d'évaluation

Président du jury et examinateur interne	Pr. David Chatenet INRS – Centre Armand-Frappier Santé et Biotechnologie
Examinateur externe	Pr. Xavier Roucou Département de biochimie Université de Sherbrooke
Examinateur externe	Dr. Yves Aubin Centre d'évaluation des produits biologiques Santé Canada
Directeur de recherche	Pr. Nicolas Doucet INRS – Centre Armand-Frappier Santé et Biotechnologie

REMERCIEMENTS

Je voudrais d'abord remercier mon directeur de thèse, Pr. Nicolas Doucet. Travailler dans son laboratoire m'a permis de développer mon indépendance en tant que chercheur, et surtout pour la première fois m'a fait me sentir comme un scientifique accompli plutôt que comme un simple étudiant. J'avais l'impression d'être à ma place et je me sentais utile. J'ai particulièrement apprécié les nombreuses opportunités qu'il m'a offertes pour collaborer sur les projets d'autres étudiants, ça a confirmé ma passion pour le travail de consultant. Finalement, Nicolas m'a grandement impressionné par ses valeurs comme chercheur, particulièrement son intégrité scientifique.

Je tiens ensuite à remercier mes collègues de laboratoire présents et passés, avec un remerciement spécial pour ceux avec lesquels j'ai travaillé sur le projet des RNases : notre merveilleuse assistante de recherche Myriam Létourneau, surnommée à juste titre « the saviour of the lab, » le travaillant et persévérant Donald Gagné Ph.D., qui a pavé la voie de mon projet, ainsi que Chitra Narayanan Ph.D., notre prolifique et souriante experte bioinformaticienne. Je remercie tous mes autres collègues, soient Maude Bourgault, Guillaume Brault Ph.D., Ying Lian Chew Fajardo, Giorgos Dalkas Ph.D., Carlos Eduardo Dulcey Jordan Ph.D., Philippe Egesborg, Donald Gagné Ph.D., Jacinthe Gagnon Ph.D., Myriam Laneville, Yossef Lopez de los Santos Ph.D., Aïssa Muamba, Chitra Narayanan Ph.D., Nhung Nguyen Thi Ph.D., Carolina Perusquía Hernández, Hang Pham, Félix Pigeon, Marie-Aude Pinoteau, Louise Roux, Andrés Mauricio Rueda Rueda Ph.D., Dominic Villemure et Carol-Anne Villeneuve pour les heures de discussions scientifiques ou non, les rires, et les cafés collectifs.

Je tiens également à remercier les chercheurs et étudiants qui ont pris un peu de leur précieux temps afin de me porter assistance sur mon projet de recherche, soient Pr. Pratul K. Agarwal, Khushboo Bafna, Purva Bhojane Ph.D., Pr. Chakra Chennubhotla, Pr. Éric Déziel, Om Choudhary Ph.D., Marie-Christine Groleau, Tim Hempel, Pr. Ewen Lescop, Tara Sprules Ph.D. et Sameer Al-Abdul Wahid Ph.D., sans qui ce projet n'aurait pas pu explorer si loin.

Parallèlement, j'aimerais remercier ceux qui m'ont à leur tour fait confiance pour les appuyer et qui m'ont permis de faire partie de leur propre projet de recherche, c'est-à-dire Ugo Dionne, Sarah Martinez et Pr. Rémy Ricoux. J'ai adoré pouvoir leur apporter mon aide.

Finalement, j'aimerais remercier mes parents Diane et Michel et ma sœur Marika, qui ont démontré une grande patience en attendant la fin de mes études.

RÉSUMÉ

Bien que le modèle clé-cadenas de fonctionnement des enzymes soit encore très prévalent de nos jours, la communauté scientifique reconnaît généralement les enzymes comme des macromolécules flexibles. Des mouvements allant de simples fluctuations browniennes de chaînes latérales à des réorganisations complètes du squelette polypeptidique sont impliqués directement ou indirectement dans la fonction de ces protéines. Toutefois, la compréhension du mécanisme derrière ces mouvements et de leur rôle dans la fonction enzymatique est généralement phénoménologique et restreinte aux mesures ayant été faites sur un nombre limité d'enzymes, et il demeure particulièrement complexe de prédire l'effet de mutations sur la fonction protéique en partie à cause du manque de compréhension sur ces mouvements intrinsèques. Le projet de recherche documenté dans la présente thèse a cherché à pallier à ces manquements via l'étude des changements de conformations au sein de nombreuses enzymes de la superfamille des ribonucléases A, et via la mise en parallèle de ces résultats avec l'analyse phylogénétique de ces enzymes. Nous présentons ici la caractérisation de la flexibilité conformationnelle de ces RNases, que nous avons effectuée au moyen de différentes expériences RMN et de simulations bioinformatiques. Les ensembles de RNases ainsi analysés englobent des enzymes phylogénétiquement éloignées et rapprochées, et nous présentons la toute première analyse quantitative des profils d'échange ainsi mesurés. Nous démontrons une divergence rapide de l'identité précise des résidus subissant des échanges conformationnels à l'échelle de la µs-ms ainsi que du taux auquel ces échanges se produisent. Cette divergence est présente à la fois entre RNases phylogénétiquement éloignées et rapprochées, et n'est pas nécessairement associée à une divergence des fonctions catalytique ou biologiques. Malgré tout, cette divergence dans les profils dynamiques des RNases n'est pas aléatoire, car une conservation de la compartimentalisation des résidus subissant des échanges est observée, et celle-ci est corrélée avec la conservation des fonctions biologiques. Les résultats de cette thèse démontrent que la pression évolutive est relativement permissive à l'intérieur de certaines limites quant à la conservation des changements de conformations chez les RNases, et cette permissivité pourrait être exploitable par l'ingénierie des protéines si les limites associées sont connues.

Mots-clés : conformations transitoires, ECP, enzymes, échanges conformationnels, évolution des protéines, paysages conformationnels, résonance magnétique nucléaire, ribonucléases

ABSTRACT

Even though the lock-and-key model for enzyme catalysis remains widespread nowadays, the scientific community usually recognizes enzymes as flexible macromolecules. Movements ranging from simple Brownian fluctuations of side chains to large-scale reorganizations of the protein backbone are directly or indirectly involved in protein function. However, the understanding of the mechanism behind these movements and of their role in enzyme function is usually phenomenological and limited to measurements that have been performed on a finite number of enzymes, and predicting mutagenesis consequences on protein function remains challenging partly because of this lack of understanding regarding these intrinsic movements. The research project documented in the current thesis tried to overcome these shortcomings through the study of conformational exchange within numerous enzymes of the ribonuclease A superfamily, and through the association of these results with the phylogenetic analysis of these enzymes. Here we present the characterization of RNase conformational flexibilities, which we performed using various NMR experiments and bioinformatics simulations. RNase ensembles thus analyzed encompass phylogenetically distant as well as close enzymes, and we present the first quantitative analysis of exchange profiles thus measured. We demonstrate a rapid divergence of exchanging residue identity and of exchange rates on the µs-ms timescale. This divergence is found between both phylogenetically distant and close RNases, and is not necessarily associated with a biological or catalytic function divergence. Nevertheless, this divergence in RNase dynamic profiles is not random, as a conservation of exchanging residue clustering is observed, and is correlated with biological function conservation. This thesis demonstrates that evolutionary pressure is relatively permissive within certain limits regarding the conservation of conformational exchanges in RNases, and this permissiveness could be used by protein engineering if the associated limits become known.

Keywords: conformational exchange, conformational landscapes, enzymes, nuclear magnetic resonance, protein evolution, transiant conformations

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES ÉQUATIONS

LISTE DES ABRÉVIATIONS

ADN / DNA	Acide désoxyribonucléique
AIDS	Syndrome d'immunodéficience acquise
ARN / RNA	Acide ribonucléique
AtR2 / AtEDN	Ribonucléase 2 / « Eosinophil-derived neurotoxin » d'*Aotus trivirgatus*
BMRB	« Biological Magnetic Resonance Bank »
BtRA	Ribonucléase A bovine
CEST	« Chemical exchange saturation transfer »
CPMG	Carr-Purcell-Meiboom-Gill
DEER	« Double electron-electron resonance »
DHFR	Dihydrofolate réductase
dNTP	Nucléotide 5'-triphosphate
ΔG	Énergie libre de Gibbs
Δω	Variation de déplacement chimique entre les conformations
EAR	« Eosinophil-associated ribonuclease »
EPR	Résonance paramagnétique électronique
FDA	« Food and Drugs Administration » (agence américaine)
FRET	Transfert d'énergie de résonance de fluorescence
HIV-1 PR	Protéase du VIH
hNOE	Effet Overhauser nucléaire hétéronucléaire
HSQC	« Heteronuclear single-quantum coherence »
HsR2 / HsEDN	Ribonucléase 2 / « Eosinophil-derived neurotoxin » humaine
HsR3 / HsECP	Ribonucléase 3 / « Eosinophil cationic protein » humaine
k_B	Constante de Boltzmann
k_{cat}	Constante de renouvellement catalytique
k_{ex}	Constante d'échange conformationnel
K_i	Constante d'inhibition
K_M	Constante de Michaelis
MD	Simulation de dynamique moléculaire
MfR3 / MfECP	Ribonucléase 3 / « Eosinophil cationic protein » de *Macaca fascicularis*
MSM	Simulation de type « Markov State Model »
p_A / p_B	Population de l'état majoritaire / minoritaire
PaR3 / PaRNase 3	Ribonucléase 3 de *Pongo abelii*
PDB	« Protein Data Bank »
Pol β	ADN polymérase β (« DNA polymerase β »)
PpR3 / PpECP	Ribonucléase 3 / « Eosinophil cationic protein » de *Pongo pygmaeus*
QAA	Analyse quasi-anharmonique
R_1 / R_2	Taux de relaxation longitudinale / transverse
$R_{2,\text{eff}}$	Taux de relaxation transverse effectif
RDC	Couplage dipolaire résiduel
R_{ex}	Taux d'échange conformationnel
RMN / NMR	Résonance magnétique nucléaire
RMSF	« Root-mean-square fluctuations »
RNase	Ribonucléase
S^2	Paramètre d'ordre
VIH-1 / HIV-1	Virus de l'immunodéficience humaine, type 1
WT	Type sauvage

1 INTRODUCTION

1.1 Aperçu du projet de recherche

1.1.1 Problématique

Depuis plusieurs années, les milieux pharmaceutique et industriel manifestent un intérêt pour les enzymes, ces protéines capables de catalyser diverses réactions chimiques, comme alternatives à la chimie classique. Leurs attraits sont effectivement multiples. L'utilisation d'enzymes permet d'éviter l'utilisation de nombreux solvants organiques potentiellement nuisibles à l'environnement ou à la santé. Elles permettent aussi un excellent rendement réactionnel, limitant les produits secondaires indésirables et maximisant l'utilisation du substrat initial. De plus, elles ont l'avantage d'être énantiosélectives, alors que les synthèses chimiques le sont beaucoup plus rarement.

Bien que ces avantages soient indéniables, il peut être désirable d'avoir une enzyme qui soit un peu plus stable, un peu plus efficace, ou qui effectue une réaction légèrement différente, que celle qui est déjà disponible. C'est à cela que sert l'ingénierie des protéines. À titre d'exemple, la lipase B de *Candida antarctica* (CalB), vendue commercialement sous le nom de Novozyme 435, est utilisée pour catalyser de très nombreuses réactions chimiques (certaines sont décrites dans la revue de Gotor-Fernández *et al*. 2006). Étant donné son intérêt industriel, l'ingénierie des protéines a été utilisée afin de rendre CalB plus tolérante aux hautes températures (Zhang *et al*. 2003), d'accélérer la réaction d'acrylation catalysée par CalB avec des alcools secondaires comme substrats (Liu *et al*. 2010), ou alors de modifier la réaction primaire de CalB en réaction de Diels-Alder (Linder *et al*. 2011).

Toutefois, l'ingénierie de réactions enzymatiques est complexe : il est difficile de prédire l'effet des modifications sur la structure, la fonction et la dynamique des protéines. Par exemple, des chercheurs tentant de changer la spécificité de la trypsine afin qu'elle clive à la suite de résidus acides ont obtenu une enzyme clivant suite à de gros résidus hydrophobes (Graf *et al*. 1987). Bien que les outils de prédiction de structure soient de plus en plus efficaces (Nikolaev *et al*. 2018), renseignant sur l'effet de mutations sur la stabilité de la protéine et la structure de sa forme de base, la confirmation de la préservation de la structure n'est qu'une seule des étapes dans l'ingénierie des protéines. Plusieurs membres de la communauté scientifique utilisant la biologie structurale, mais ne travaillant pas dans ce domaine, considèrent encore la structure

d'une protéine comme étant statique, et il n'est pas rare d'entendre parler de la structure cristalline d'une protéine comme de la « vraie » structure (ce qui omet déjà l'information donnée par le facteur B, qui décrit l'étalement de la densité électronique). Cette perception erronée est confortée par le fait que plusieurs outils d'arrimage moléculaire (« molecular docking ») utilisent une structure immobile de la protéine afin de faire les calculs auxquels s'intéressent les « utilisateurs occasionnels » de la biologie structurale. Or, les protéines sont des entités très complexes et flexibles. Elles ne possèdent pas une structure unique mais bien un ensemble de conformations limites qu'elles explorent au fil du temps, avec différents temps de vie et populations. L'ensemble des structures ainsi accessibles peut être décrit par ce qu'on appelle le paysage conformationnel de la protéine (Frauenfelder *et al*. 1988; Ramanathan *et al*. 2014), concept décrit plus en détail dans la section 1.4.

Ces conformations alternatives sont fréquemment nécessaires à la fonction des protéines, mais leur faible population les rend difficiles à détecter expérimentalement. Ceci a pour conséquence de limiter fortement notre compréhension de ces changements de conformations, de cette flexibilité parfois inhérente à la fonction. Dans l'état actuel des connaissances sur le sujet, les échanges entre conformations ont été abondamment caractérisés chez des systèmes protéiques précis, par exemple la dihydrofolate réductase (DHFR) de *E. coli* (Bhabha *et al*. 2013), la cyclophiline A (Fraser *et al*. 2009), l'adénylate kinase (Henzler-Wildman *et al*. 2007a), ou la RNase A (Watt *et al*. 2007). Par contre, aucun lien, aucune règle n'existe pour former un tout cohérent sur ce sujet. Ceci a pour conséquence dans le cadre de l'ingénierie des protéines qu'une mutation aura souvent des conséquences imprévues, de par notre incapacité à prévoir ses effets sur les conformations transitoires de l'enzyme et/ou sur les échanges permettant d'y accéder. Ainsi, la communauté scientifique ne peut que constater les effets des échanges conformationnels plutôt que de les utiliser à son avantage. Améliorer la compréhension des échanges entre les conformations des protéines pourrait donc faciliter l'ingénierie de nouvelles réactions enzymatiques.

1.1.2 Hypothèse

S'il est si complexe de prédire l'effet de mutations sur la fonction et sur les conformations transitoires d'une enzyme, peut-être la nature a-t-elle des réponses à fournir. Il a d'ores et déjà été démontré que l'évolution exerce une pression de sélection sur les protéines afin de conserver ou d'améliorer les fonctions importantes (Liu *et al*. 2008). Il est bien connu que le site actif d'une enzyme constitue l'ensemble d'acides aminés le mieux conservé, ce qui découle

directement de cette importance fonctionnelle. Toutefois, bien qu'il soit connu que les conformations transitoires d'une protéine peuvent être elles aussi cruciales à sa fonction (tel que discuté dans la section 1.4.2), peu d'informations sont disponibles quant à leur conservation par l'évolution. En effet, il est logique de supposer que l'accès d'une protéine à des conformations plus énergétiques sera évolutivement avantageux à conserver si cette caractéristique est importante pour l'organisme. L'hypothèse globale de cette thèse se résume donc ainsi : les changements conformationnels importants pour la (les) fonction(s) d'enzymes auront été conservés par l'évolution si la structure et la (les) fonction(s) sont conservées.

Vérifier la présence ou l'absence de cette conservation et identifier les déterminants qui y sont reliés impliqueraient idéalement de pouvoir discriminer l'effet des structures primaire et tertiaire, de la fonction, et de la distance évolutive. Ceci est bien sûr impossible, puisque toutes ces caractéristiques sont intrinsèquement interreliées, mais il devrait être possible de minimiser l'effet de la structure tertiaire et de la fonction en utilisant comme modèles plusieurs protéines d'une même famille.

Afin de comprendre la direction de cette thèse, il importe de mentionner ceci : les différentes expériences et analyses présentées dans les articles de cette thèse ont été planifiées afin de vérifier l'hypothèse susmentionnée, mais au fil de l'accumulation des résultats, celle-ci s'est avérée erronée. En effet, bien que des ressemblances aient pu être identifiées entre plusieurs des enzymes étudiées, un trop grand nombre d'exceptions aurait été nécessaire pour justifier la conservation de l'hypothèse originale. La seule conclusion logique permettant d'expliquer à la fois les ressemblances et les différences est la divergence au fil de l'évolution de la façon dont les mouvements intrinsèques se produisent au sein de la superfamille des RNases.

1.1.3 Objectifs

Les modèles utilisés dans ce projet sont des ribonucléases (RNases) de la famille de la RNase A. Ce sont des enzymes ayant toutes une structure similaire mais aux fonctions biologiques variées, qui sont décrites plus en détail dans la section 1.2. Ces enzymes-modèles permettront d'étudier différents aspects du lien entre les changements de conformation, la fonction catalytique, les fonctions biologiques et l'évolution, ce qui sera effectué en plusieurs étapes : 1- étudier l'effet de la distance évolutive sur les échanges conformationnels des RNases, 2- étudier le lien entre les changements conformationnels et l'activité ribonucléolytique, 3- étudier les changements conformationnels chez des orthologues d'une même enzyme. Plusieurs méthodes d'analyse utilisées pour atteindre ces objectifs sont décrites dans l'article de revue

présenté au chapitre 2, « Role of conformational motions in enzyme function : selected methodologies and case studies » (Narayanan, Bernard et Doucet 2016).

Donald Gagné, un ancien collègue du laboratoire, a pu démontrer que les mouvements coordonnés que subissent les différentes RNases humaines étaient uniques à chaque protéine et qu'ils peuvent être compartimentalisés similairement (Gagné *et al*. 2012; Gagné *et al*. 2013). Afin d'interpréter correctement ces résultats, le premier objectif du projet a été de remettre ces résultats, ainsi que d'autres résultats similaires obtenus par la suite, dans le contexte de l'évolution de ces protéines, et d'y ajouter la perspective de la distance évolutive. En d'autres mots, il s'agit de comprendre ce qui a mené aux arrangements actuels d'acides aminés subissant des échanges conformationnels chez les RNases humaines, et voir ce qui peut réellement être comparé. Les résultats des analyses correspondant à cet objectif sont présentés dans l'article « Conservation of dynamics associated with biological function in an enzyme superfamily » (Narayanan, Bernard *et al*. 2018a), reproduit au chapitre 3.

L'aspect le mieux conservé chez les RNases est leur capacité à cliver l'ARN. Or, le phénomène de liaison implique fréquemment des réorganisations structurales (sélection conformationnelle et/ou ajustement induit) qui seront donc liées à la fonction de l'enzyme associée. Ces mouvements impliquent des conformations transitoires qui peuvent être accessibles ou non à la protéine dans sa forme libre. Il est logique de penser que si des changements conformationnels sont conservés par l'évolution, ceux liés à une caractéristique essentielle des RNases ont de fortes chances d'être conservés. Le deuxième objectif de cette thèse consistait donc à déterminer la similitude des changements conformationnels subis par différentes RNases humaines suite à la liaison de différents ligands. Cet objectif est exploré en détails par l'article « Ligand-Induced Variations in Structural and Dynamical Properties Within an Enzyme Superfamily » (Narayanan, Bernard *et al*. 2018b), reproduit au chapitre 4.

Tel qu'il le sera détaillé dans la section 1.2.3, les différentes sous-familles de RNases possèdent toutes des fonctions biologiques différentes. Vérifier le lien entre conservation de la fonction et conservation des échanges conformationnels implique donc l'étude d'enzymes différentes possédant une même fonction biologique. C'est pourquoi des RNases provenant d'une seule sous-famille ont servi de modèles pour le troisième objectif, qui cherche à vérifier ce lien. Ces résultats sont présentés dans l'article « Divergence of conformational motions along the evolutionary pathway in the eosinophil ribonuclease subfamily » (Bernard *et al*. en préparation), reproduit au chapitre 5.

1.2 **Ribonucléases**

Cette section vise à présenter les protéines servant de modèles dans le présent projet de recherche. Toutes sont membres de la superfamille des ribonucléases, parfois appelée les RNases pancréatiques (pour des raisons purement historiques) ou la superfamille de la RNase A, d'après son membre le plus connu.

1.2.1 Ribonucléase A : une protéine historique

La superfamille étudiée ici a comme membre emblématique l'une des protéines les plus étudiées du 20^{e} siècle : la ribonucléase A bovine (RNase A), également appelée la RNase pancréatique bovine (Raines 1998). Dans les années 1940, la compagnie américaine d'emballage de viande Armour and Company a pu isoler une énorme quantité de RNase A, soit plus d'un kilogramme de protéine pure, qu'ils ont décidé de vendre à bas prix à de nombreux chimistes des protéines à travers le monde, ce qui a mené à des avancées à la fois nombreuses et rapides en science des protéines (Richards 1972). Les travaux effectués sur cette protéine étaient d'une importance telle qu'ils ont mené à deux prix Nobel de chimie. En 1972, cette récompense a été attribuée à Christian Anfinsen, Stanford Moore et William Stein. Anfinsen a montré que la structure tertiaire d'une protéine était entièrement codée dans sa séquence d'acides aminés – et prouvant de surcroît que la structure correctement repliée d'une protéine était adoptée à son énergie libre de Gibbs (ΔG) minimale (Anfinsen 1959). De leur côté, Moore et Stein ont utilisé la RNase A comme modèle pour démontrer le lien entre la structure d'une enzyme et son mécanisme catalytique (Crestfield *et al*. 1963). En 1984, le prix Nobel de chimie a été décerné à Bruce Merrifield pour l'invention de la synthèse chimique en phase solide de protéines, et dans ses travaux il a justement réussi à effectuer la synthèse peptidique totale de la RNase A catalytiquement active (Gutte et Merrifield 1971).

Le mécanisme catalytique de la RNase A est le tout premier mécanisme catalytique enzymatique à avoir été déduit, ce qui représente une avancée majeure en biochimie (Findlay *et al*. 1961; Cuchillo *et al*. 2011). Ce mécanisme endonucléolytique, conservé par toutes les RNases catalytiquement actives, est reproduit ici à la figure 1.1. Deux histidines, l'une protonée servant d'acide général, et l'autre non protonée servant de base générale, catalysent le clivage d'un brin simple d'ARN en effectuant une réaction de transphosphorylation, ce qui mène à un phosphomonoester cyclique. Une lysine permet la stabilisation de l'état de transition. Les

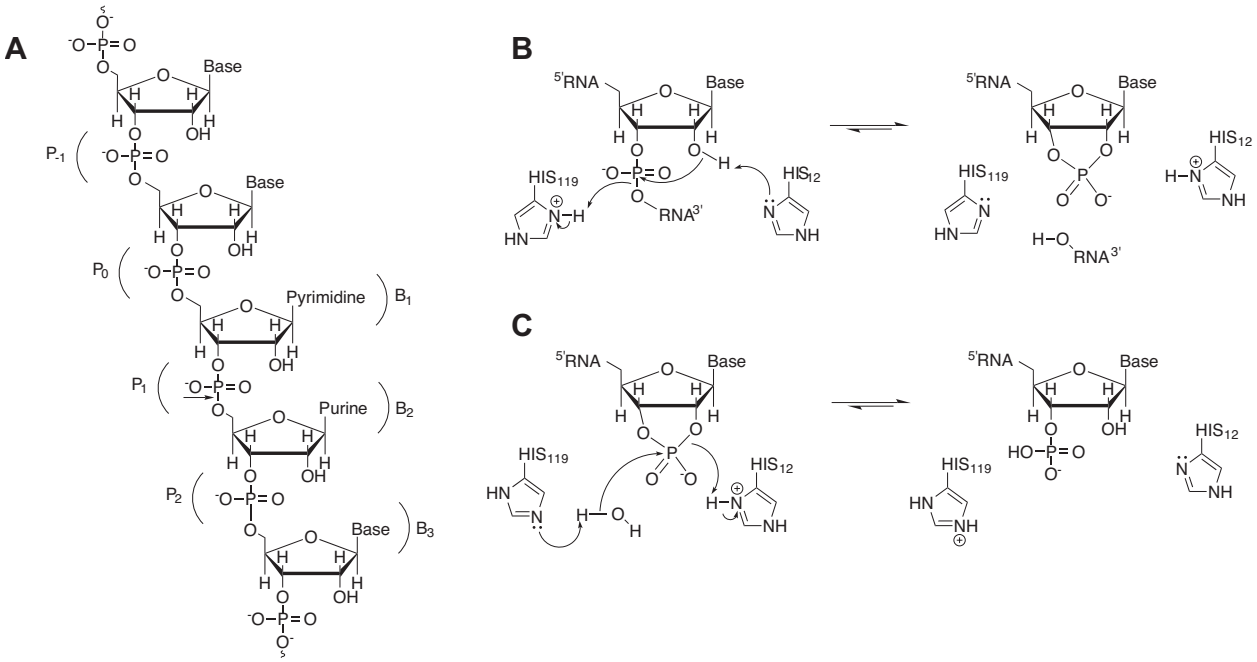


Figure 1.1 Mécanisme réactionnel de la ribonucléase A (tiré de Gagné et Doucet 2013)

Schéma du mécanisme chimique catalysé par la RNase A. En (A) sont représentés les différents sous-sites de liaison que la protéine reconnaîtra sur le brin d'ARN. La liaison chimique clivée est indiquée par une flèche, dans le sous-site P_1. En (B) et (C) sont illustrées les étapes de transphosphorylation (B) et d'hydrolyse (C) effectuées par la RNase A.

mêmes résidus peuvent par la suite catalyser la réaction d'hydrolyse permettant de reformer un ribose-3'-phosphate, mais il arrive que les produits de la réaction soient libérés sans que l'hydrolyse soit effectuée (Gagné *et al*. 2013). La RNase A préférera cliver une séquence formée d'une pyrimidine suivie d'une adénosine (Raines 1998), lui permettant de cliver à peu près n'importe quel ARN simple brin considérant l'abondance et la faible spécificité de cette séquence.

1.2.2 Structure des ribonucléases

Toutes les RNases sont de petites protéines de ~15 kDa et ~120 acides aminés, possédant une triade de résidus catalytiques (pour les enzymes actives à tout le moins) formée de deux histidines et d'une lysine retrouvées dans les séquences hautement conservées FXXQH, CKXXNTF et PVHXD (voir l'alignement de séquences des huit RNases humaines et de la RNase A à la figure 1.2). Elles possèdent également quatre ponts disulfures très bien conservés, à l'exception de la RNase 5 qui n'en possède que trois (Sorrentino *et al*. 2010), et la

RNase 8 qui possède deux cystéines vraisemblablement non appariées, ou alors permettant une potentielle multimérisation (Zhang *et al*. 2002).

La RNase A fut historiquement la troisième enzyme et la quatrième protéine toutes catégories confondues dont la structure a pu être déterminée par cristallographie (Kartha *et al*. 1967; Raines 1998). Toutes les enzymes de sa superfamille possèdent un repliement similaire à celui représenté à la figure 1.3. À l'exception de la RNase 8, au moins une structure est retrouvée sur la PDB pour chacune des RNases humaines, avec par exemple les identifiants 2k11, 1gqv, 2kb5, 1rnf, 1awz, 5oab et 2hky respectivement pour les RNases humaines 1 à 7, et 7rsa pour la RNase A. Ces protéines comportent deux feuillets β orientés en forme de V (et nommés V_1 et V_2), trois hélices α, et la triade catalytique est prise en sandwich entre les deux branches du V et l'hélice α1. De six à sept boucles complètent la structure.

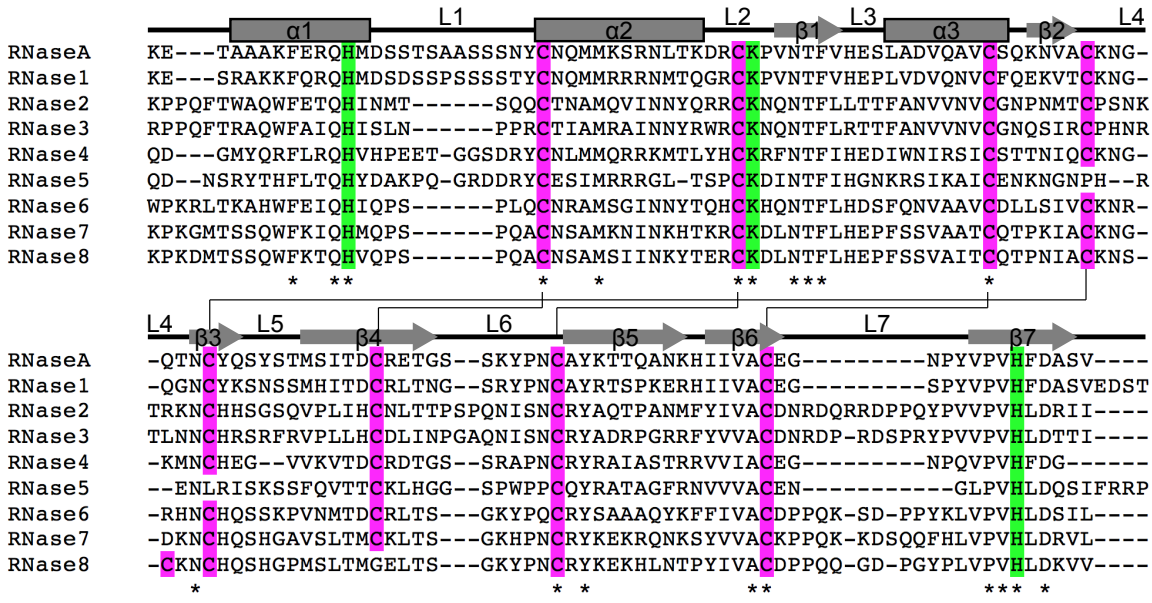


Figure 1.2 Alignement de séquences des RNases humaines et de la RNase A

Alignement de séquences entre les huit RNases humaines et la RNase A (séquences tirées de UniProt) réalisé par le serveur T-Coffee (http://tcoffee.crg.cat/). Les acides aminés de la triade catalytique sont surlignés en vert, les cystéines participant aux ponts disulfures sont surlignées en magenta, et les acides aminés conservés à 100% sont indiqués par une astérisque *. Les cystéines appariées sont reliées par un trait, et les éléments de structure secondaire sont indiqués au-dessus.

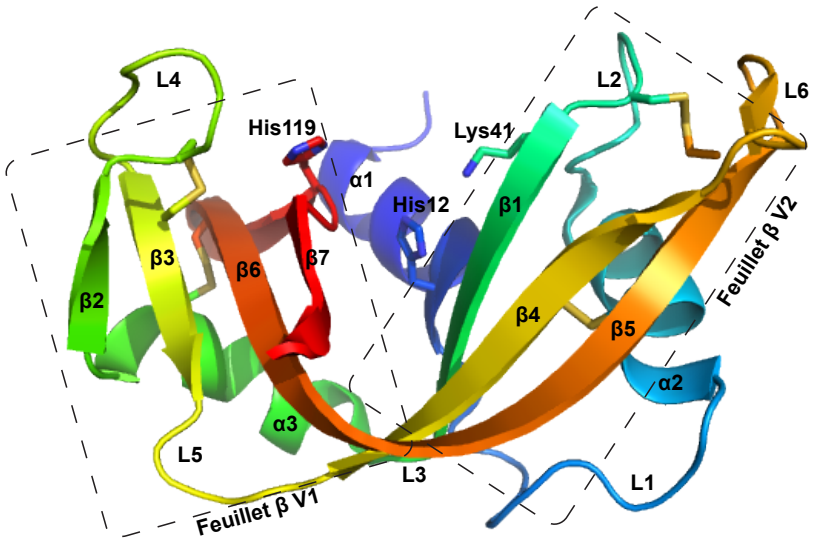


Figure 1.3 Structure tridimensionnelle d'une ribonucléase

Représentation « cartoon » de la structure d'une RNase typique, la RNase A (identifiant PDB 7rsa). Les différents éléments de structure secondaire sont annotés, les lettres α, β et L représentant respectivement les hélices α, les brins β et les boucles. Les résidus de la triade catalytique et les cystéines formant les ponts disulfures stabilisant la structure de la protéine sont visibles en bâtonnets, et les résidus catalytiques sont identifiés. La structure a été illustrée grâce au logiciel PyMOL.

1.2.3 Ribonucléases humaines : membres de la famille et fonctions

Les RNases de la superfamille de la RNase A sont des enzymes retrouvées exclusivement chez les vertébrés (Sorrentino 2010), bien que des preuves de l'existence d'une distante branche bactérienne de la superfamille commencent à émerger (Cuthbert *et al*. 2018). Chez l'humain ont été identifiés huit gènes codant pour des RNases (RNases 1 à 8), ainsi que cinq

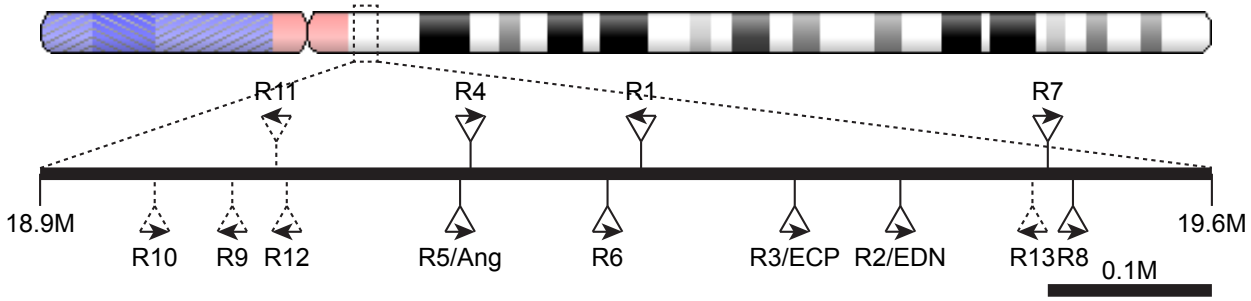


Figure 1.4 Localisation des gènes codant pour les RNases humaines

Représentation du chromosome 14 de *Homo sapiens*. Le centromère est identifié en rose, et la région codant pour les RNases est encadrée en pointillés. Les localisations sont à l'échelle, et les flèches représentent le sens de la transcription. Les localisations représentées en pointillés correspondent aux pseudogènes. Modifié d'après (Cho *et al*. 2005), l'image du chromosome est tirée de NCBI.

pseudogènes codant pour des enzymes homologues mais catalytiquement inactives (RNases 9 à 13), et tous sont regroupés sur le chromosome 14 (figure 1.4) (Cho *et al*. 2005). Malgré le fait qu'elles soient issues d'un unique gène qui s'est dupliqué à plusieurs reprises au fil de millions d'années d'évolution, les fonctions biologiques et l'activité catalytique de ces RNases sont très différentes les unes des autres. Différentes constantes catalytiques de plusieurs membres de la famille sont comparées dans le tableau 1.1.

Tableau 1.1 Constantes catalytiques de différentes RNases sur le dinucléotide UpA

Protéine	pI	k_{cat}	k_{cat} relatif à RNase A	K_M	K_M relatif à RNase A	k_{cat}/K_M	k_{cat}/K_M relatif à RNase A
		s^{-1}		µM		$M^{-1}s^{-1}$	
RNase A	8.64	471±26	1.000	130±35	1.00	$3.62*10^{6}$	1.000
RNase 2	9.20	87.83±3.90	0.186	520±54	4.00	$1.69*10^{5}$	0.047
RNase 3	10.47	0.97±0.06	0.0021	1230±220	9.46	789	$2.18*10^{-4}$
RNase 4	9.18	16.4±2.1	0.035	440±160	3.38	$3.73*10^{4}$	0.010
RNase 5	9.73	0.031±0.006	$6.58*10^{-5}$	2500±880	19.23	12.4	$3.43*10^{-6}$
RNase 6	9.22	5.70±0.14	0.0121	140±20	1.077	$4.07*10^{4}$	0.011

Ces paramètres ont été mesurés dans le cadre du présent projet de recherche par Purva Prashant Bhojane et Khushboo Bafna dans le laboratoire de Elizabeth E. Howell.

La RNase 1 humaine a la particularité d'être la plus active catalytiquement de toutes les RNases caractérisées jusqu'à présent, et elle est l'une des très rares RNases à cliver l'ARN double brin. Cette dernière caractéristique lui serait conférée par des résidus basiques situés à proximité du site actif (Kövér *et al*. 2008). Elle est exprimée par le pancréas afin de digérer l'ARN contenu dans la nourriture (Sorrentino 2010), elle contribuerait à l'homéostasie vasculaire en digérant l'ARN extracellulaire lorsqu'elle est sécrétée par les cellules endothéliales vasculaires (Lu *et al*. 2018), elle permet la maturation des cellules dendritiques et augmente leur production de cytokines (Kövér *et al*. 2008). D'autres tissus expriment également cette enzyme, et la variété de fonctions qu'elle effectue dépendrait des modifications post-traductionnelles qu'elle subit (Sorrentino 2010).

La RNase 2, également appelée « eosinophil-derived neurotoxin » (EDN), et la RNase 3, également appelée « eosinophil cationic protein » (ECP) sont deux des quatre toxines majeures sécrétées par les éosinophiles. La RNase 2 est de plus exprimée par les cellules dendritiques et les monocytes. Les éosinophiles faisant partie du système immunitaire inné, leur mode d'action est de libérer des agents antimicrobiens peu spécifiques mais capables d'agir sur une large gamme d'agents pathogènes contre lesquels le corps ne possède pas d'anticorps (Gleich et

Adolphson 1986; Hogan *et al*. 2008). Dans cette optique, la RNase 2 possède une activité antivirale efficace contre plusieurs virus à ARN, incluant le VIH, le virus respiratoire syncytial (RSV) ou le virus de l'hépatite B (Lu *et al*. 2018). Cette activité antivirale est due en partie à sa forte efficacité catalytique, bien que celle-ci ne soit pas suffisante pour expliquer totalement cette activité puisque la RNase A, qui a pourtant une efficacité catalytique supérieure, ne possède aucune activité antivirale (Rosenberg 2008). La RNase 2 agit de plus comme une alarmine, et permet l'activation, la maturation et le chimiotactisme des cellules dendritiques (Yang *et al*. 2003). La RNase 3, de son côté, participe au système immunitaire inné via ses propriétés antibactérienne, antivirale, helminthotoxique et cytotoxique. Son activité antivirale fonctionne de la même manière que celle de la RNase 2, mais est dix fois moins puissante (Zhang et Rosenberg 2002). Elle est capable de combattre la prolifération des bactéries *E. coli* et *Staphylococcus aureus* (Lehrer *et al*. 1989), le parasite helminthique *Leishmania donovani* (Singh et Batra 2011) et est cytotoxique pour des cellules humaines (Navarro *et al*. 2008). Ces trois activités seraient dues au même mécanisme, qui est indépendant de l'activité catalytique de la protéine, et qui impliquerait l'agrégation de l'enzyme à la membrane dans un mécanisme « carpet-like » (Salazar *et al*. 2014). Structurellement, les RNases 2 et 3 sont caractérisées par une courte boucle 1, par l'absence du réseau de ponts hydrogènes médiés par une histidine (His48 chez la RNase A) reliant cette boucle au sous-site B_1 (Boix *et al*. 1999a; Swaminathan *et al*. 2002; Doucet *et al*. 2011), ainsi que par la présence d'une longue boucle 7, tel que montré dans l'alignement de séquences à la figure 1.2.

La particularité de la RNase 4 est son haut degré de conservation comparativement aux autres RNases : celui-ci oscille entre 81 et 94% d'identité de séquence entre les RNases 4 des différents vertébrés (Hofsteenge *et al*. 1998). Comme beaucoup de RNases, elle est exprimée par de nombreux tissus : cœur, reins, poumons, muscles, ovaires, rate, et son expression est la plus forte dans le foie (Hofsteenge *et al*. 1998). Sa fonction exacte demeure nébuleuse, mais elle a une action angiogénique et elle semble impliquée dans la neuroprotection des neurones moteurs. En effet, un polymorphisme détecté dans le peptide signal a été identifié chez des patients atteints de sclérose latérale amyotrophique (Li *et al*. 2013).

La RNase 5 est plus connue sous le nom d'angiogénine, ayant d'abord été identifiée comme facteur angiogénique avant d'être associée à la superfamille des RNases suite à son séquençage (Sorrentino 2010). Elle possède une boucle 4 tronquée dépourvue de pont disulfure, et son activité ribonucléolytique est la plus faible de toutes les RNases humaines à cause entre autres d'une séquence supplémentaire en C-terminal (voir figure 1.2) qui adopte

une conformation hélicale. Un résidu glutamine vient alors créer un encombrement stérique limitant l'accès au site catalytique (Leonidas *et al*. 1999). Cette activité catalytique minime, qui de plus cible des substrats plus spécifiques que les autres RNases (Lyons *et al*. 2017), est toutefois nécessaire à son activité angiogénique (Leonidas *et al*. 1999). Son mode de fonctionnement commence par sa sécrétion par les cellules endothéliales vasculaires et les cellules musculaires lisses, et elle sera internalisée par ces mêmes types cellulaires de manière autocrine ou paracrine (Lyons *et al*. 2017). La RNase 5 subira alors une translocation vers le noyau. L'un des mécanismes par lesquels elle promeut la prolifération cellulaire est alors de cliver un petit ARN associé à un promoteur, et la dégradation de cet ARN empêche l'extinction (« silencing ») de l'ADN ribosomal (Hoang et Raines 2017). Cette protéine est dérégulée dans de nombreux cas de cancer (surexpression) ainsi que, comme pour la RNase 4, chez des patients atteints de sclérose latérale amyotrophique (sous-expression) (Hoang et Raines 2017).

L'expression de la RNase 6 (ou RNase k6) est induite par des infections bactériennes dans de nombreux types cellulaires, dont les monocytes, les neutrophiles, les macrophages et des cellules épithéliales du système urinaire (Sorrentino 2010). Elle serait impliquée dans le système immunitaire également, ayant une action antibactérienne (quoique moins puissante que celle de la RNase 3) contre *E. coli*, *Enterococcus faecalis* et *Staphylococcus saprophyticus* (Becknell *et al*. 2015). La RNase 6 a la particularité de posséder une paire d'histidines placées adéquatement pour former un deuxième site catalytique sur sa boucle 2 (Prats-Ejarque *et al*. 2016).

Bien qu'elle puisse être exprimée par de nombreux tissus (Zhang *et al*. 2003), la RNase 7 est majoritairement exprimée sur la peau par les kératinocytes (Harder et Schroder 2002). Ceci en fait l'une des raisons pour lesquelles travailler avec de l'ARN est si laborieux, car le simple contact d'un échantillon avec la peau mettra ledit échantillon en présence d'une enzyme capable de dégrader quasiment n'importe quelle séquence. Tout comme la RNase 3, la RNase 7 est hautement cationique (pI de 10.5), ce qui lui est conféré par le grand nombre de lysines à sa surface (Huang *et al.* 2007), et tout comme la RNase 3, elle possède une activité bactéricide – indépendante de son activité catalytique – contre des bactéries à Gram positif et négatif (Harder et Schroder 2002).

L'ARN messager de la RNase 8 a d'abord été identifié comme étant exprimé exclusivement dans le placenta (Zhang *et al*. 2002), mais des expériences subséquentes ont permis sa détection dans les poumons, la rate et les testicules (Chan *et al*. 2012). Elle n'existe que chez les primates, mais sa fonction demeure pour l'instant inconnue (Chan *et al*. 2012).

En bref, les RNases font d'excellents modèles pour une étude sur l'évolution des protéines de par la conservation de leur structure tertiaire associée à la pluralité de leurs rôles biologiques.

1.3 Évolution des protéines

Afin de mettre en contexte les analyses de conformations alternatives et d'échanges conformationnels, il vaut la peine de comprendre en quoi consiste le contexte en question. L'évolution des protéines est un processus souvent étudié dans le but de comprendre l'émergence de nouvelles fonctions (Akiva *et al*. 2017), ou afin de déterminer l'importance de certaines sections d'une protéine en particulier (Zhang et Rosenberg 2002).

1.3.1 Phénomènes impliqués dans l'évolution des RNases

Au cours des milliards d'années de l'existence de la vie sur Terre, les êtres vivants sont passés du stade unicellulaire à des organismes complexes comprenant des dizaines de milliers de gènes, codant pour un nombre encore largement supérieur de protéines grâce à des phénomènes tels l'épissage alternatif (Nilsen et Graveley 2010) ou les cadres de lecture alternatifs (Vanderperre *et al*. 2013), et ceci au sein d'un seul organisme. De nombreux mécanismes ont permis l'apparition et la diversification des protéines existant de nos jours, mais cette section se limitera à mentionner ceux jouant un rôle majeur pour la superfamille des RNases (Goo et Cho 2013).

Tel que mentionné précédemment, les RNases sont un ensemble d'enzymes exclusif aux vertébrés, et les gènes codant pour celles-ci se retrouvent tous sur le même chromosome (figure 1.4). La majorité des vertébrés ont deux ou trois gènes codant pour des RNases dans leur génome, mais les mammifères en présentent une grande diversité, allant jusqu'à une vingtaine de gènes et de pseudogènes chez la souris (Cho *et al*. 2005). Cette multiplicité proviendrait de duplications successives des gènes d'origine (Goo et Cho 2013). Généralement, suite à une duplication de gènes, les deux copies pourront subir au fil des générations plusieurs destins possibles : une conservation de la fonction, la pseudogénisation, une sous-fonctionnalisation, ou alors l'apparition d'une nouvelle fonction (Magadum *et al*. 2013).

Les deux copies d'un gène dupliqué peuvent toutes deux conserver leur fonction originale, pouvant servir de « copie de sûreté » pour protéger un gène de mutations délétères, ou alors afin de permettre une augmentation de l'expression de ce gène en particulier (Magadum *et al*. 2013). Il s'agit toutefois d'une conséquence rare des duplications. La majorité des gènes

dupliqués finissent par disparaître par pseudogénisation (Lynch et Conery 2000). Le procédé par lequel une copie d'un gène devient un pseudogène implique l'accumulation de mutations ou de délétions qui perturbent l'expression ou l'épissage du gène, ou alors la structure ou la fonction de la protéine pour laquelle code ce gène. Ceci mène à une protéine qui sera alors non exprimée, ou sans fonction (Magadum *et al*. 2013 ; Lynch et Conery 2000). Ainsi, cinq RNases (les RNases 9 à 13) sont considérées comme des pseudogènes chez l'humain, n'ayant pas conservé leur activité catalytique (Cho *et al*. 2005).

Les gènes dupliqués qui arrivent à éviter la pseudogénisation auront divergé l'un de l'autre et finiront avec des séquences et des modes de régulation parfois très différents, donnant lieu à des protéines complètement indépendantes l'une de l'autre (Diss *et al*. 2017). Il arrive que les deux copies se séparent la (les) fonction(s) du gène original, un processus désigné comme la sous-fonctionnalisation (Magadum *et al*. 2013). Bien qu'il n'ait pas été formulé que ce phénomène soit directement impliqué, il est raisonnable de présumer que la diversification de plusieurs sous-familles de RNases se soit produite par sous-fonctionnalisation. En effet, alors que la majorité des mammifères possèdent un seul gène de RNase 5, les rongeurs (dont *Mus musculus* et *Rattus norvegicus*) en possèdent plusieurs (Codoñer *et al*. 2010; Hooper *et al*. 2003), et l'activité catalytique de même que l'activité angiogénique de chacune d'entre elles est inférieure à celles de la protéine humaine (Crabtree *et al*. 2007; Iyer *et al*. 2013). Similairement, la sous-famille des RNases 1 n'a qu'un seul représentant chez la majorité des mammifères, mais les mustéloïdes en possèdent une grande diversité (Liu *et al*. 2014).

Finalement, il arrive que les copies de gènes développent une fonction entièrement nouvelle. Dans le cas d'enzymes, ceci signifie souvent de catalyser la même réaction chimique sur des substrats différents, ou alors catalyser une réaction différente mais qui possède une caractéristique commune, par exemple un intermédiaire réactionnel (Newton *et al*. 2017).

Suite aux duplications, les gènes ancestraux des RNases ont subi des mutations afin d'acquérir leur séquence actuelle. Les polymorphismes nucléotidiques (« single-nucleotide polymorphisms ») créent des substitutions dans la séquence d'un gène qui peuvent affecter (polymorphismes non-synonymes) ou pas (polymorphismes synonymes) la séquence de la protéine (Pál *et al*. 2006). Les polymorphismes non-synonymes mèneront à des mutations, qui peuvent être délétères (30 à 50% des mutations), bénéfiques dans de rares cas (0.01 à 1%), ou alors quasi neutres (50 à 70%) (Liu *et al*. 2008; Romero et Arnold 2009) quant à la fonction de la protéine pour laquelle code le gène muté. Ceci peut être conceptualisé dans le « fitness landscape » décrit par la lauréate du prix Nobel Frances Arnold comme étant le phénotype en

fonction du génotype, ou alors la cartographie de l'activité catalytique en fonction de l'espace des séquences (Romero et Arnold 2009). Une mutation peut alors être considérée comme un déplacement dans l'espace des séquences, et celle-ci sera potentiellement accompagnée d'un changement dans le phénotype associé (figure 1.5). Dans le cas d'une protéine dont la fonction est nécessaire à l'organisme, des pressions de sélection s'appliqueront pour empêcher les mutations délétères et favoriser les mutations bénéfiques (Pál *et al*. 2006), mais c'est plutôt suite à l'accumulation de mutations neutres que des nouvelles fonctions peuvent apparaître. En effet, ces mutations ne subissent que peu de pression de sélection, et il arrive que des combinaisons particulières de mutations ayant un effet neutre sur la fonction originale permettent l'apparition d'une fonction secondaire (Romero et Arnold 2009). Par conséquent, c'est ainsi que la diversification de gènes dupliqués peut se produire.

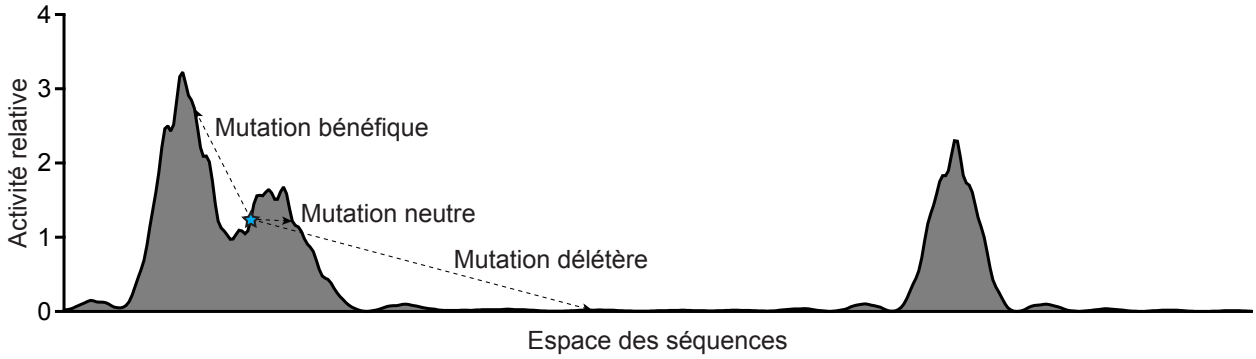


Figure 1.5 Effets de mutations sur l'activité d'une enzyme

Exemple du « fitness landscape » d'une enzyme. L'enzyme WT est située à l'étoile cyan. Une mutation correspond à un déplacement latéral dans l'espace des séquences, et pourrait potentiellement être accompagnée d'un changement dans l'activité.

Dépendamment du rôle de l'acide aminé subissant la mutation dans la protéine WT, certains polymorphismes seront mieux tolérés que d'autres. En effet, des réseaux d'acides aminés spatialement contigus, nommés « secteurs, » peuvent être identifiés à l'intérieur de la structure tridimensionnelle d'une protéine, et ces réseaux régissent des rôles indépendants, tels la stabilité de la protéine, la catalyse, ou la liaison d'un cofacteur (Halabi *et al*. 2009; Reynolds *et al*. 2011; Narayanan *et al*. 2017). Une mutation dans un secteur associé à une fonction autre que la stabilité de la structure a donc davantage de chance d'être tolérée (Marsh 2009), en plus de favoriser des mutations compensatoires ou épistatiques (Storz 2018) au sein du même secteur. En effet, ces secteurs sont caractérisés par la tendance de leurs acides aminés à co-évoluer (Halabi *et al*. 2009). Ceci permet en outre d'accélérer la dérive génétique lorsque le

secteur est approprié. Des travaux précédents effectués par notre laboratoire ont d'ailleurs identifié la présence de tels secteurs chez la superfamille des RNases (Narayanan *et al*. 2017).

Outre la présence de secteurs, une dérive génétique accélérée est aussi favorisée par la présence de ponts disulfures dans la structure tertiaire d'une protéine. En effet, leur effet stabilisateur offre une contrepartie à de nombreuses mutations parfois délétères et permet la conservation d'un repliement tridimensionnel particulier malgré une entropie conformationnelle accentuée. Ceci a pour conséquence une plus grande tolérance aux mutations chez ces protéines, autorisant un rythme accéléré d'évolution (Feyertag et Alvarez-Ponce, 2017). Incidemment, les RNases font partie de ces types de protéines, tel que mentionné à la section 1.2.2.

1.3.2 Méthodes d'analyse de l'évolution des protéines

L'étude de l'évolution des protéines est rendue possible grâce à la connaissance d'un large ensemble de séquences de protéines, de gènes et de génomes entiers, qui permettent l'identification d'homologues et l'alignement de leurs séquences. Plusieurs méthodes bioinformatiques peuvent ensuite être utilisées pour déterminer lesquels des homologues d'une protéine sont les plus proches les uns des autres, et à quelle distance évolutive se sont séparés leurs ancêtres. Aucune de ces méthodes n'est exacte, par contre, il s'agit d'extrapolations utilisant statistiques et probabilités dont les résultats peuvent varier d'une méthode à l'autre (Gabaldón 2007). Une fois ces prédictions faites, les séquences seront schématisées sous forme d'arbre, appelé « arbre phylogénétique. » Un exemple est illustré à la Figure 1.6.

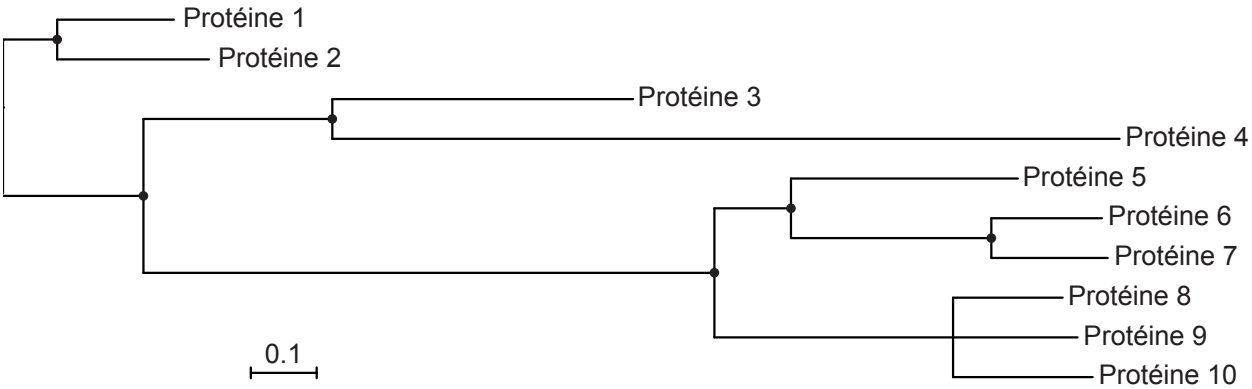


Figure 1.6 Exemple d'arbre phylogénétique

Arbre phylogénétique fictif montrant les relations entre dix protéines. L'échelle en bas à gauche correspond au nombre de mutations par site. Les points noirs correspondent aux ancêtres communs les plus récents.

Les protéines représentées sur un arbre phylogénétique ont évolué à partir de la gauche vers la droite (ou du centre vers les extrémités), afin d'obtenir les protéines actuelles identifiées au bout de chacune des branches. La distance horizontale correspond au nombre de mutations par site, et la distance verticale n'est pas importante. Les embranchements correspondent aux ancêtres communs les plus récents, et la distance évolutive entre deux protéines est égale à la distance entre celles-ci et leur ancêtre commun. Un embranchement peut être la résultante d'une duplication ou alors de la séparation de deux espèces (Loes *et al*. 2018). Des embranchements multiples, comme le triple embranchement entre les protéines 8, 9 et 10 à la figure 1.6, signifient que le système de prédiction a été incapable de déterminer laquelle des différenciations s'est produite avant l'autre. Des branches d'un arbre phylogénétique regroupées par espèce dénotent des duplications du gène au sein de l'espèce en question, alors que des branches regroupées par famille de protéines en mélangeant les espèces montre que les gènes ont divergé avant que les espèces ne le fassent (Goo et Cho 2013). Dans le premier cas, les homologues peuvent être appelés des paralogues, alors qu'ils seront désignés comme orthologues dans le deuxième cas (Gabaldón 2007).

L'analyse d'arbres phylogénétiques permet de comprendre un grand nombre d'événements directement liés à l'évolution d'un ensemble de protéines. Ils permettent d'établir les liens d'orthologues ou de paralogues, de détecter les transferts horizontaux de gènes (fréquents chez les bactéries), d'identifier les duplications de gènes ou de génomes, d'identifier des acides aminés ayant subi une pression sélective positive, ou d'estimer les taux de mutation (Gabaldón 2007). De plus, la connaissance des proches homologues et de l'ordre dans lequel ils se sont éloignés les uns des autres permet d'inférer la séquence des protéines ancestrales (Thornton 2004; Joy *et al*. 2016). De là, il est possible d'exprimer de manière recombinante ces protéines ancestrales et de les caractériser, révélant ainsi de nombreuses propriétés de prédécesseurs des protéines modernes, et ainsi délivrant plusieurs secrets de l'évolution (Thornton 2004). Encore une fois, cette méthode d'analyse consiste en une extrapolation qu'aucune expérience ne peut confirmer ou infirmer, et elle dépend de la qualité du modèle phylogénétique utilisé comme entrée (Thornton 2004; Joy *et al*. 2016).

En bref, les propriétés des protéines n'apparaissent pas spontanément, elles sont des conséquences logiques de phénomènes biologiques et biochimiques regroupés sous le nom d'évolution. Ces phénomènes, bien que liés au hasard, peuvent être compris, étudiés, et même utilisés à l'avantage de la science.

1.4 Paysages conformationnels

Le but du projet de recherche présenté dans cette thèse tourne autour de la caractérisation des échanges entre les conformations qu'une enzyme peut adopter. Une molécule aussi complexe qu'une protéine ne peut être entièrement décrite par une unique structure statique. Des mouvements allant de la rotation d'une unique chaîne latérale à une réorganisation complète de son squelette polypeptidique peuvent survenir au fil du temps, et elle adoptera plusieurs conformations limites. La Protein Data Bank (PDB) contient par exemple 168 entrées pour la RNase A bovine de type sauvage (« wild-type, » donc en excluant les mutants et la RNase S, en date du 17 octobre 2019). Qu'elles soient en présence ou en absence de ligand, ce sont toutes des conformations accessibles à cette protéine, potentiellement nécessaires à sa fonction. Une description exhaustive de la structure de la RNase A devrait donc pouvoir mentionner chacune de ces 192 conformations connues, et laisser la place à de potentielles conformations inconnues. Une telle description exhaustive peut être représentée par ce qu'on appelle le paysage conformationnel de la protéine. Il est à noter que les équations présentées dans cette section sont utilisées principalement à titre théorique et informatif, elles ne sont pas nécessairement utilisées pour des calculs explicites.

1.4.1 Généralités sur un paysage conformationnel

Le paysage conformationnel (« conformational landscape » ou « energy landscape, » à ne pas confondre avec le « fitness landscape » (Romero et Arnold 2009)) d'une protéine peut être décrit par un graphique de l'énergie libre de Gibbs (ΔG) du système en fonction de différents degrés de liberté (Dill et Bromberg 2011). Lorsque celle-ci est négative, la structure sera stable, car ΔG représente l'énergie nécessaire pour former une structure à partir d'un état entièrement dénaturé, dont l'énergie est définie à zéro (Dill et Bromberg 2011). Déjà, ceci montre l'intérêt de l'utilisation du paysage conformationnel dans les calculs de repliement de protéines. Fait intéressant, la compagnie Google a récemment consacré leur intelligence artificielle DeepMind à la prédiction de structures de protéines pour la compétition Critical Assessment of Structure Prediction (Sample 2018). Elle a utilisé l'algorithme du gradient (« gradient descent algorithm »), qui fonctionne conceptuellement en déplaçant la structure à travers le paysage énergétique de manière à toujours faire descendre son ΔG, donc à trouver le minimum du paysage conformationnel.

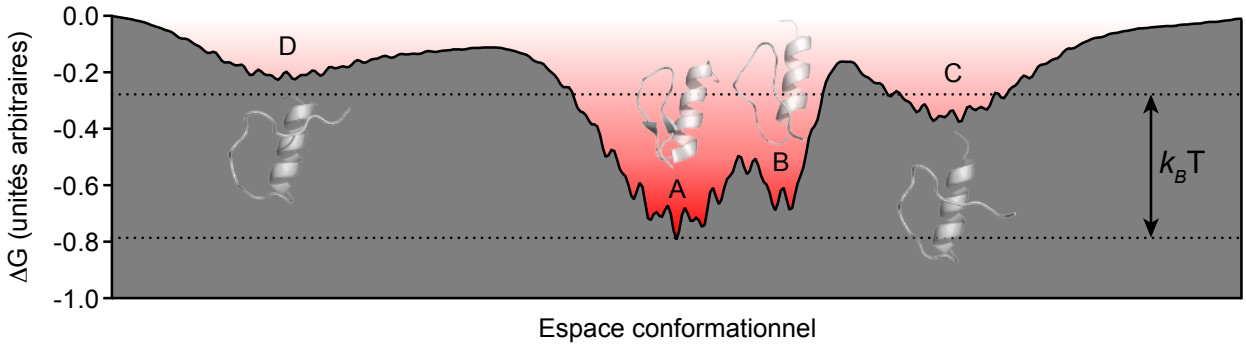


Figure 1.7 Exemple de paysage conformationnel

La protéine fictive décrite par le paysage conformationnel ci-dessus peut adopter quatre conformations « majeures, » représentées par les lettres A, B, C et D, en plus de diverses fluctuations autour de celles-ci. Le gradient rouge correspond à la population d'un état particulier, plus un état est rouge foncé, plus sa population sera élevée. Les exemples de structures proviennent des PDB 2moe et 2mof.

La figure 1.7 montre un paysage conformationnel fictif à une seule dimension, donc en fonction d'un seul degré de liberté. De possibles degrés de liberté pourraient être l'angle de torsion d'un lien, la distance entre deux atomes particuliers, la projection sur une composante principale de « principal component analysis » (PCA) ou sur un mode propre (« normal mode analysis »), ou toute autre façon de décrire une structure (Maria-Solano *et al*. 2018). Chaque puits correspond à une conformation, et la largeur d'un puits représente des fluctuations ou des vibrations de ce degré de liberté autour de l'état correspondant à ce puits (Dill et Bromberg 2011; Maria-Solano *et al*. 2018). La protéine peut en tout temps adopter n'importe laquelle de ces conformations (Dill et Bromberg 2011; Ramanathan *et al*. 2014; Agarwal *et al*. 2016), et la population de chaque état (de l'état *i*) peut être déterminée selon la thermodynamique par la distribution de Boltzmann (Henzler-Wildman et Kern 2007; Dill et Bromberg 2011) :

Équation I

$$p_i = \frac{e^{-\frac{\Delta G_i}{k_B T}}}{\sum_i e^{-\frac{\Delta G_i}{k_B T}}}$$

Cette équation montre que les conformations ayant une énergie libre de Gibbs inférieure à l'énergie thermique k_BT, telles les conformations A, B et C de la figure 1.7, seront plus facilement accessibles que les conformations plus énergétiques, comme la conformation D. Un paysage conformationnel particulier, et surtout la population de chacun des états, est unique à chaque séquence primaire et variera en fonction de la température, de la pression, de la

présence ou l'absence de ligands, ainsi que de l'environnement dans lequel se retrouve la protéine (Maria-Solano *et al*. 2018).

Il est important de rappeler, toutefois, que l'exemple de la figure 1.7 présente un paysage énergétique à un seul degré de liberté. Il s'agit là d'une simplification, la majorité des paysages conformationnels étant multidimensionnels (Agarwal *et al*. 2016). En ne comptant par exemple que les distances entre C_α, une protéine de 100 acides aminés compterait 4950 degrés de liberté.

Des barrières de potentiel (représentées par les bosses sur la figure 1.7) séparent les différents états, et leur amplitude détermine la facilité avec laquelle une protéine peut passer d'un état à l'autre – autrement dit, l'échelle de temps sur laquelle la transition se produira, qui est donnée par l'équation d'Arrhenius (Henzler-Wildman et Kern 2007; Dill et Bromberg 2011; Maria-Solano *et al*. 2018) :

Équation II

$$k_{ij} = Ae^{-\frac{\Delta\Delta G_{ij}^{bar}}{k_B T}}$$

où k_{ij} est la constante de vitesse entre l'état *i* et l'état *j*, *A* est une constante propre à la réaction, et $\Delta\Delta G_{ij}^{bar}$ est la différence énergétique entre l'état *i* et le sommet de la barrière énergétique le séparant de l'état *j* (voir figure 1.8). Selon cette équation, passer de l'état A à l'état B et inversement dans l'exemple de la figure 1.7 implique de surmonter une barrière de potentiel inférieure à l'énergie thermique k_BT, donc la protéine pourra alterner très rapidement

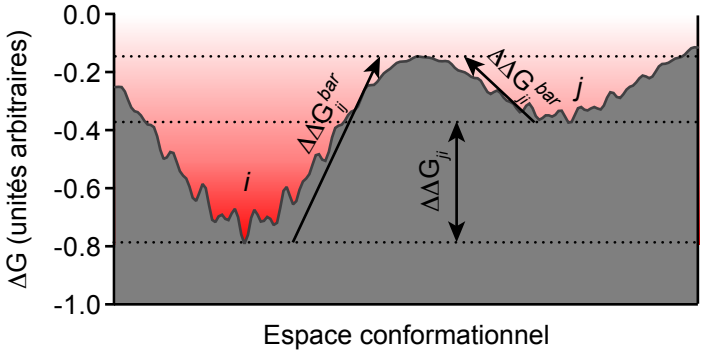


Figure 1.8 Énergies impliquées dans une barrière de potentiel

Le paysage conformationnel de cette protéine fictive montre deux états majeurs *i* et *j* séparés par une barrière de potentiel. Les deux états ont une différence énergétique de $\Delta\Delta G_{ji}$, mais une énergie supplémentaire de $\Delta\Delta G_{ji}^{bar}$ doit être fournie à l'état *i* pour lui permettre de transiter vers l'état *j*. Le gradient rouge correspond ici aussi à la population des états.

(potentiellement dans l'ordre de la nanoseconde) entre ces deux conformations. Atteindre les états C et D à partir de l'état de base A implique une barrière de potentiel supérieure à k_BT, et donc ces transitions seront beaucoup plus lentes (possiblement de l'ordre de la milliseconde ou plus lentes). *A contrario*, retourner à l'état de base A à partir de l'état D sera facile, puisque la différence énergétique entre le minimum de l'état D et le maximum le séparant de l'état A est faible, et donc le temps de vie d'une protéine dans l'état D sera court (Maria-Solano *et al*. 2018).

1.4.2 Conformations et fonctions des protéines

Quelle est l'utilité pour une protéine de pouvoir adopter d'autres conformations plus énergétiques et transitoires? En fait, plusieurs réponses existent à cette question (Ramanathan *et al*. 2014), et celles présentées ici n'en sont pas une liste exhaustive. En reprenant l'exemple de la figure 1.7, la conformation A pourrait correspondre à l'état de base, et afin d'effectuer sa fonction principale sur son ligand, la protéine devrait alterner avec l'état B. Elle aurait la possibilité de se lier à un autre ligand en adoptant la conformation C, ou alors d'effectuer une deuxième fonction en adoptant l'état D. Bien que cette protéine soit fictive, de nombreux exemples plus concrets peuvent illustrer l'utilité de la diversité des conformations accessibles à une protéine.

Il arrive fréquemment que les différentes conformations soient critiques à la fonction. Chez plusieurs protéines, les multiples états accessibles serviront à promouvoir différentes sous-étapes de leur fonction. Dans le cas de l'enzyme dihydrofolate réductase de *Escherichia coli*, la protéine adopte séquentiellement cinq conformations différentes au cours de son cycle de renouvellement catalytique avec un taux de transition spécifique à chaque étape, variant de 1200 s^{-1} à 12 s^{-1} (Boehr *et al*. 2006). Dans ce cas particulier, la présence ou l'absence du cofacteur, du substrat ou du produit modifie les populations des cinq états, et donc l'allure du paysage conformationnel, au fil du cycle (Boehr *et al*. 2006; Agarwal *et al*. 2016).

Pour d'autres protéines, les différents états permettront la reconnaissance de ligands. À titre d'exemples, la protéine STARD6 doit adopter une conformation où une boucle est ouverte pour permettre l'entrée de stéroïdes à l'intérieur d'une cavité hydrophobe, puis retrouver la conformation fermée afin de transporter ces molécules hydrophobes à l'intérieur du cytoplasme (Létourneau *et al*. 2016). L'adénylate kinase de *E. coli* adopte elle aussi une conformation fermée afin de faciliter la liaison de son substrat. Cette liaison causera par la suite une augmentation de la population de cet état, permettant à la catalyse enzymatique de s'effectuer (Kovermann *et al*. 2017). La conformation de base de l'haloalkane déhalogénase DhlA de

Xanthobacter autotrophicus permet la liaison (et l'hydrolyse subséquente) du 1,2-dichloroéthane, mais la transition à une conformation ouverte permet la catalyse de substrats beaucoup plus volumineux que ce qui serait permis par la taille de la cavité (Kokkonen *et al*. 2018).

Il est également possible que les différents états excités d'une protéine puissent lui permettre de carrément effectuer des fonctions accessoires à leur fonction principale. Une phosphotriestérase de *Pseudomonas diminuta* possède par exemple un état excité lui conférant une activité arylestérase, bien que l'efficacité catalytique (k_{cat}/K_M) de cette fonction accessoire soit plus faible de quatre ordres de grandeur que celle de la fonction primaire (Campbell *et al*. 2016). Chez les récepteurs couplés aux protéines G, la pluralité des états actifs permet la sélectivité fonctionnelle (le « biased signalling »). Le récepteur AT1 de l'angiotensine-II, par exemple, a la possibilité d'adopter des conformations permettant l'activation des voies de signalisation associées à $G_{q/11}$ ou alors aux β-arrestines. L'intensité de la réponse dans l'une de ces voies dépendra des populations des états correspondants, et ces populations dépendent à leur tour du ligand reconnu par le récepteur (Cabana *et al*. 2015).

La connaissance du paysage énergétique d'une protéine peut être extrêmement utile en ingénierie des protéines. Encore une fois, les exemples énumérés ici ne forment pas une liste exhaustive, mais ont pour but de donner une idée de l'importance du concept. Tel que mentionné plus tôt, le paysage conformationnel est unique à chaque séquence de protéine. Il est donc possible de l'altérer grâce à des mutations, et potentiellement de modifier ce paysage jusqu'à l'obtention de l'effet désiré (Chen *et al*. 2014; Campbell *et al*. 2018; Maria-Solano *et al*. 2018). Des exemples de modifications possibles de paysages énergétiques sont montrés à la figure 1.9.

Le premier exemple de modification du paysage conformationnel qui vient à l'esprit est d'augmenter l'efficacité énergétique d'une enzyme en augmentant la population de la conformation critique à la réaction, ou alors en diminuant la barrière énergétique responsable de la transition limitante. Une telle altération du paysage énergétique via des mutations distales au site actif déterminées par des prédictions bioinformatiques a permis d'augmenter significativement l'efficacité d'une rétro-aldolase, augmentant son k_{cat} d'un facteur 4400 (Romero-Rivera *et al*. 2017).

Dans le cas de protéines possédant une conformation permettant une activité secondaire, il peut être désirable de modifier la protéine de manière à rendre majoritaire cette dernière

1.4.3 Paysages conformationnels et ingénierie des protéines

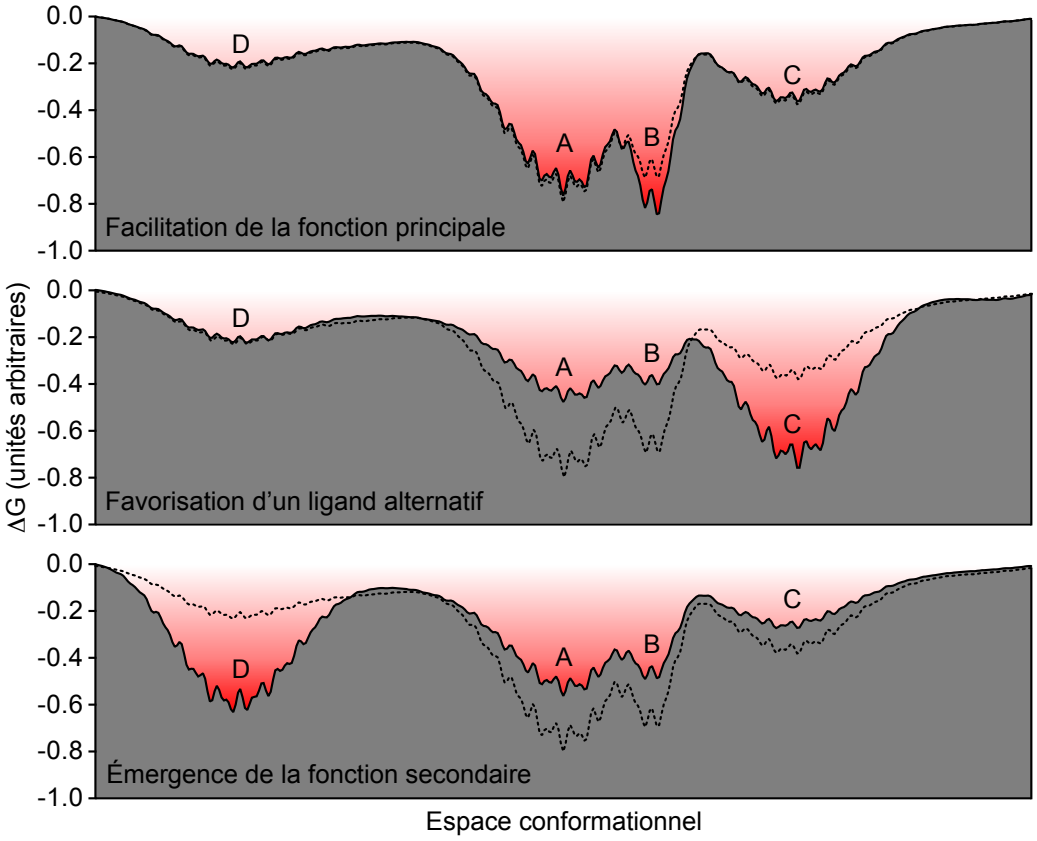


Figure 1.9 Altérations potentielles du paysage conformationnel par ingénierie des protéines

Les trois paysages conformationnels ci-dessus sont des modifications de celui présenté à la figure 1.7, qui est reproduit en pointillés. Le gradient rouge montre encore une fois la population des états, avec une couleur plus foncée correspondant à une population plus élevée.

conformation. Une série de générations d'évolution dirigée a justement permis d'augmenter l'activité arylestérase de la phosphotriestérase de *P. diminuta*, de manière à ce que le k_{cat}/K_M relatif de chacune de ces deux activités soit augmenté de neuf ordres de grandeur par rapport à sa valeur initiale en faveur de l'activité arylestérase (Campbell *et al*. 2016). Un autre exemple est la modification d'une β-lactamase ancestrale qui a enrichi une conformation permettant une activité d'éliminase de Kemp. Cette β-lactamase ancestrale avait la particularité d'être sujette à des échanges conformationnels, contrairement aux β-lactamases modernes, et implanter les mêmes mutations chez la protéine moderne n'a pu générer d'activité éliminase de Kemp (Risso *et al*. 2017), montrant l'importance de bien connaître le paysage conformationnel d'une enzyme avant de procéder à une ingénierie de sa fonction.

1.4.4 Caractérisation des paysages conformationnels

L'étude du paysage conformationnel d'une protéine peut instruire grandement sur la (ou les) fonction(s) de celle-ci. Il existe de nombreuses techniques bioinformatiques permettant d'évaluer ce paysage (plusieurs sont décrites dans Maria-Solano *et al*. 2018), mais considérant le degré de complexité de ces molécules que sont les protéines, il est important de prendre en compte que ces paysages ne seront que partiels. D'un point de vue expérimental, par contre, il est beaucoup plus complexe d'étudier l'ensemble des structures accessibles à une protéine. Tel que mentionné plus tôt, les conformations plus énergétiques sont pour la très grande majorité transitoires, avec un temps de vie fréquemment inférieur à la seconde. La population de ces états excités sera donc faible, et par conséquent, ils seront difficiles à observer par les techniques actuelles qui reposent sur les valeurs/structures moyennes dans un ensemble de population. Dans un avenir potentiellement rapproché, la cryo-microscopie électronique pourra vraisemblablement déterminer l'ensemble des conformations accessibles à une protéine d'un échantillon donné de par sa capacité à étudier les molécules individuelles (« single-particle analysis ») (Zhang *et al*. 2019), mais pour l'instant cette technique est encore limitée aux complexes protéiques massifs. À l'heure actuelle, il est plus aisé d'étudier les phénomènes d'échanges qui se produisent entre les conformations. Ceux-ci peuvent renseigner sur les barrières énergétiques séparant l'état de base et les états excités malgré l'ignorance de la structure des différents états impliqués, ainsi que sur leur population. Les constantes de vitesse données par l'équation II sont difficiles à extraire, mais il est possible de déterminer la constante d'échange k_{ex}, qui dépend de la barrière énergétique ainsi que de la différence énergétique entre les deux états :

Équation III

$$k_{ex} \propto e^{-\frac{\Delta\Delta G_{ji}^{bar}}{k_B T}} \left(e^{-\frac{\Delta\Delta G_{ji}}{k_B T}} + C \right)$$

où C est une constante. La section 7.1 en annexe présente la preuve mathématique de cette équation. L'article de revue « Role of conformational motions in enzyme function : selected methodologies and case studies » (Narayanan, Bernard et Doucet 2016), présenté au chapitre 2, discute plus en détails de différentes manières d'étudier ces échanges conformationnels.

2 PREMIER ARTICLE

Les conformations transitoires qu'adoptent les protéines sont difficilement observables, et les techniques permettant de les étudier donnent des informations précises, mais souvent insuffisantes à tracer un portrait global. La sélection de la technique appropriée pour aller chercher l'information nécessaire devient donc d'autant plus cruciale dans ces circonstances. Les façons les plus appropriées d'étudier les conformations transitoires sont de les observer par des méthodes bioinformatiques, ou alors d'en détecter la présence expérimentalement via l'étude des changements conformationnels. L'article de revue présenté dans ce chapitre résume différentes techniques utilisées pour étudier les mouvements internes et les échanges conformationnels chez les protéines. De plus, il approfondit le rôle de ces échanges conformationnels chez deux exemples de systèmes enzymatiques, et ce qui arrive lorsque ceux-ci sont modifiés suite à des mutations de ces protéines. Cet article constitue donc une excellente entrée en matière pour mon projet de recherche. Sa référence complète est la suivante : **C Narayanan*, DN Bernard* et N Doucet (2016) Role of conformational motions in enzyme function : selected methodologies and case studies. Catalysts 6, 81**.

L'article a été publié dans l'édition de juin 2016 de la revue Catalysts. Chitra Narayanan et moi-même en sommes co-premiers auteurs, et nous avons contribué également à sa rédaction. Je me suis concentré sur la description des différentes techniques expérimentales ainsi que sur la revue des connaissances sur les échanges conformationnels chez la protéase du VIH-1, alors que Chitra Narayanan s'est occupée des sections sur la bioinformatique et la polymérase β. Nicolas Doucet a révisé et supervisé le processus de rédaction.

2.1 Résumé

Il est maintenant couramment accepté par la communauté scientifique que les enzymes sont des entités mobiles qui s'appuient sur des mouvements complexes à l'échelle atomique ainsi que sur des échanges conformationnels coordonnés pour effectuer correctement la liaison du substrat et la catalyse. Par contre, le rôle exact de ces mouvements internes dans la fonction de ces enzymes demeure soit mal compris, soit difficile à interpréter. Ce court article de revue tente de réconcilier les observations biophysiques et les fonctionnalités biologiques d'abord en décrivant un certain nombre de méthodologies expérimentales et computationnelles communes qui sont employées pour caractériser ces mouvements de résidus à l'échelle atomique sur plusieurs échelles de temps chez les enzymes, et ensuite en présentant comment la

connaissance de ces mouvements peut être utilisée pour décrire le comportement fonctionnel d'enzymes et même y agir. Deux exemples biologiques pertinents seront soulignés, c'est-à-dire les systèmes enzymatiques de la protéase du VIH-1 et de l'ADN polymérase β.

2.2 Abstract

It is now common knowledge that enzymes are mobile entities relying on complex atomic-scale dynamics and coordinated conformational events for proper ligand recognition and catalysis. However, the exact role of protein dynamics in enzyme function remains either poorly understood or difficult to interpret. This mini-review intends to reconcile biophysical observations and biological significance by first describing a number of common experimental and computational methodologies employed to characterize atomic-scale residue motions on various timescales in enzymes, and second by illustrating how the knowledge of these motions can be used to describe the functional behavior of enzymes and even act upon it. Two biologically relevant examples will be highlighted, namely the HIV-1 protease and DNA polymerase β enzyme systems.

2.3 Introduction

Proteins and enzymes are essential components of living cells. Among a variety of other functions, they act as hormones in cell signaling, protein transporters, antibodies in host defense, etc. Concomitantly, enzymes are some of the most efficient catalysts known to date, enhancing biochemical reaction rates up to 19 orders of magnitude relative to uncatalyzed reactions (Garcia-Viloca *et al*. 2004; Wolfenden 2006). However, factors contributing to these large rate enhancements in enzyme-catalyzed reactions remain largely uncharacterized. The original structure-function paradigm, popular for many decades, portrayed enzymes as rigid structures with shapes that facilitate substrates, ligand and/or cofactor binding. This original model evolved over time, and theories were developed that led to the now-accepted induced fit, conformational selection, and transition-state stabilization models to explain the behavior of protein-ligand recognition and catalysis in the molecular function of enzymes (Hammes *et al*. 2009). Increasing evidence suggests that proteins sample a variety of distinct conformations (or sub-states) enabled by concerted atomic-scale dynamical fluctuations occurring over a wide range of timescales and acting on the primary, secondary, tertiary, and quaternary organization of their molecular structure (Henzler-Wildman and Kern, 2007). Conformational transitions

between highly and rarely populated states have been shown to play important roles in substrate recognition, binding, and product release, among others (Venkitakrishnan *et al*. 2004; Watt *et al*. 2007; Boehr 2012; Berlow *et al*. 2012). Advances in experimental and computational methodologies continue to offer new insights into enzyme conformational motions over functionally relevant timescales (Kay 2005; Vallurupalli *et al*. 2008; Vanwart *et al*. 2012; Ramanathan *et al*. 2014). Interestingly, experimental (Eisenmesser *et al*. 2005; Gagné *et al*. 2012), computational (Agarwal *et al*. 2004; Goodey and Benkovic 2008), and sequence-based (Suel *et al*. 2003; Reynolds *et al*. 2011) approaches have also revealed functional networks of concerted residue motions distant from the active site in selected enzyme systems. Correlations between the timescale of conformational fluctuations and that of catalytic turnover have been well established in a variety of enzyme systems, including cyclophilin A (Eisenmesser *et al*. 2005), RNase A (Cole and Loria 2002; Watt *et al*. 2007; Gagné and Doucet 2013), triosephosphate isomerase (Pareek *et al*. 2016), and HIV-1 protease (Torbeev *et al*. 2011), among others. Further, the rate of conformational exchange has been shown to coincide with the rate-limiting step, such as product release, in some of these systems (Cole and Loria 2002; Gagné and Doucet 2013). Conformational exchange between sub-states allows enzymes to sample higher energy conformations with structural and dynamical properties important for function such as ligand binding and allosteric regulation (Goodey and Benkovic 2008; Kar *et al*. 2010; Ramanathan *et al*. 2014). In addition to the millisecond conformational exchange, dynamics on faster timescales modulate the chemical environment through rearrangements in the active site, thus affecting enzyme function (Agarwal 2006). Taken together, a view is emerging whereby conformational fluctuations occurring over a range of timescales can affect enzyme function through conformational sampling along preferred pathways.

Over the past few years, a number of controversial statements relating ‘catalysis’ with ‘dynamics’ have been published in the literature, often giving rise to heated debates between experimentalists and theoreticians (Kamerlin and Warshel 2010; Kohen 2015). As recently outlined, these debates are often semantic in nature and can be traced back to actual definitions (and research field perceptions) of what ‘protein dynamics’ represents, in addition to which atomic-scale events are being observed and/or over which timescales they occur during enzyme-catalyzed reactions (Kohen 2015). In the current report, we consider motions affecting enzyme function in a broad sense, i.e., “any motion that, if impeded, would reduce the ability of the enzyme to function by the mechanism that it has evolved to execute” (Nagel and Klinman 2009). As such, we take into account motions on a wide range of timescales, from the fast, local picosecond-nanosecond (ps-ns) fluctuations of individual residues to the slower microsecond-

millisecond-second (μs-ms-s) global conformational exchange experienced by enzymes. Conserved motional events occurring on a broad range of timescales could be involved in many and/or a selected subset of events occurring along the catalytic cycle of an enzyme, including ligand binding, recognition, substrate discrimination, structural rearrangements leading to the transition state, product release, etc. (Kohen 2015). To clarify and prevent confusion, we never imply that 'motions contributing to catalysis' exclusively describe the fast femtosecond atomic motions involved in transition-state chemistry.

In the current account, we draw attention to a number of experimental and computational methodologies that have recently improved our understanding of catalytic and functional properties in two enzyme systems that closely rely on conformational dynamics for proper biological function. We do not pretend to cover the overwhelming number of theoretical approaches, methodologies, or enzymatic systems that previously illustrated the role of protein dynamics in enzyme function, but instead focus on recent reports where the combination of experimental NMR and computational techniques have emphasized the role of conformational exchange in inhibitor binding (HIV-1 protease) and fidelity in DNA repair through the selection of correct nucleotides (DNA polymerase β).

2.4 **Selected Experimental and Computational Approaches for Sampling Conformational Motions in Proteins**

A variety of approaches, including (but not limited to) NMR spectroscopy, X-ray diffraction, single-molecule FRET, and computational simulations, have been used to probe conformational dynamics in proteins over nanosecond to millisecond timescales, offering the means to extract motions potentially relevant to biological function (Henzler-Wildman and Kern 2007; Hanoian *et al*. 2015). One of the most commonly used experiments to characterize local and global conformational events experienced by enzymes on the timescale of their catalytic rate k_{cat} — typically occurring on the order of milliseconds — is the NMR Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion experiment (Loria *et al*. 1999a; Loria *et al*. 1999b; Kleckner and Foster 2011; Palmer 2014; Grutsch *et al*. 2016). This methodology garnered considerable attention in recent years since motions experienced by enzymes on this physiologically relevant timescale can improve our understanding of catalysis and allostery, as the flexibility events uncovered can be analyzed in light of catalytic events. The CPMG method enables the detection of atomic-scale millisecond dynamics of enzymes in solution, either in the free form or bound to biologically significant ligands/analogues that mimic intermediate states along the

chemical reaction (Gagné *et al*. 2012; Bhabha *et al*. 2015; Oyen *et al*. 2015). On this time frame, molecular flexibility may involve the rearrangement of single residues and/or entire secondary structures moving in a coordinated fashion (for reviews see Palmer (2001); Palmer (2004); Kay (2005); Doucet (2011)) (Figure 2.1). In its typical form, the method involves the characterization of an equilibrium exchange process experienced by the N–H amide bond of amino acid residues, performed by recording the transverse relaxation rate constant (R_2) as a function of ν_{CPMG}, the frequency with which refocusing pulses are applied in a CPMG pulse train (Manley and Loria 2012; Palmer 2014). These refocusing pulses are effectively invisible to amino acid residues that do not undergo conformational exchange in the protein, but they add an additional contribution (R_{ex}) to R_2 for exchanging residues — those experiencing motions on this millisecond time frame — and thus causing a decrease in NMR resonance intensity. Plotting this decrease in signal intensity as a function of ν_{CPMG} not only allows the qualitative identification of residues experiencing conformational exchange in the enzyme, but fitting the curves to relaxation dispersion equations also allows the quantitative extraction of structural information related to the excited (less populated) sub-states (differences in chemical shifts, Δω), in addition to their population dynamics in solution (conformational exchange rates, k_{ex}, and equilibrium populations, p_A and p_B) (see Manley and Loria (2012) and references therein).

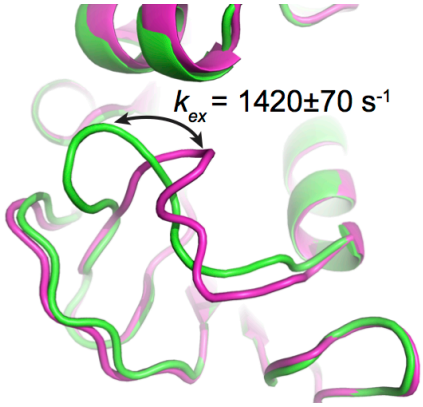


Figure 2.1 Conformational exchange experienced by the Met20 loop in *E. coli* DHFR, as probed by ^{15}N-CPMG NMR relaxation dispersion experiments.

The CPMG experiment is particularly well suited to extract exchange rates (k_{ex}) between two or more conformations in solution on the timescale of catalysis (ms) in many enzyme systems, offering a measure of comparison between conformational exchange experienced by the enzyme and its catalytic rate (k_{cat}). This method can also provide quantitative information on low-populated "invisible" excited sub-states and their populations in solution (p_A and p_B). The cartoon representation illustrates a simplified view of the closed (green) and occluded (magenta) conformations sampled by the Met20 loop in *E. coli* DHFR as it catalyzes hydride transfer, an atomic-scale movement essential to bacterial DHFR function and correlated with substrate/cofactor recognition and turnover in this enzyme (reviewed in Bhabha *et al*. (2015); Hanoian *et al*. (2015); and Luk *et al*. (2015)). Depicted PDB structures are 1RX2 (green) and 1RX7 (magenta).

With regards to the timescale of protein motions, NMR is not limited to the µs-ms range. To this day, the model-free formalism developed by Lipari and Szabo in the early 1980s remains one of the preferred approaches to determine the atomic-scale dynamics of a protein in solution (Lipari and Szabo 1982a; Lipari and Szabo 1982b). Using spin relaxation NMR data, the method primarily provides information on residue dynamics occurring on the ps-ns timescale, which are most often attributed to single residue side-chain rotations or local residue motions. The method measures spin relaxation rates for each residue amide N–H pair to extract order parameters (S^2), conformational exchange parameters (R_{ex}), and the correlation time (τ_c), offering an interpretation of the internal amplitude of the N–H bond vector in solution (Mandel *et al*. 1995; Morin 2011; Fisette *et al*. 2012). The order parameter S^2 provides an easily interpretable measure of the rigidity of each protein residue, where the S^2 of a completely unrestricted N–H bond would theoretically correspond to a value of 0, while that of a fully rigid residue would be 1. For a number of residues that do not fit to simple calculation models, an R_{ex} exchange parameter can be estimated to achieve an adequate fit to the model-free analysis (Mandel *et al*. 1995).

While the aforementioned NMR relaxation analyses provide detailed information on the molecular flexibility experienced by enzymes in solution, other experimental methodologies offer additional details pertaining to potentially significant biological motions. Single-molecule Förster Resonance Energy Transfer (single-molecule FRET) (Henzler-Wildman and Kern 2007; Stephenson *et al*. 2016) and site-directed spin labeling Double Electron-Electron Resonance (DEER) spectroscopy (Huang *et al*. 2012) — also called pulsed Electron Paramagnetic Resonance (pulsed-EPR) — are both used to determine the distance between two labels, ranging up to a few tens of Å. In DEER, a pair of cysteine residues is chemically modified with electronic spin labels, namely the nitroxide radical of a methylthiosulfonate moiety (Torbeev *et al*. 2011; Huang *et al*. 2014). The magnitude of the dipolar coupling between the two spins can be extracted from the measured DEER echo curves, and this magnitude is inversely proportional to the cube of the distance between the two spins (Jeschke and Polyhach 2007), thus producing a distance profile. Populations of the different states of the protein can then be evaluated by deconvoluting the distance profile. Much like with protein crystallography, a disadvantage of the technique is the requirement of cryogenic temperatures (Huang *et al*. 2012), which does not provide an accurate description of biological conditions. On the other hand, single-molecule FRET can be used in solution and at physiologically relevant temperatures, much like NMR. This technique also requires labeling of two distinct sites on a protein, this time using two different fluorophores. Since these fluorophores are selected so that

the absorption spectrum of the first (the acceptor) matches the emission spectrum of the second (the donor), the fluorescence energy can be transmitted from the donor to the acceptor if they are in close proximity to each other. Thus, the emission wavelength of the donor will disappear and the emission wavelength of the acceptor will be detected as a function of distance. By measuring the FRET efficiency, the inter-fluorophore distance can be calculated. The single-molecule version of this experiment can thus allow one to witness whole domain reorientations in a protein. This technique also has the advantage of being performed in real-time (Henzler-Wildman and Kern 2007; Stephenson *et al*. 2016).

While most experimental methodologies yield quantitative information on flexibility events, they often fail to provide a clear visual representation of the actual atomic-scale motions experienced by an enzyme in solution (Fisette *et al*. 2012). In contrast, computational molecular dynamics (MD) simulations trace the position and interaction of atoms over time based on classical mechanics principles, to provide an atomic-scale time evolution of conformational properties in proteins and other biomolecules. Until recently, MD simulations typically sampled over hundreds of nanoseconds, while protein motions span a much wider range of timescales (Henzler-Wildman and Kern 2007), with many catalytically relevant motions occurring on the same timescale as the catalytic constant k_{cat}, i.e., typically milliseconds or slower (Cole and Loria 2002; Boehr *et al*. 2010; Boehr *et al*. 2014; Kohen 2015). This presents an existing limitation of computational MD simulations. Advances in software and hardware capabilities have facilitated access to longer timescales, with several recent studies reporting ms timescale simulations (Voelz *et al*. 2010; Bowman *et al*. 2011; Voelz *et al*. 2012). While a single ms trajectory has been achieved using Anton (Shaw *et al*. 2010), a specialized supercomputer for protein simulations, access to the ms timescale with traditional, unbiased MD simulations still remains limited.

A variety of enhanced sampling techniques such as accelerated MD (aMD) (Hamelberg *et al*. 2004; Bucher *et al*. 2011), umbrella sampling (Wojtas-Niziurski *et al*. 2013), steered MD (Da *et al*. 2012), and metadynamics (Tagle 1998; Sutto and Gervasio 2013; Abrams and Bussi 2014), among others, have been developed to improve sampling of the conformational landscape of proteins. In aMD, the potential energy landscape is modified by raising energy minima that are below a threshold, therefore minimizing the energy barriers separating states, and facilitating access to conformational states not easily accessible through traditional MD (Hamelberg *et al*. 2004). This approach has been used to characterize dynamics on the μs-ms timescale in a variety of protein systems (Doshi and Hamelberg 2015). A combination of normal and

accelerated MD simulations was recently used to characterize the role of conformational dynamics in stabilizing the transition state in the enzyme cyclophilin A (Doshi *et al*. 2012). Umbrella sampling uses a biasing potential to characterize the higher energy states along a reaction coordinate (Torrie and Valleau 1977). Metadynamics involves the use of a history-dependent biasing potential that acts on collective variables corresponding to select degrees of freedom (Tagle 1998). A combination of NMR and metadynamics was recently used to characterize the conformational ensemble and the free energy landscape of the highly dynamic helix 1 of the prion protein (Camilloni *et al*. 2012b). While these approaches have clear benefits from enhanced sampling, the requirement of the knowledge of start and end structures (umbrella sampling), or the selection of suitable reaction coordinates (metadynamics), among others, limit their application (Andersen *et al*. 2015).

Other advanced sampling methods, such as the Markov State Model (MSM, also known as the kinetic network model) (Noé and Fischer 2008; Bowman *et al*. 2009; Singhal *et al*. 2004) and Quasi Anharmonic Analysis (QAA) (Ramanathan and Agarwal 2011; Ramanathan *et al*. 2014), facilitate the identification of conformational sub-states, equilibrium populations of sub-states, and transition probabilities between these sub-states. MSMs are constructed from trajectories obtained from MD simulations. In these models, the conformational space is discretized into multiple states, and a network connecting these states is created from the MD simulations. A transition matrix corresponding to the probability of transitions between different states is constructed to characterize the kinetics of transition between states. MSMs have been applied to sample a wide range of timescales associated with protein folding, functional dynamics, ligand binding and for characterizing intrinsically disordered proteins (Morkos *et al*. 2010; Chodera and Noé 2014; Shukla *et al*. 2015). QAA uses higher order statistics to describe positional fluctuations and the coupling of these fluctuations in different regions of the protein (Ramanathan and Agarwal 2011). It allows the characterization of conformational sub-states along a reaction pathway and the identification of sub-states with structural and dynamical properties important for function (Ramanathan *et al*. 2011). This approach was used to identify the hierarchical organization of conformational sub-states in ubiquitin and the T4 lysozyme, and the conformational sub-states associated with the cis/trans isomerization catalyzed by cyclophilin A were also characterized (Ramanathan *et al*. 2011). This technique was recently used to characterize conformational sub-states associated with the interconversion between reactant and product states in wild-type and mutant forms of RNase A (Gagné *et al*. 2015b). Recent years have also seen advances in the integration of experimental and computational approaches for characterizing the dynamics of proteins. NMR chemical shifts and residual

dipolar couplings have been used as replica-averaged structural restraints to characterize the conformational fluctuations in RNase A and the hen lysozyme, respectively (Camilloni *et al*. 2012a; Camilloni *et al*. 2013; De Simone *et al*. 2013).

2.5 Case Studies

2.5.1 HIV Protease

The protease of the type 1 human immunodeficiency virus (HIV-1 PR) is an aspartyl proteinase that cleaves the polyprotein product of the *gag-pro-pol* gene and the Gag protein itself, both of which are vital steps in the maturation of the virus (Sundquist and Krausslich 2012). The three most prevalent subtypes of the virus are those whose proteases have been the most studied: subtypes B, C, and CRF_01 A/E, which are the prevalent HIV subtypes in North America and Europe, Africa, and Asia, respectively. Even though all three subtypes have a distinct wild-type (WT) protease, subtype B is the most widely characterized (Huang *et al*. 2012). To date, nine Food and Drug Administration (FDA)-approved inhibitors targeting this protein have been used to treat HIV/acquired immune deficiency syndrome (AIDS)-infected patients. However, multidrug resistance to protease inhibitors has sparked renewed interests in uncovering distinct inhibitors targeting HIV-1 PR, and thus understanding drug resistance acquisition mechanisms is of utmost importance (Ghosh *et al*. 2016).

HIV-1 PR is only active in its homodimeric form. The 99-amino-acid protomers assemble in a symmetric configuration, forming a tunnel. The protomers each provide one of the catalytic aspartate residues (Asp25, located in the tunnel), while access to the catalytic site is allowed through the opening and closing of a pair of β-hairpins, termed “flaps” (residues 43–58) (colored red in Figure 2.2). The protomers interact with each other through their N- and C-termini, as well as at the tip of the flaps (Spinelli *et al*. 1991). A previous report briefly discusses the atomic-scale dynamics experienced by WT HIV-1 PR, describing how movements of the flaps originating from the isomerization of Gly51 are transmitted to the active site (Pastor and Amero 2015). This vital conformational transition, identified as early as 1995, involves the 180° reorientation of the Gly51 backbone N–H vector, exchanging between β-turn types I and II, a process taking place over a 10-μs timescale, as determined from R_{ex} values extracted using the Lipari-Szabo model-free formalism mentioned above (Nicholson *et al*. 1995). This unique residue reorganization represents the sole significant structural difference between otherwise symmetrical protomers in the crystal structures. Most commonly, Gly51 is found in the *L*-

conformation in one protomer and in the *D*-conformation in the opposing protomer (Torbeev *et al*. 2011). Model-free analyses also confirmed that the flap region, along with the hinge loop (residues 34–42), are the primary regions in which sub-nanosecond timescale residue dynamics are localized (Ishima and Louis 2008).

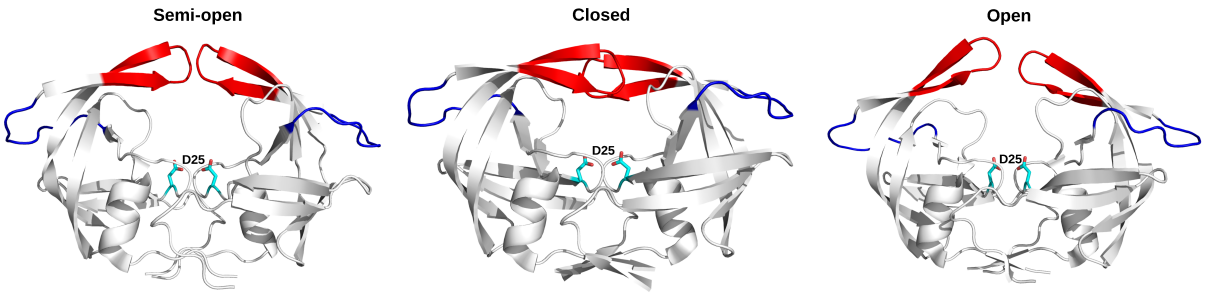


Figure 2.2 Conformational states experienced by HIV-1 PR as it proceeds through its catalytic cycle.

Crystal structures of HIV-1 PR in the semi-open (PDB 1HHR), closed (PDB 1HVR) and wide open (PDB 1TW7) conformations. The catalytic Asp25 residues from each protomer are depicted as cyan sticks and the "flaps" are depicted in red. The hinge loop, corresponding to residues 34–42, is shown in blue. The curled/tucked conformation of HIV-1 PR is not depicted, since it has never been crystallographically resolved.

NMR relaxation and computational MD simulations demonstrated that flaps can adopt a variety or conformations as the enzyme proceeds through its catalytic cycle (Ishima *et al*. 2003; Huang *et al*. 2012; Xia *et al*. 2013). These flaps alternate between a "closed" conformation, a "semi-open" conformation, a "wide open" conformation, and a "curled/tucked" conformation (Figure 2.2). The flaps close over the active site to maintain the substrate inside the catalytic pocket and to provide a gating mechanism for substrate binding, but they also interact with each other in the absence of substrate or inhibitor (Ishima and Louis 2008). A combination of DEER measurements, NMR relaxation experiments, and MD simulations provided information on Asp25 and flap dynamics in each protomer during catalytic turnover. The μs-ms timescale motions of Asp25 and Gly27 were found to correlate with those of the flaps (Torbeev *et al*. 2011). It was also observed that the more flexible *L*-Gly51 must be located on the same protomer as the protonated general acid Asp25-COOH for efficient catalysis to occur, while the more rigid *D*-Gly51 must be located on the protomer bearing the unprotonated general base Asp25-COO^-. The rate-limiting step of catalytic turnover was shown to be linked to isomerization of the flap tips towards the type I and type II β-turn configurations (Torbeev *et al*. 2011), further illustrating the functional importance of conformational exchange in this enzyme system.

The kinetic network model, constructed by discretizing the conformational space of atomistic simulations into a series of related sub-states, was used to determine the transitions between the semi-open, closed, and open conformations of HIV-1 PR (Deng *et al*. 2011). This approach facilitates the characterization of interconverting rates between kinetically relevant sub-states, and can predict kinetic quantities on slower (millisecond) timescales (Shukla *et al*. 2014; Shukla *et al*. 2015). Deng *et al*. (2011) combined replica exchange molecular dynamics simulations with transition path theory to characterize the diversity and temperature-dependence of kinetic pathways between the three conformational states of HIV-1 PR. They showed that transitions between semi-open and closed states occur on a fast timescale (~33 ns), while transitions between semi-open and open states were infrequent (~375 ns), consistent with NMR observations showing motions on the fast sub-nanosecond and slower microsecond timescales.

Not all studies agree on which conformation is the most likely populated form of the apo WT protease in solution. Studies using the DEER method concluded that the semi-open conformation is the most populated state, independent of the subtype (Torbeev *et al*. 2011; Huang *et al*. 2012), while NMR residual dipolar coupling (RDC) measurements suggested that the closed conformation was predominant (Roche *et al*. 2015). However, all studies agree that the populations of these conformers change upon inhibitor binding, with the most potent inhibitors favoring the “closed” conformation (Yamazaki *et al*. 1996; Huang *et al*. 2012; Leonis *et al*. 2013; Gupta *et al*. 2015). In all common subtypes of WT HIV-1 PR (subtypes B, C, and CRF_01 A/E), the residence time of inhibitors in the active-site pocket, and thus the inhibitory strength, is correlated with the degree to which said inhibitor induces flap closure. Consequently, many drug resistance-inducing mutations perturb the dynamic behavior of the flaps relative to the WT enzyme (Huang *et al*. 2012). Different inhibitors cause varying population shifts among the four possible flap conformations. According to DEER measurements, while inhibitors indinavir, nelfinavir, and atazanavir trigger very small differences in populations in all three subtypes—the semi-open conformation being the most populated — all other six FDA-approved inhibitors strongly shift the population equilibrium towards the closed configuration. Interestingly, the same six inhibitors also make the curled/tucked conformer much more likely in subtype CRF_01 A/E (Huang *et al*. 2012). While smaller in scope than those of the flaps, conformational changes in the hydrophobic core predicted by MD simulations were also confirmed to impact protease activity (Mittal *et al*. 2012). In the WT HIV-1 PR, the two halves of the hydrophobic core slide over each other during these rearrangements, and cross-linking these two halves — thus impeding hydrophobic core rearrangements — causes a 150-fold reduction in catalytic efficiency (Mittal *et al*. 2012; Goldfarb *et al*. 2015). Also, mutations

distal to the active site in the hydrophobic core are hypothesized to cause drug resistance by altering the flexibility of the protein (Mittal *et al*. 2012).

Previous reports have highlighted that many distinct mutations can arise in patients developing drug resistance to HIV treatment (Lee and Goodey 2011). Some of these mutations are located in the active-site pocket and directly impair inhibitor binding (primary mutations), while others are more distal and indirectly induce resistance (secondary mutations) (Figure 2.3). None of these distal mutations cause significant changes in either k_{cat} or K_M in HIV-1 PR, suggesting that conformational events experienced by the enzyme play a significant role in resistance acquisition, which was confirmed by NMR, MD simulations, and DEER spectroscopy. Mutations in 45 of the 99 residues in HIV-1 PR have been linked to drug resistance in vitro (Wu *et al*. 2003), and a quarter of the residues have been linked to drug resistance in patients, arising through evolutionary selective pressure driven by protease inhibitor therapy. However, mutations at only seven sites can be linked to direct active-site modifications (Figure 2.3a) (Wu *et al*. 2003; Leonis *et al*. 2013; Ragland *et al*. 2014). These primary mutations often decrease ligand binding affinity, both for inhibitors and natural substrates. Secondary mutations often arise to restore activity towards substrates, sometimes as a result of conformational behavior. For instance, incorporation of I15V, E35D, R41K, and R57K was demonstrated by DEER to increase backbone flexibility in WT HIV-1 PR (Huang *et al*. 2014). These mutations, which are all believed to act as function-restoring secondary replacements, result in the increased population of the curled/tucked flap conformation, which does not normally occur in apo subtype B. Other secondary mutations can favor drug resistance. For example, 13 out of the 19 hydrophobic core residues are associated with drug resistance acquisition, despite the absence of direct contacts with inhibitors (Mittal *et al*. 2012). Importantly, the accumulation of mutations in HIV-1 PR does not necessarily lead to increased drug resistance (Huang *et al*. 2014; Tzoupis *et al*. 2014; Gupta *et al*. 2015). Indeed, alterations in inhibitor-protein interactions caused by secondary mutations correlate only partly with K_i variations. The explanation most often put forward is that these secondary mutations alter the dynamic ensemble of the protein, which propagates through the entire structure (Figure 2.3b), sometimes even affecting the dynamically constrained catalytic Asp25 residue (Leonis *et al*. 2013; Ragland *et al*. 2014; Gupta *et al*. 2015; Roche *et al*. 2015). Thus, if a secondary mutation does not alter protein dynamics, it will have no effect on inhibitor binding. Nonetheless, increased drug resistance is the usual outcome of mutational accumulation. For instance, a DEER study comparing a pediatric AIDS patient-derived HIV-1 PR before and after protease inhibitor treatment found that the accumulation of secondary mutations shifts the populations of the conformers to favor the open-like

conformations, when it originally favored the closed conformation (Carter *et al*. 2014). These authors emphasize the importance of the L63P replacement on the behavior of the protein in the pre-treatment sample, arguing that it likely affects hydrophobic core packing, and thus favoring the closed conformation in the apo protein.

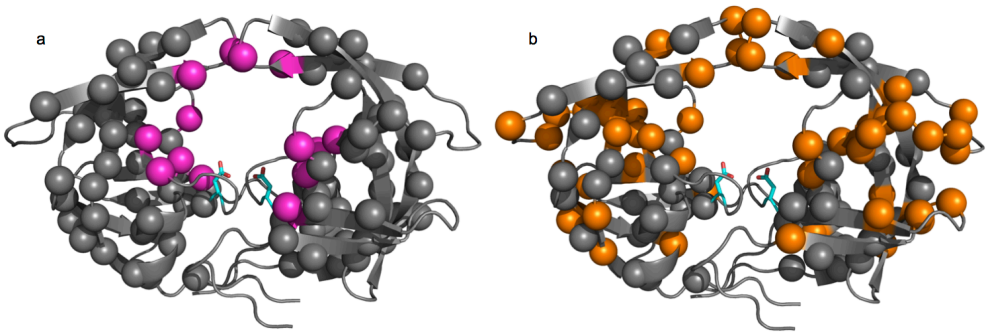


Figure 2.3 Location of drug resistance and conformational mutations in HIV-1 PR.

Structural mapping of various mutations on the crystal structure of apo HIV-1 PR (1HHR). The catalytic Asp25 residue is depicted as cyan sticks on each enzyme protomer. All spheres represent a residue documented to undergo mutation in HIV-1 PR (Wu *et al*. 2003). (**a**) Location of drug resistance-inducing mutations. Magenta spheres represent mutations that impair inhibitor binding through direct active-site contacts (primary mutations) and gray spheres represent distal indirect mutations (secondary mutations); (**b**) Location of mutations documented to affect the conformational sampling of HIV-1 PR, depicted as orange spheres. Gray spheres are other drug resistance-inducing mutations. All mutations were compiled from references listed in the text.

These results illustrate how knowledge of the conformational transitions in HIV-1 PR may help develop new drug strategies against AIDS, further exemplifying the importance of concerted and organized molecular flexibility for the proper catalytic function in this enzyme. Targeting conformational transitions has already provided promising treatment alternatives for new HIV-1 PR drug leads. The use of MD to screen for new competitive or allosteric inhibitors that target MD-identified, transiently open surface cavities was shown to be effective against mutants of HIV-1 PR (Leonis *et al*. 2013; Kunze *et al*. 2014; Ung *et al*. 2014). In one case, a designed allosteric inhibitor was demonstrated to be effective in vitro using enzymatic essays, even against drug-resistant variants (Ung *et al*. 2014). Interestingly, the allosteric inhibitor strategy used in these cases was designed to lock a conformation in place, thus preventing the catalytic cycle from talking place.

2.5.2 DNA Polymerase β

DNA polymerase β (Pol β) is a 39 kDa monomeric enzyme involved in the base excision repair of damaged DNA. Pol β prefers short-gapped or single-base-gapped DNA substrates (Beard and Wilson 2006) and is responsible for the repair of around 20,000 DNA lesions per cell per day (Barnes and Lindahl 2004). Mutations of some residues in Pol β have been shown to increase the error rate of the enzyme; further, overexpression of Pol β has been implicated in nearly 30% of human cancers (Starcevic *et al*. 2004; Loeb and Monnat 2008). Incorporation of correct nucleotides is an essential step in the catalytic mechanism of the enzyme. The structure of Pol β consists of two domains: an 8-kDa lyase domain involved in the removal of deoxyribose 5′-phosphate, and a 31-kDa polymerase domain which performs template-directed DNA synthesis (Beard *et al*. 2002). The polymerase domain is further composed of nucleoside 5′-triphosphate selection, nucleotide transferase, and DNA-binding subdomains, corresponding to the fingers, palm, and thumb regions, respectively. The catalytic mechanism of Pol β, which requires two Mg^{2+} ions as cofactors, is shown in Figure 2.4.

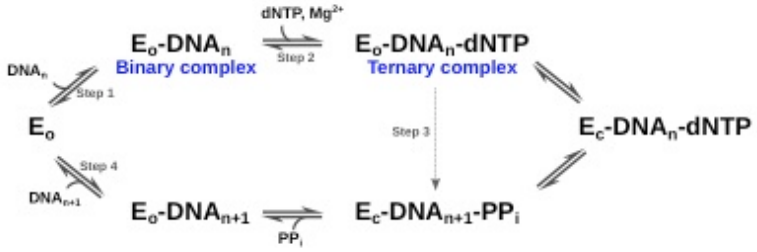


Figure 2.4 Schematic representation of the catalytic mechanism of Pol β.

E_o and E_c correspond to the open and closed conformations of the enzyme. Sequential binding of the DNA (DNA_n) and dNTP leads to the formation of the binary (step 1) and ternary complexes (step 2), respectively. Subsequent conformational rearrangement of the enzyme (E_c) aligns the active site for catalysis (step 3). The chemical step of catalysis is followed by the release of pyrophosphate (PP_i) and the DNA product (DNA_{n+1}) (step 4). Figure adapted from (Towle-Weicksel *et al*. 2014; Moscato *et al*. 2016).

Crystallographic studies have shown that Pol β undergoes large conformational changes upon substrate binding (Sawaya *et al*. 1994; Sawaya *et al*. 1997; Batra *et al*. 2006; Batra *et al*. 2008; Freudenthal *et al*. 2012; Freudenthal *et al*. 2013) and suggest an induced fit mechanism for the binding of correct nucleotides (Sawaya *et al*. 1997). The apo form of the enzyme is characterized by an extended conformation of the lyase domain (Figure 2.5a), which adopts a compact open conformation upon DNA binding (Figure 2.5b). Binding of the nucleotide 5′-

triphosphate (dNTP) promotes closing of the N-helix in the finger subdomain (Figure 2.5c). NMR studies were recently performed to characterize the millisecond timescale conformational exchange experienced by the apo and DNA-bound binary complex forms of Pol β (Berlow *et al*. 2012). The authors of that study showed that conformational exchange in the apo form occurs primarily in the lyase domain and the base of the thumb in the polymerase domain, and corresponds to residues that are either in direct contact or in close proximity to the substrate in the binary complex. In contrast, the number of residues displaying conformational exchange in the binary complex with DNA was significantly reduced. These residues are primarily located in

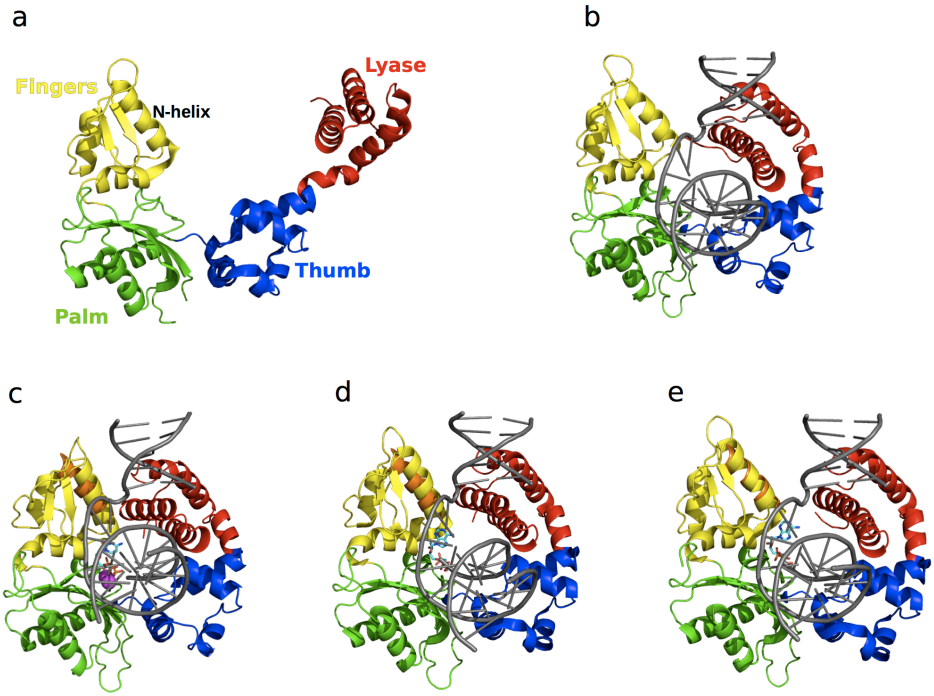


Figure 2.5 Structure of free and ligand-bound states of Pol β.

The lyase domain is shown in red while the thumb, palm, and fingers subdomains of the polymerase domain are shown in blue, green and yellow, respectively. Pol β in the (**a**) apo (PDB 1BPD) form is characterized by an open conformation; (**b**) binary complex (PDB 1BPX) formation is accompanied by a 40 Å conformational shift of the lyase domain to form a compact "open" conformation; (**c**) matched dNTP-bound ternary (PDB 1BPY) complex shows a 10 Å shift of the N-helix of the fingers domain relative to the binary form (in orange), leading to the formation of a "closed" conformation. Binding of mismatched dNTPs has been shown to adopt (**d**) closed (PDB 3C2M) and (**e**) open (PDB 4F5P) conformations of the N-helix. For comparison, the N-helix from the binary form is shown in orange in the ternary complexes in **c–e**, illustrating subtle yet functionally significant conformational exchange in this enzyme. DNA substrates in all the binary and ternary forms are shown in grey (**b**–**e**), the dNTPs in **c**–**e** are shown as cyan sticks, and the Mg^{2+} ions are represented as magenta spheres in c.

the polymerase domain, suggesting that DNA binding limits the motions in the lyase domain (Figure 2.6). Chemical shift variations were located in the lyase domain, associated with large conformational changes experienced by Pol β upon DNA binding (Figure 2.5a,b). Comparison of the conformational exchange relaxation rates with chemical shift variations between the apo and binary forms indicates that the enzyme in the apo form samples conformations adopted by the DNA-bound binary form, thus contradicting the induced fit mechanism proposed by the aforementioned crystallography experiments, at least for the formation of the binary complex (step 1 in Figure 2.4). Interestingly, residues displaying conformational exchange in the apo and binary complex forms correspond to mutational sites known to affect Pol β activity and identified in colon cancer tumors (Donigan *et al*. 2012). Taken together, these results hint at the potential role of conformational motions in the catalytic function of Pol β.

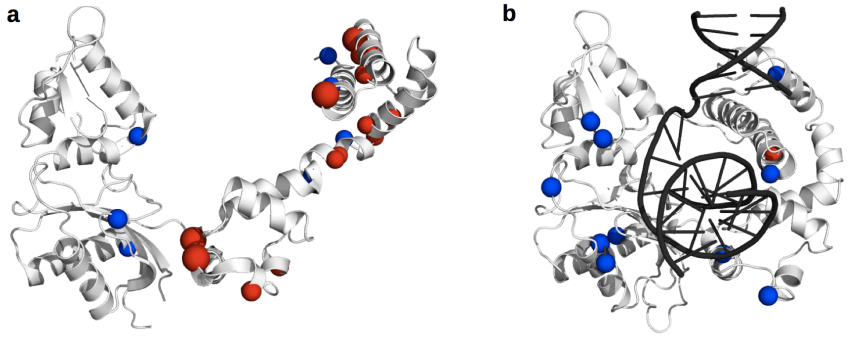


Figure 2.6 Conformational exchange experienced by Pol β.

Residues displaying elevated R_2 and ^{15}N-CPMG relaxation dispersion curves are shown as blue and red spheres, respectively, corresponding to C_α atoms. (**a**) The apo form shows millisecond dynamics predominantly in the lyase domain (residues 1**–**90). Residues with k_{ex} ~1400 s^{-1} are shown as smaller spheres while residues with k_{ex} ~4000 s^{-1} are shown as larger red spheres. Few residues in the other polymerase domains show elevated R_2 values; (**b**) Only residue Glu21 displays ^{15}N-CPMG relaxation dispersion in the binary form. Other residues displaying elevated R_2 values are located in the polymerase domain of Pol β. DNA in the binary form is shown in black. The relaxation data used for preparing this figure were taken from reference (Berlow *et al*. 2012).

Conformational changes in both Pol β and substrates have been suggested to play a critical role in the selection and incorporation of correct nucleotides (Beard *et al*. 2014). Experimental (Sawaya *et al*. 1997; Bakhtina *et al*. 2004; Batra *et al*. 2006; Batra *et al*. 2008; Moscato *et al*. 2016) and computational (Arora *et al*. 2005; Kirmizialtin *et al*. 2012) studies suggest an induced fit mechanism for the selection of correct nucleotides (step 2 in Figure 2.4), which form a Watson-Crick base pair with the DNA substrate. The mechanism involved in the selection of correct dNTP binding and discrimination is a hotly debated topic (Moscato *et al*. 2016). Crystal

structures with open (Figure 2.5d), closed (Figure 2.5e), and intermediate conformations have been obtained with mismatched or incorrect dNTPs, such as dA that cannot form a correct base pair with dG of the DNA substrate. Previous fluorescence studies (Zhong *et al*. 1997; Balbo *et al*. 2011; Bakhtina *et al*. 2005) showed movements in Pol β upon incorporation of dNTPs and cofactors, while more recent studies (Towle-Weicksel *et al*. 2014) indicate the rapid movement (ms timescale) of the fingers domain upon binding to correct dNTPs. Incorporation of incorrect dNTPs did not induce these conformational changes.

Moscato *et al*. (2016) characterized the effects of Pol β binding to correct and incorrect dNTPs on the millisecond timescale using NMR relaxation dispersion experiments. Binding of correct dNTP to Pol β led to large chemical shift variations localized in the fingers subdomain, similar to previous observations (Sawaya *et al*. 1997; Freudenthal *et al*. 2012). No residue displayed conformational exchange, suggesting a stable conformation of the ternary complex on the millisecond timescale, similar to that observed for the binary complex. In contrast, chemical shift changes upon addition of the incorrect dNTP were much smaller than those observed for the correct dNTP. Comparison of the direction of chemical shift variations during titrations, performed using chemical shift perturbation (Williamson 2013), revealed that the conformation of the mismatching complex is not on-pathway to the closed conformation observed upon binding of the correct nucleotide. Further, binding of the mismatched dNTP led to enhanced millisecond dynamics of 10 residues with conformational exchange rates varying between 500 and 2700 s^{-1}. These results support an induced fit mechanism for the ternary complex (step 2 in Figure 2.4), where binding of the correct nucleotide promotes a shift from the open binary complex form to a closed ternary complex conformation.

2.6 **Concluding Remarks**

The intrinsic dynamical properties of enzymes have been shown to impact the catalytic function in a variety of well-characterized systems. Advances in experimental and computational approaches continue to offer insights into the important role of conformational fluctuations in function on timescales relevant for enzyme catalysis and other biological events. Conformational exchange on the catalytically relevant timescale facilitates sampling of excited-state conformations or sub-states, aiding in various steps along the reaction pathway, such as substrate recognition, binding, and product release. In this review, we depicted the important effects of conformational dynamics on correct substrate selection and substrate/inhibitor recognition and binding in two enzyme systems. These observations also emphasize the role of

dynamic fluctuations in tuning the conformational landscape for efficient catalysis, and present a potential approach for using dynamics to modulate function and for designing better inhibitors in drug design.

2.7 Acknowledgments

This work was supported by the National Institute of General Medical Sciences (NIGMS) of the US National Institutes of Health (NIH) under award number R01GM105978 (to N.D.) and a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant under award number RGPIN-2016-05557 (to N.D.). D.N.B. is the recipient of an NSERC Postgraduate Scholarships-Doctoral (PGS D) Award and N.D. holds a Fonds de Recherche Québec-Santé (FRQS) Research Scholar Junior 2 Career Award.

3 DEUXIÈME ARTICLE

La structure tertiaire d'une protéine ne suffit pas à elle seule à expliquer les conformations transitoires que cette protéine peut adopter. Ceci a été démontré par Donald Gagné, qui a étudié l'ensemble des RNases humaines lors des travaux qu'il a effectués au laboratoire. Il a montré que malgré la conservation de la structure tertiaire de ces protéines, les échanges conformationnels qu'elles subissent étaient différents d'une RNase à l'autre, et qu'ils pouvaient être compartimentalisés (Gagné *et al*. 2012; Gagné et Doucet 2013; Gagné *et al*. 2015a; Gagné *et al*. 2015b; Gagné 2015; Gagné *et al*. 2017). La première étape pour faire suite à ces résultats était de les remettre en contexte, de déterminer selon l'évolution quelles RNases sont les plus proches les unes des autres, et de comparer les mouvements que subissent les enzymes de la famille à la lumière de ces informations. C'est pourquoi nous avons comparé les mouvements intrinsèques de 23 RNases, incluant celles caractérisées précédemment par Donald Gagné, suite à une analyse phylogénétique complète. Ces analyses ont été rapportées dans un article sous la référence suivante : **C Narayanan*, DN Bernard*, K Bafna, D Gagné, CS Chennubhotla, N Doucet et PK Agarwal (2018) Conservation of dynamics associated with biological function in an enzyme superfamily. Structure 26, 1-11**.

L'article a été publié le 6 mars 2018. Chitra Narayanan et moi en sommes encore une fois co-premiers auteurs, et nous avons contribué également à sa rédaction. J'ai travaillé sur toutes les parties expérimentales, c'est-à-dire l'expression et la purification des protéines et la RMN, à l'exception des résultats de RMN qui ont été repris d'expériences de Donald Gagné qui n'avaient pas été publiées (le CPMG sur la RNase 4 par exemple) mais que j'ai réanalysées. J'ai également effectué les modélisations par homologie. Chitra Narayanan et Khushboo Bafna se sont séparé les nombreuses analyses bioinformatiques. Le projet était cosupervisé par Chakra Chennubhotla, Pratul Agarwal et bien entendu Nicolas Doucet.

3.1 Résumé

Les membres de superfamilles d'enzymes qui partagent des fonctions chimiques et/ou biologiques ont également des caractéristiques en commun. Bien que le rôle de la structure soit bien caractérisé, le lien entre la fonction et la dynamique des enzymes est mal compris. Nous présentons une caractérisation systématique des mouvements intrinsèques de plus de 20 membres de la superfamille des RNases pancréatiques, qui ont toutes un repliement similaire. Cette étude est motivée par le fait que la plage d'activité chimique ainsi que les mouvements

moléculaires chez les homologues des RNases s'étend sur cinq ordres de grandeur. Les mouvements ont été caractérisés à la fois par résonance magnétique nucléaire et par des simulations computationnelles. Une analyse phylogénétique a mené au regroupement des séquences en sous-familles fonctionnellement distinctes. La caractérisation détaillée des diverses RNases a montré la conservation de comportements dynamiques chez des enzymes de même sous-famille. Ces résultats suggèrent que la pression sélective pour la conservation des comportements dynamiques, entre autres facteurs, peut être liée à la spécificité des fonctions chimiques et biologiques chez une superfamille d'enzymes.

3.2 **Abstract**

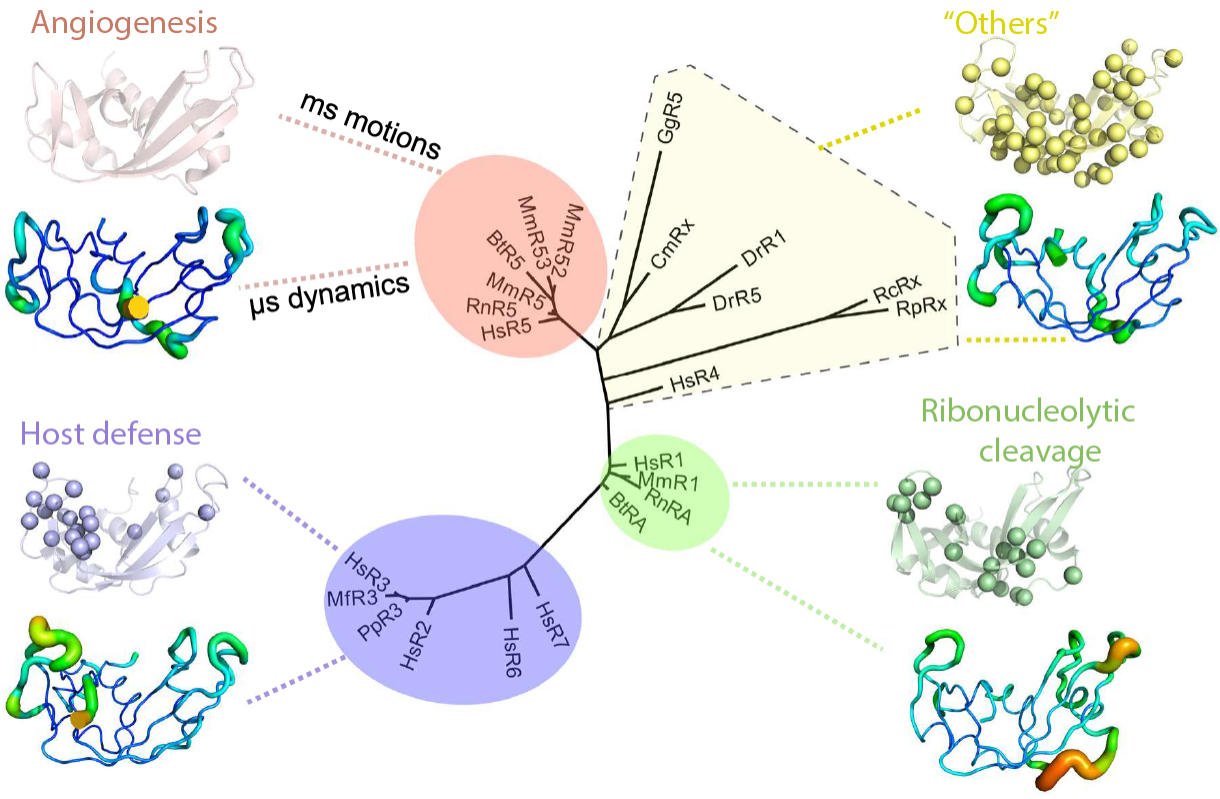


Figure 3.1 Graphical Abstract

Phylogenetic clustering of the 23 RNases generated based on maximum-likelihood analysis. Sequences are classified into four clusters, identified using different colors. Next to each cluster is a depiction of the exchange patterns on the ms and on the µs timescales, as extracted from ^{15}N-CPMG and MD simulation data.

Enzyme superfamily members that share common chemical and/or biological functions also share common features. While the role of structure is well characterized, the link between

enzyme function and dynamics is not well understood. We present a systematic characterization of intrinsic dynamics of over 20 members of the pancreatic-type RNase superfamily, which share a common structural fold. This study is motivated by the fact that the range of chemical activity as well as molecular motions of RNase homologs spans over 10^5 folds. Dynamics was characterized using a combination of nuclear magnetic resonance experiments and computer simulations. Phylogenetic clustering led to the grouping of sequences into functionally distinct subfamilies. Detailed characterization of the diverse RNases showed conserved dynamical traits for enzymes within subfamilies. These results suggest that selective pressure for the conservation of dynamical behavior, among other factors, may be linked to the distinct chemical and biological functions in an enzyme superfamily.

3.3 Introduction

The linkage between dynamics and enzyme catalysis has been intensely debated (Kohen 2015). Enzymes, like other molecules, experience internal motions over a wide range of timescales. These motions are driven by temperature and surrounding environment, including the solvent (Agarwal 2005; Frauenfelder *et al.* 2006). Evidence continues to emerge from an increasing number of systems emphasizing the involvement of dynamics in enzyme function (Eisenmesser *et al.* 2005; Boehr *et al.* 2006; Nashine *et al.* 2010; Holliday *et al.* 2017). However, the presence of motions even near/in the active site, or their occurrence at timescales that coincide with rate-limiting events in enzyme mechanisms, does not automatically imply that these motions play a role in enzyme function (Gobeil *et al.* 2014). One of the challenges associated with investigating enzyme dynamics is that structure and dynamics are inter-related, thus making it difficult to decouple their individual contributions to enzyme catalysis.

The enzyme fold has been proposed to serve as a scaffold that enables optimal positioning of the catalytic residues to provide a unique environment, which is very different from solvent, and provides structural and electrostatic features necessary for transition state stabilization (Warshel *et al.* 2006). Catalytically important residues are conserved across enzymes that share the same fold and catalyze the same (or similar) chemical reaction (Ojha *et al.* 2007). It has been suggested that if dynamics plays a functional role in enzyme catalysis, then it must also be subjected to evolutionary pressure and be conserved as a part of the enzyme fold topology (Ramanathan and Agarwal 2011). Therefore, characterization of dynamics of enzymes within a superfamily could provide vital information on the evolutionary conservation of dynamics (or lack thereof) among structural and functional homologs that catalyze the same chemical reaction.

This approach could provide much needed information for teasing apart the role of structure and dynamics in enzyme catalysis.

Pancreatic-type ribonucleases (RNases) are a superfamily of enzymes found mostly in vertebrates that catalyze the hydrolysis and transphosphorylation of RNA substrates (Singhania *et al.* 1999; Sorrentino 2010; Batot *et al.* 2017). Nearly 650 sequences corresponding to members of this superfamily are found in the Pfam database (Finn *et al.* 2016). Eight catalytically active canonical RNases (and five supplemental pseudo-genes that lack the conserved active-site residues essential for ribonucleolytic activity) were identified in the initial sequencing of the human genome (Sorrentino 2010). All structurally characterized RNases display a conserved structural fold and share the common chemical function of cleaving the phosphodiester bond of RNA substrates, albeit with varying degrees of catalytic efficiencies and specificities. The canonical RNases have further been reported to perform a variety of biological functions such as angiogenesis, anti-pathogenicity, and immunosuppressivity, in addition to the common ribonucleolytic activity (Sorrentino 2010). While the biological functions are broadly known, and have been characterized for some members identified in the human genome, a significant number of members remain largely uncharacterized to date. The unique similarity in structure (fold), while displaying significant diversity in function, makes the RNase superfamily an excellent candidate for characterizing the effect of conserved dynamics on function.

A variety of biophysical, nuclear magnetic resonance (NMR), and computational approaches have indicated the role of conformational dynamics in regulating catalysis in bovine RNase A (BtRA), the archetypal member of this family (Gagné and Doucet, 2013). Millisecond motions of a loop located over 20 Å away from the active site, suggested to be propagated by a conserved hydrogen-bonding network (Cole and Loria 2002; Doucet *et al.* 2009; Gagné and Doucet 2013), were shown to be essential for efficient catalysis (Doucet *et al.* 2011). Mutations disrupting this hydrogen-bonding network were shown to result in the uncoupling of dynamics on the catalytic timescale and a corresponding decrease in the catalytic turnover (Doucet *et al.* 2009). Furthermore, reduced catalytic efficiency relative to bovine RNase A was observed in two RNase A variants lacking this hydrogen-bonding network (Watt *et al.* 2007; Doucet *et al.* 2009; Doucet *et al*. 2011). Interestingly, the global conformational exchange rates of the diverse RNase members vary over five orders of magnitude (microseconds to hundreds of milliseconds) coinciding with the range of substrate turnover (and chemical efficiency), which also varies up to five orders of magnitude between the most efficient and least efficient members of the human canonical forms (Gagné and Doucet 2013). For example, human RNase 5 (angiogenin), which

lacks the dynamical loop 4 region of RNase A shown to affect catalysis, displays 10^5 fold reduction in ribonucleolytic activity compared with RNase A (Raines *et al.* 1995).

As a step toward characterizing the relationship between dynamics, chemical, and biological functions of various members of this broad vertebrate family, we performed a systematic characterization of the intrinsic dynamics of over 20 RNase homologs. The selected RNases correspond to a representative set of homologous vertebrate sequences whose structures have been determined previously by X-ray crystallography or NMR and are available in the PDB. Using a combination of NMR and computational approaches, we characterized the dynamics of these proteins over the microsecond-to-millisecond (μs-ms) timescale to probe the relationship between sequence diversity and dynamics, and its correlation with the designated biological functions. The present study focuses on characterizing and comparing the intrinsic dynamics of apo proteins, as these enzymes bind to and catalyze a wide variety of RNA substrates. Our results show that while the diverse RNase homologs used for the analysis share a common structural fold, the dynamical properties of these proteins vary significantly. Clustering superfamily members into evolutionarily distinct subfamilies showed conservation of dynamical properties within subfamily members and notable variations between different subfamilies. Categorizing homologous RNase superfamily sequences into subfamilies, which also exhibit similar dynamics, provides a unique opportunity for further investigation to tease apart the role of structure and dynamics in enzyme function.

3.4 Method Details

3.4.1 Genes, Expression and Purification

Escherichia coli codon-optimized sequences of *Pongo pygmaeus* ribonuclease 3 (PpR3) and *Macaca fascicularis* ribonuclease 3 (MfR3) (UniProt entries P47781 and P47779) were synthesized in IPTG-inducible pJExpress414 vectors containing the ampicillin resistance gene (ATUM, Newark, CA). These vectors were transformed into *E. coli* BL21 (DE3) for all protein expression experiments. All human RNases were cloned, expressed, and purified according to previous protocols (Doucet *et al*. 2009; Gagné *et al*. 2012; Gagné *et al*. 2015b; Gagné *et al*. 2017). ^{15}N- and [^{15}N,^{13}C]-labeled protein expression and purification was conducted as described previously (Doucet *et al*. 2009; Gagné *et al*. 2012), with the following modifications. The temperature was lowered to 30°C following addition of IPTG, the volume of culture media was of 1 L, and bacteria were grown overnight before being harvested by centrifugation. Protein

concentrations were determined using extinction coefficients of 11 960 and 14 940 $M^{-1}cm^{-1}$, respectively, as estimated by ExPASy ProtParam.

3.4.2 Solution NMR Experiments

2D ^{15}N-HSQC, 3D-HNCACB and 3D-CBCA(CO)NH assignment experiments were performed on a Varian INOVA 500 MHz spectrometer (298 K), while ^{15}N-CPMG relaxation dispersion experiments were performed on the same spectrometer as well as a Varian INOVA 800 MHz spectrometer (298 K), as described earlier (Doucet *et al*. 2009; Gagné *et al*. 2012; Gagné and Doucet 2014). NMR processing and analysis was performed using NMRPipe (Delaglio *et al.* 1995), CcpNmr Analysis (Vranken *et al.* 2005) and Sparky (Goddard and Keneller 2008).

3.4.3 Phylogenetic Analysis

Multiple sequence alignment of the 23 RNase sequences was performed using ClustalΩ (Sievers *et al.* 2011). A complete list of sequences and sequence identifiers is provided in Table 7.1. The phylogenetic tree of the aligned sequences was determined based on maximum likelihood approach using RaxML v8.0.26 with the WAG amino acid substitution model (Stamatakis 2014). Reliability of the branching was assessed using the bootstrapping method over 100 iterations. The resulting phylogenetic tree in Newick format was visualized using Figtree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

3.4.4 System Preparation for Computer Simulations

Coordinates of 21 RNases were obtained from the corresponding PDB files for the apo model preparations (see Table 7.1 in Supplemental Information). The structures for two hominidae RNase 3 are not available; therefore, homology models were generated as described below. System preparation for molecular dynamics (MD) simulations were performed using the leap module in AMBER (Case *et al.* 2005) v14 and AMBER's ff14SB force-field. Equilibrated systems were used for each of the 23 proteins to run a total of 1ms of production MD under constant energy conditions (NVE ensemble). A PME cut-off of 8 Å was used as suggested by AMBER input files for GPU-enabled simulations (Salomon-Ferrer *et al.* 2013). To address an important concern regarding the possible bias of results from a single MD trajectory and ensure reproducibility, for each of the 23 systems two independent MD trajectories each 0.5 ms were performed.

3.4.5 Homology Modeling of Hominidae RNase 3 Sequences

Structures of two hominidae RNase 3 sequences, *Pongo pygmaeus* (PpR3) and *Macaca fascicularis* (MfR3), were generated through homology modeling using MODELLER v9.16 (Webb and Sali 2014). The two hominidae sequences share 84.3% and 85.8% sequence identity with human RNase 3 (PDB ID: 1QMT), which was used as the template for modeling. The automodel option in MODELLER was used to generate models, of which two models with the lowest DOPE scores were chosen for each of the two sequences.

3.4.6 Molecular Dynamics Simulations

The protein systems were neutralized by the addition of counter ions and the resulting systems were then immersed in a rectangular box of SPC/E water, with a 10 Å minimum distance between the protein and the edge of the periodic box. The prepared systems were equilibrated using a protocol described previously (Ramanathan *et al.* 2011). The equilibrated system was used for each of the 23 proteins to run a total of 1 ms of production MD (two independent 0.5 ms MD trajectories, see details below) under constant energy conditions (NVE ensemble). The Particle-mesh Ewald (PME) method was used for long-range electrostatics and a time-step of 2 femtoseconds was used during production runs. Coordinates of the system were saved every 5 picoseconds, for a total of 200 000 conformational snapshots for each protein (as a sum of 100 000 snapshots from each independent trajectory). All simulations were performed using graphical processing units (GPU) enabled version of AMBER's MD simulation engine pmemd (v14). A PME cut-off of 8 Å was used as suggested by AMBER input files for GPU-enabled simulations (Salomon-Ferrer *et al.* 2013). To address an important concern regarding the possible bias of results from a single MD trajectory and ensure reproducibility, for each of the 23 systems two independent MD trajectories each 0.5 ms were performed. The first trajectory after equilibration was used for production at 300 K, while for the second trajectory the structure was slowly heated to 310 K using MD with 50 000 steps and minimized to generate conformational and structural diversity. Production run for the second trajectory was also performed at 300 K.

3.4.7 RMSF and RMSF10 Calculations

All-atoms root-mean-square fluctuations (RMSF) were computed based on the conformational snapshots collected during the MD simulations. To identify global motions on slower time-scales from MD, for each of the 23 systems we separately computed and aggregated the fluctuations associated with the first (slowest) 10 quasi-harmonic modes (RMSF10). It is well known that

slowest 10 modes contribute to the majority of fluctuations in proteins (>80%) and the use of RMSF10, instead of all modes (RMSF), removes the faster stochastic motions of the protein, allowing focus on intrinsic dynamics of proteins (Ramanathan and Agarwal 2009). Both these calculations were performed using AMBER's ptraj analysis program. All trajectory conformations were first aligned to a common structure, to remove any translation and overall molecular rotation during the simulations.

3.4.8 Quantitative Comparison of Dynamical Properties

Pearson's correlations were used for the quantitative characterization of dynamical similarities (and differences) between protein pairs. Correlation between RMSFs for each residue a_i and b_i of RNases A and B, respectively, were calculated using the equation:

Équation IV

$$Pairwise\ Pearson's\ correlation = \frac{\sum_{i=1}^{n}(a-\bar{a})(b-\bar{b})}{\sqrt{\sum_{i=1}^{n}(a-\bar{a})^2}\sqrt{\sum_{i=1}^{n}(b-\bar{b})^2}}$$

where n corresponds to the number of residues including the gaps in the alignment. a and b correspond to the average RMSF values for RNase A and B. Pairwise-correlations were calculated for each of the 23 RNases using RMSF10 excluding residues corresponding to gaps in the pairwise alignment. Average correlations within and between sub-families were calculated by averaging correlation coefficients of sequences being compared.

3.5 Results

As a representative set of the pancreatic RNase superfamily, 21 diverse RNases, whose three-dimensional structures are available in the PDB, were selected for detailed dynamical characterization. Human RNases 1–7 were included in this analysis. Additionally, two hominidae RNase 3 sequences that are closely related to human RNase 3 were also analyzed using structures generated based on homology modeling (see Supplemental Information). A list of all RNase sequences and sequence identifiers is provided in Table 7.1. Comparison of the 23 RNase dataset shows that the sequences share an average sequence identity of ~37.4%, and an average root-mean-square deviation of 2.47 Å for the alignment of all backbone atoms of the 21 members with solved structures. Sequence and structural alignments indicated that most variable regions are, unsurprisingly, associated with loop regions of the protein fold (Figures

3.2A and 3.2B). Catalytic triad residues (His12, Lys41, and His119, bovine RNase A residue sequence numbering), which are essential for the ribonucleolytic activity, are conserved in all 23 proteins analyzed.

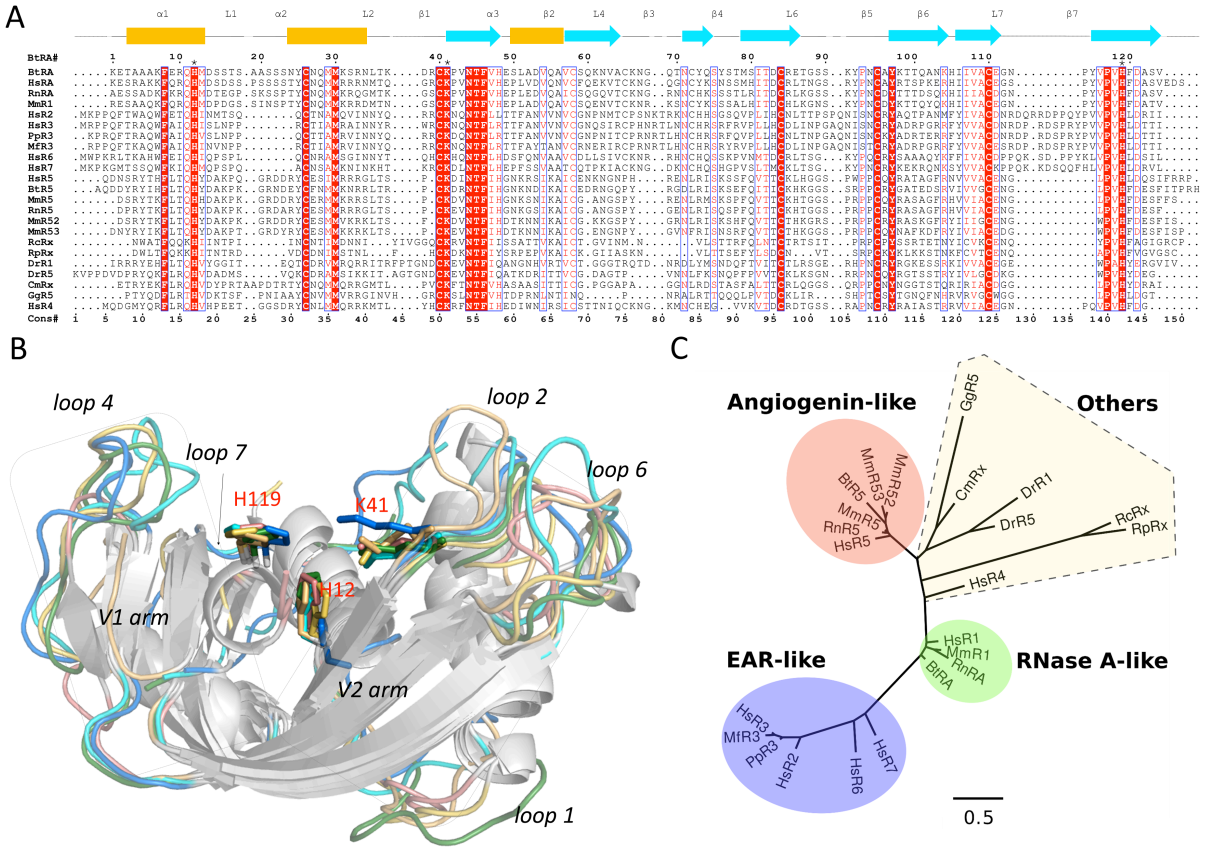


Figure 3.2 Structural and Phylogenetic Analysis of RNase Homologs

(A) Consensus sequence alignment of the 23 RNase homologs, with the corresponding sequence identifiers displayed to the left of each sequence. Detailed information about the sequences is provided in Table 7.1. The secondary structural elements α helices (orange rectangles), β strands (cyan arrows), and loop regions (L), corresponding to the bovine RNase A structure, are displayed above the alignment. Conserved active-site residues His12, Lys41, and His119 (BtRA numbering) are identified with an asterisk above the alignment. BtRA (PDB: 7RSA) sequence numbering (BtRA#) is shown at the top and consensus sequence numbering (Cons#) below the multiple sequence alignment. A list of all consensus sequence positions is presented in Figure 7.2. Gaps in the alignment are represented using dotted lines. (B) Structural alignment of representative RNase homologs. Structures include 7RSA (green), 1QMT (blue), 2HKY (cyan), 1ANG (salmon), 3SNF (yellow), and 1RNF (gold). The V_1 and V_2 arms are identified by rectangular boxes. Active-site residues are displayed as sticks and labeled with BtRA numbering. (C) Phylogenetic clustering of the 23 RNases generated based on maximum-likelihood analysis. Sequences are classified into four clusters, identified using different colors. Alignments were prepared using EsPript (Robert and Gouet 2014).

Phylogenetic analysis of the 23 RNase sequences led to the identification of four distinct sequence clusters, henceforth referred to as subfamilies (Figure 3.2C). The RNase A-like subfamily (four members) includes members that share similarity with bovine RNase A (BtRA), whose primary function is associated with the cleavage of RNA substrates. Human and bovine ribonucleases of this subfamily are the most efficient ribonucleolytic catalysts characterized to date within the RNase superfamily. Members of the Eosinophil-associated ribonuclease-like (EAR-like) subfamily (six members) are characterized by insertion of a loop (L7), and have been suggested to play an important role in host defense and inflammatory response (Boix 2001; Gupta *et al.* 2013), in addition to preserving the common ribonucleolytic function. Members of this subfamily show catalytic turnover rates of 5- to >400-fold lower relative to BtRA, for those with known catalytic turnover rates (data not shown). The Angiogenin-like subfamily (six members) is characterized by a truncated loop 4 (L4), a loop region that displays conformational dynamics on the millisecond timescale in bovine RNase A and was shown to play an important role in its ribonucleolytic function. Catalytic efficiency of human angiogenin (HsR5) is 10^5-fold lower than bovine RNase A. However, its ribonucleolytic activity was shown to be essential for biological function (angiogenesis) (Kelemen *et al.* 1999). Previous studies also showed that replacing the truncated loop 4 of HsR5 with that of bovine RNase A increases catalytic efficiency of this chimeric protein by over 200-fold (Harper and Vallee 1989), while swapping loop 4 from RNase A with that of HsR5 reduced the activity of this chimera by 60-fold (Allemann *et al.* 1991). The other seven RNases not forming a part of the three groups described above were classified as Others, and include sequences that are not necessarily closely related to each other. The functional characteristics of this cluster are not well understood and could therefore display functional and dynamical properties similar to the three subfamilies described above.

3.5.1 Conformational Dynamics on the Microsecond-to-Millisecond Timescale

The timescale of catalysis varies widely in the RNase superfamily. BtRA is known to catalyze cleavage of RNA with a k_{cat} ~1700 s^{-1}, with product release being the rate-limiting step of the reaction (Doucet *et al.* 2011; Watt *et al.* 2007). Other members are slower catalysts, showing up to 10^5-fold weaker catalytic efficiency (k_{cat}/K_M). We used a combination of NMR CPMG (Carr-Purcell-Meiboom-Gill) relaxation dispersion experiments and molecular dynamics (MD) simulations to probe the µs-ms dynamical properties of these enzymes. Figure 3.3 shows the conformational exchange of various RNases from each of the four subfamilies probed using ^{15}N-CPMG relaxation dispersion experiments. Residues exhibiting conformational exchange on the

μs-ms timescale are represented as spheres. Comparison of millisecond motions of these selected RNases showed significant differences in regions displaying conformational exchange for members from different subfamilies. In contrast, residues exhibiting conformational exchange are clustered in the same regions for four members within a subfamily (Figure 3.3A). Members of the EAR-like RNase subfamily display motions primarily clustered in the V_1 arm and loop 4. In contrast, BtRA from the RNase A-like subfamily showed conformational exchange in both V_1 and V_2 arms, including loops L1 and L4 (Figure 2B). Human RNase 5 (HsR5) of the Angiogenin-like subfamily exhibits no conformational exchange on the CPMG timescale (Figure 3.3C), confirming results from a previous report (Gagné *et al.* 2015b). To investigate whether HsR5 experiences motions on a faster μs-ms timescale window than the one probed by the CPMG experiment, we also analyzed elevated transverse relaxation rates at short spin-echo

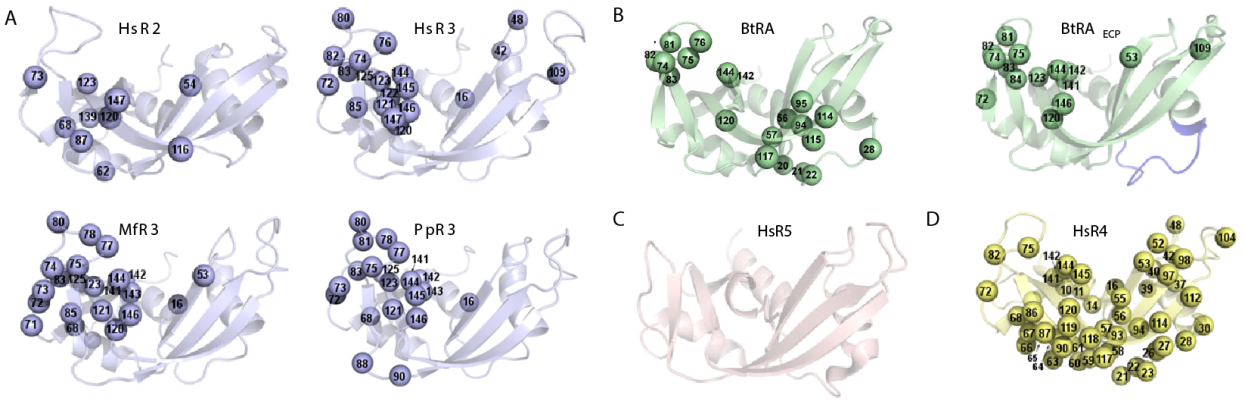


Figure 3.3 Millisecond Conformational Exchange Experienced by Select Hominidae and Bovine RNases

Cartoon representation of: (A) EAR-like subfamily members: *Homo sapiens* EDN (HsR2), *H. sapiens* ECP (HsR3), *Pongo pygmaeus* ECP (PpR3), *Macaca fascicularis* ECP (MfR3); (B) RNase A-like subfamily member *Bos taurus* RNase A (BtRA) and the $BtRA_{HsR3}$ chimera (Doucet *et al.* 2009); (C) Angiogenin-like subfamily member *H. sapiens* angiogenin (HsR5); and (D) *H. sapiens* RNase 4 (HsR4) from the Others cluster. Structures are colored based on the phylogenetic group shown in Figure 3.2C. Residues identified by NMR as undergoing conformational exchange on the millisecond timescale are depicted as spheres, and numbered according to their position in the sequence alignment (see Figure 3.2) to facilitate comparison. Conformational exchange was probed by ^{15}N-CPMG NMR relaxation dispersion experiments at 500 and 800 MHz (298 K). Residues were considered for further analysis only if the difference in measured R_2 ($1/\tau_{cp}$) values at fast (τ_{cp} = 0.625 ms) and slow (τ_{cp} = 10 ms) refocusing pulse delays was greater than 2 s^{-1}, similar to previous reports (Doucet *et al.* 2009; Gagné *et al.* 2012). $BtRA_{HsR3}$, which displays subsector conformational exchange similar to HsR3, is an artificial chimeric hybrid of BtRA in which the 12-residue long loop 1 of BtRA (14-DSSTSAASSSNY-25, BtRA numbering) was replaced by the 6-residue long loop 1 of HsR3 (17-SLNPPR-22, HsR3 numbering) (Gagné et. al. 2012). Structures color-coded based on conformational exchange rates are shown in Figure 7.4.

pulse delays in the CPMG relaxation dispersion profiles (i.e. $R_{2,eff}$ at ν_{CPMG} = 1600 s^{-1}, or τ_{cp} = 0.625 ms). Analysis for apo and ligand-bound HsR5 showed that only one residue per state exhibited elevated $R_{2,eff}$ values (Asn68 in the apo form and Asn109 in the ligand-bound form), which could indicate that these two residues experience conformational exchange on a faster µs-ms timescale (Figure 7.1). Overall, these observations further confirm that HsR5 is a fairly rigid protein on the µs-ms timescale and that no significant conformational exchange is observed on this regime. Human RNase 4 (HsR4), from the Others cluster, displayed extensive conformational exchange throughout the protein (Figure 3.3D). NMR data from other members of the RNase subfamily are presently unavailable on the µs-ms timescale.

Interestingly, insertion of loop 1 (L1) from an EAR-like RNase (HsR3) into BtRA ($BtRA_{HsR3}$, Figure 3.3B) (Doucet *et al.* 2009; Gagné *et al.* 2012) switched the dynamical profile of BtRA into that of HsR3 (Figure 3.3A) in the chimeric protein. The motivation for the design of this chimera was to develop and characterize a control system for investigating the dependence of dynamics on a region based on type of residues rather than secondary structure. Previous studies showed a detailed comparison of the rates of ligand binding and release (Doucet *et al.* 2009), in addition to conformational exchange profiles of BtRA, $BtRA_{HsR3}$, and HsR3 (Gagné *et al.* 2012). To gain insights into changes in the relaxation dispersion of these enzymes in the L1 region, here we compared the fitted relaxation dispersions for residues of loop 1 from HsR3 (EAR-like subfamily) with BtRA (RNase A-like subfamily) and the chimera $BtRA_{HsR3}$ (Figure 3.4A). Our results show that loop 1 residues of BtRA exhibit a two-site conformational exchange at these positions, as observed from the presence of relaxation dispersion profiles (solid lines in Figure 3.4, left panel). In contrast, no motions on this timescale were observed for loop 1 positions in HsR3 and the chimera $BtRA_{HsR3}$, as seen from the flat dispersion profiles (shown as dashed lines in Figure 3.4A, middle and right panels). Residues in structurally equivalent positions in the three enzymes are represented using the same colors. Furthermore, a chemical-shift comparison between BtRA and the chimera showed that chemical-shift variations arising from loop swapping are primarily localized in the V_2 domain (Figure 3.4B). However, residues that acquire conformational exchange in the V_1 domain (Figure 3.3B) do not exhibit significant chemical-shift variations. These observations suggest that alterations in regions displaying functionally relevant motions results in the modulation or loss of conformational exchange for residues in these regions. Overall, these results suggest that conformational exchange profiles are conserved between members within the EAR-like subfamily, perhaps suggesting a role in the

shared biological or chemical function, while distinctly different dynamical fluctuations are observed for functionally distinct homologs from other phylogenetic clusters.

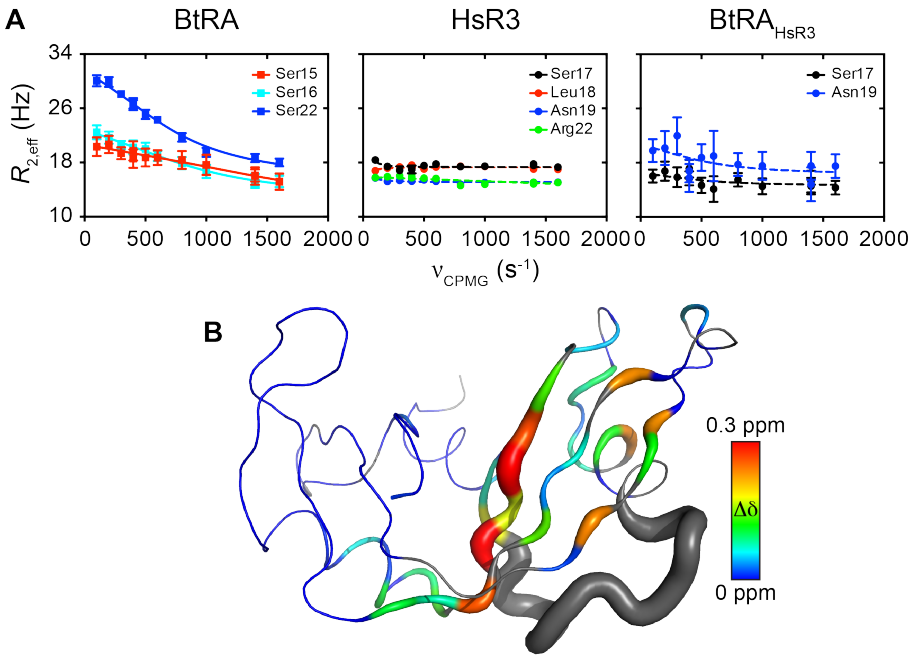


Figure 3.4 Dynamical Effects of Loop 1 Swapping on BtRA

(A) Relaxation dispersion profiles of various loop 1 residues obtained from ^{15}N-CPMG experiments at 800 MHz and 25°C for bovine RNase A (BtRA), human RNase 3 (HsR3), and a previously described (Doucet *et al.* 2009) chimera of RNase A (BtRA$_{HsR3}$) in which loop 1 has been replaced by the corresponding residues of HsR3. ^{15}N-CPMG data was also acquired at 500 and/or 600 MHz for these proteins (not shown for clarity). Solid curves are shown for residues displaying conformational exchange, and dashed curves belong to non-exchanging residues. Curves of the same color represent structurally equivalent residues in the sequence alignment. (B) Compounded ^{1}H-^{15}N chemical-shift variations (Grzesiek *et al.* 1996) induced on BtRA by loop 1 swapping in BtRA$_{HsR3}$. Residues in gray are either unassigned or belong to the swapped loop. Chemical-shift variations for each residue are indicated by both the color and the width of the cartoon putty. Figure was made with PyMOL using the bovine RNase A structure (PDB: 7RSA).

3.5.2 Microsecond Dynamics

Atomic-level dynamical properties of all 23 RNase homologs were probed using MD simulations to determine the detailed dynamical characteristics on the μs timescale (Figure 3.5). Root-mean-square fluctuations (RMSF) associated with the slowest ten quasi-harmonic modes (RMSF$_{10}$) were used to characterize the intrinsic dynamics of the proteins. Previously, such an approach was used for other family members and was shown to match the dynamics from NMR studies and other techniques (Ramanathan and Agarwal 2011). One of the issues with

comparing dynamical patterns between proteins sequences of different lengths is the presence of gaps associated with insertions and deletions. To address this, we represented dynamical profiles of consensus sequences that include gaps in the multiple sequence alignment (Figure 3.2A).

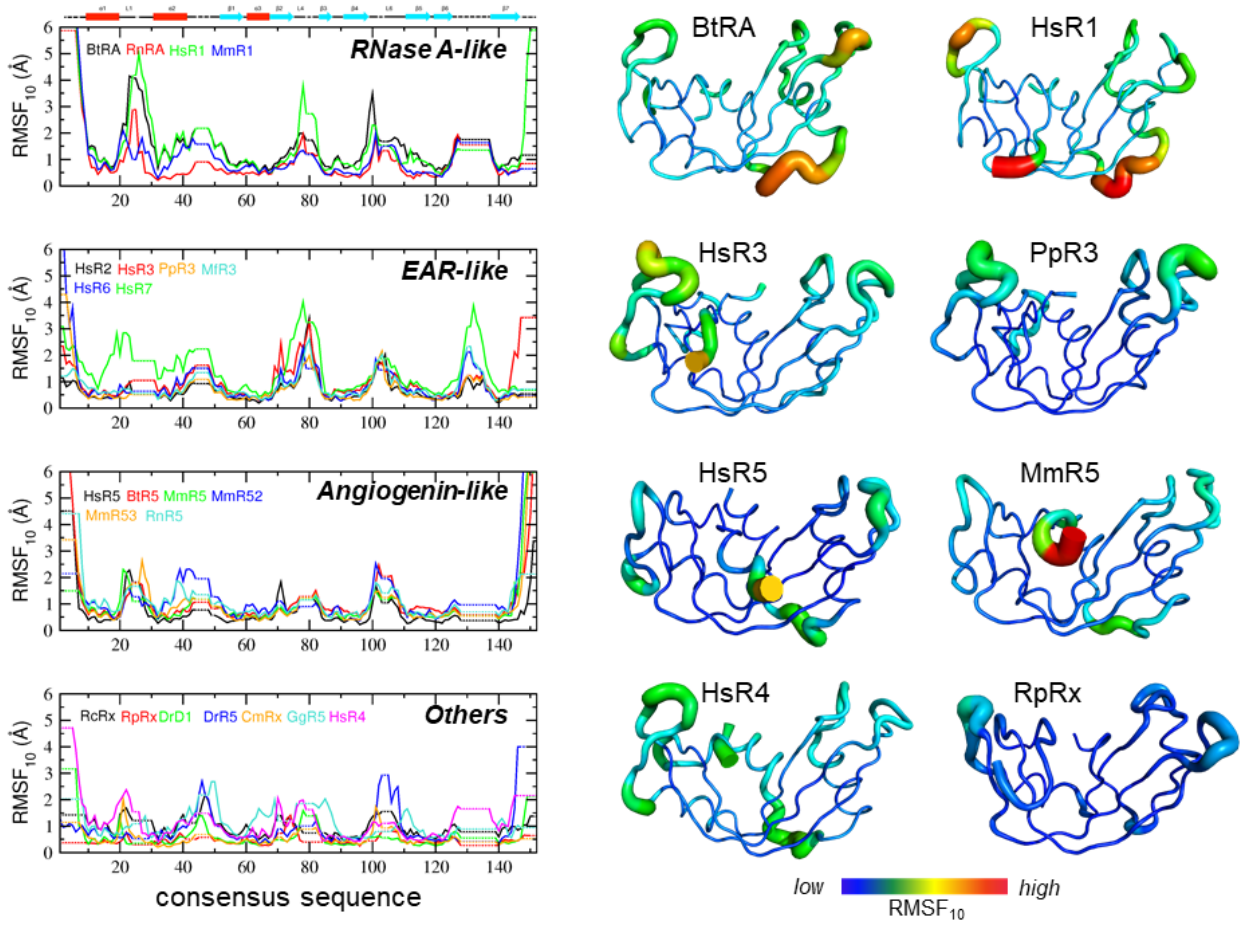


Figure 3.5 Microsecond Conformational Dynamics of RNase Homologs

Root-mean-square fluctuations corresponding to the C_α displacements in the top ten quasi-harmonic modes ($RMSF_{10}$), for the 23 proteins grouped into the four phylogenetic clusters described in Figure 3.2. A consensus sequence with gaps is used (see Figure 3.2A), where gaps are represented as dotted regions for each protein. The panel on the right shows representative RNase homologs from each phylogenetic group using a tube representation where the thickness of the tube corresponds to the flexibility of residues in each protein, with thicker tubes corresponding to flexible regions and thinner tubes representing less flexible regions. The dynamical range represented using the color spectrum is consistent across all sequences with the blue and red ends of the spectrum corresponding to low and high dynamic regions, respectively. The representative structures correspond to *B. taurus* RNase A (BtRA) and *H. sapiens* RNase 1 (HsR1); *H. sapiens* RNase 3 (HsR3) and *P. pygmaeus* RNase 3 (PpR3); *H. sapiens* angiogenin (HsR5) and *Mus musculus* RNase 5 (MmR5); and *H. sapiens* RNase 4 (HsR4) and *Rana pipiens* RNase (RpRx), respectively. N-terminal residues, which showed larger RMSFs (see left panel), are hidden for clarity. Tube representations for all RNases are shown in Figure 7.5.

A comparison of the conformational dynamics on the ms timescale showed similarity in the RMSF profiles of RNases within the same subfamily members, while differences were observed between subfamilies (Figure 3.5), consistent with the qualitative comparison of the conformational exchange profiles from experiments (Figure 3.3). The RNase A-like proteins, which share significant sequence similarity with BtRA, showed greater similarity in dynamical patterns between each other. Differences in the dynamical patterns were restricted to the loop L1 region (residues 20–30, consensus sequence numbering, Figure 7.2), which showed the largest fluctuations for most members in this group except for *Mus musculus* (mouse) RNase A (MmR1), which showed diminished dynamics in this region relative to other members. Characterization of hydrogen-bonding lifetimes in MmR1 showed the interaction between the side chains of Ser19 of loop1 and Thr82 of the β4 strand, occurring in ~81% of the simulation ensemble. This strong hydrogen-bonding interaction may be dampening the dynamics of loop L1 observed for MmR1.

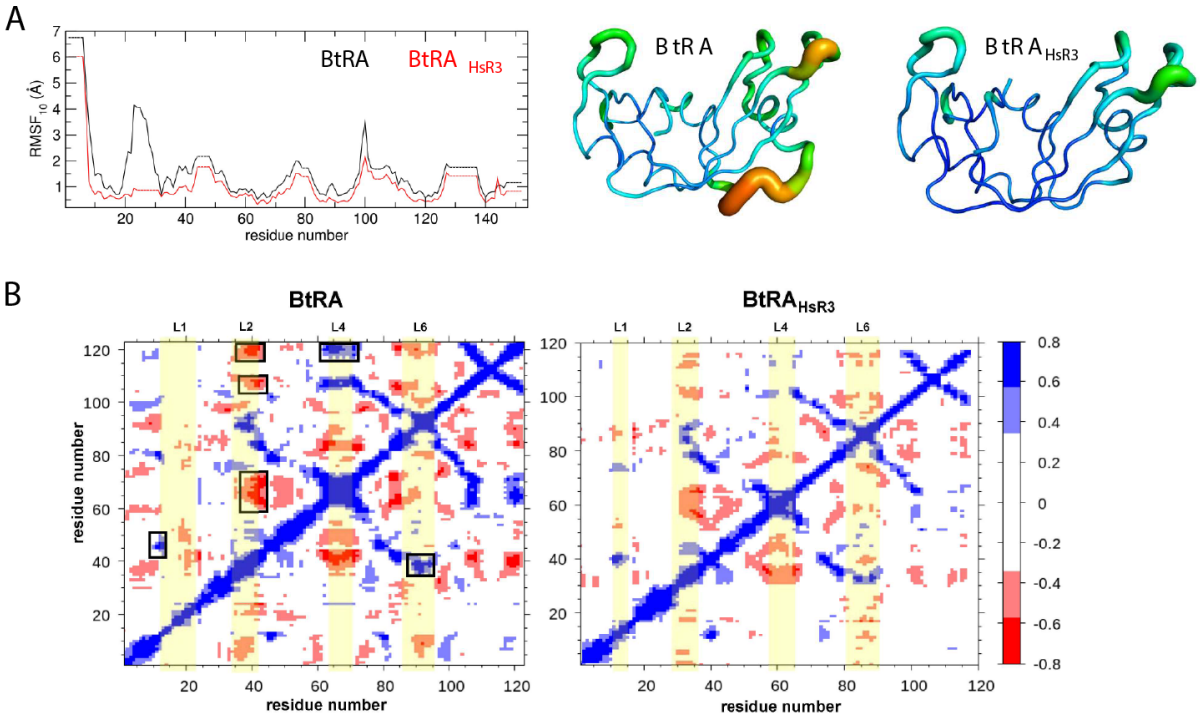


Figure 3.6 Dynamical Properties of $BtRA_{HsR3}$ Chimera

(A) Comparison of the root-mean-square fluctuations of the ten slowest modes ($RMSF_{10}$) between BtRA (black) and the chimera (red). On the right, RMSFs are shown using a tube representation for BtRA and the chimera, where the thickness of the tube corresponds to the flexibility of residues in each protein, with thicker tubes corresponding to flexible regions and thinner tubes representing less flexible regions. (B) Dynamical cross-correlation maps for BtRA (left), with regions displaying significant correlations and anti-correlations identified using boxes, and the $BtRA_{HsR3}$ chimera (right). Loops 1, 2, 4, and 6 are highlighted in yellow.

EAR-like subfamily members exhibit similar dynamical patterns in all regions. RNases of this group are characterized by a truncated loop L1, which incidentally showed the largest fluctuations in the RNase A-like subfamily. Swapping loop 1 from HsR3 into BtRA resulted in a chimera ($BtRA_{HsR3}$), which displayed a loss of dynamics in this loop, in addition to an overall reduction in the dynamics throughout the protein relative to BtRA (Figure 3.6). Human RNase 7 (HsR7) showed larger fluctuations throughout the protein while preserving the dynamical profile exhibited by other members of the subfamily. Angiogenin-like members are characterized by a truncated loop L4, and showed differences in dynamics of loop L2. Members not classified in the groups above (clustered into Others), as expected, showed large variations within the group compared with members within the other three subfamilies. Large dynamical fluctuations were observed in loops L2, L4, and L6 for some members of the group, while other members showed deletions in these regions.

3.5.3 Quantitative Characterization of Dynamical Conservation

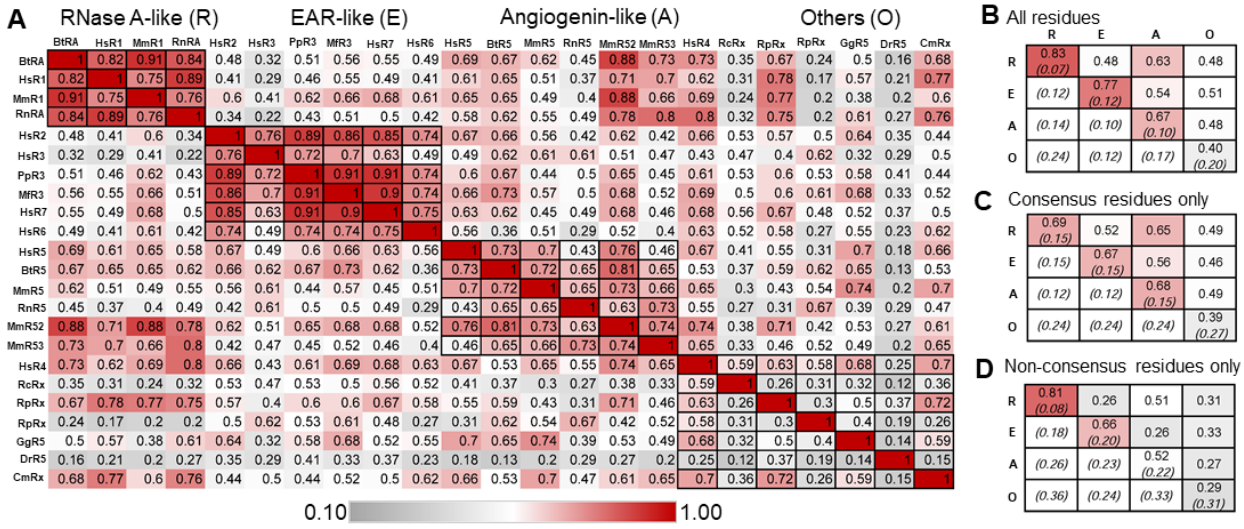


Figure 3.7 Quantitative Characterization of Dynamical Similarities

Pairwise Pearson's correlations for the 23 RNase homologs classified into four phylogenetic clusters. (A) Correlation coefficients for each pair of sequences, color-coded based on strong correlations (red) to weaker or no correlations (gray). (B–D) Average correlations, calculated by averaging the correlation coefficients for sequences compared for: all residue positions (B), consensus positions (C), and non-consensus positions, corresponding to insertions or deletions in sequences (D). The diagonal elements correspond to correlations within phylogenetic subfamilies while off-diagonal elements correspond to correlations between the subfamilies. Standard deviations (SD) for average correlation calculations are shown in parentheses. Abbreviations R, E, A, and O correspond to phylogenetic clusters RNase A-like, EAR-like, Angiogenin-like, and Others, respectively.

To determine whether the similarities (and differences) observed within (between) subfamilies are statistically significant, we performed a quantitative comparison of the dynamical profiles of sequences by performing pairwise correlations for all pairs of sequences (Figure 3.7). The pairwise Pearson's correlation coefficient for all 23 sequences (Figure 3.7A) showed larger correlations between members within, rather than between the three functionally distinct phylogenetic subfamilies, except for the Others cluster, which showed small intracluster correlations. To better elucidate these comparisons within and between subfamilies, we calculated the average value for these correlations, with the SDs presented in parentheses (Figures 3.7B–D). Comparison of these averages showed that correlations within subfamilies are significantly larger than those between subfamilies for the three functionally distinct RNase subfamilies (Figure 3.7B). Overall, these observations highlight the conservation of dynamical patterns in homologous sequences sharing distinct functional properties, further supporting NMR experiments within and between subfamily members.

Sequence and structural alignments showed insertions/deletions of residue positions in different regions along the primary sequence of the RNases (Figure 3.2). To determine whether these insertions and deletions influence the functional identity of these homologous sequences, we calculated the average correlations for consensus and non-consensus residue positions separately. Consensus positions correspond to residue positions that are preserved in all sequences (86 amino acid positions) while non-consensus positions represent positions with insertions or deletions in at least one of the sequences in the dataset (66 amino acid positions). A comparison of the correlations within subfamilies showed similar trends for both consensus and non-consensus positions (Figures 3.7C and D). Interestingly, a comparison of correlations between subfamilies for non-consensus residues showed markedly smaller correlations (gray in Figure 3.7) between the three functionally distinct subfamilies, except for the Angiogenin-like subfamily, which showed similar average correlations within the subfamily and with RNase-A-like subfamily on this timescale (Figure 3.7D). Overall, these observations suggest that non-consensus positions influence the distinct dynamical patterns observed among the functionally divergent phylogenetic subfamilies. It could potentially be counter-argued that dynamical patterns will always be most influenced at non-consensus residue positions when comparing phylogenetically distinct proteins. However, it is important to remember that the physicochemical identity of residues determines the extent and nature of dynamical motions and rotameric conformer populations. Experimental characterization of the $BtRA_{HsR3}$ chimera (with L1 loop region from HsR3 inserted into BtRA) revealed significant changes in the dynamics of this region and the V_2 domain relative to BtRA (Figure 3.4). It is tempting to speculate that as loop

regions are flexible, any sequence in the L1 loop region of BtRA could be expected to show large dynamical motions. On the contrary, the results from chimera characterization indicate that L1 loop residues in $BtRA_{HsR3}$ show dynamical behavior similar to L1 in HsR3 rather than BtRA.

Analysis of MD simulations allows the characterization of atomic-scale details, including residue-level dynamical interactions. A comparison of dynamic residue-residue cross-correlations for the different RNase homologs shows distinct variations among members. Detailed characterization of dynamical crosscorrelations was performed for members with experimental dynamical information (correlation maps with black outlines in Figure 7.3). BtRA of the RNase A subfamily showed strong anti-correlations (value less than -0.4) for residues 65–72 of loop L4 (positions 75–84, consensus sequence numbering, Figure 7.2) with residues near the active site and the catalytic K41 (position 51) and correlations (value greater than 0.4) with residues near His119 (position 142). Anti-correlations were also observed between residues near Lys41 and residues 105–110 (positions 119–124) and those near catalytic His119. Furthermore, strong correlations were observed for residues near, and including His12 (position 17) and residues 13–23 of loop 1 (positions 18–29) with residues around His48 and Thr82 (positions 58 and 94, respectively), which form a hydrogen-bonding network shown previously to be involved in the allosteric modulation of catalysis and product release in this enzyme (Doucet *et al.* 2011; Watt *et al.* 2007).

Interestingly, loop swapping in the chimera ($BtRA_{HsR3}$) resulted in changes throughout the protein, in regions far from loop 1, including significant changes near active-site residues. Specifically, we observed a loss of correlated motions between loop 4 and the C-terminal region in the vicinity of the active-site residue His119, and loss of anti-correlated motions of loop 4 and the vicinity of Lys41 (another active-site residue) and between residues near His119 and Lys41 (Figure 3.6B). HsR3 of the EAR-like subfamily showed strong anti-correlations between residues near the catalytic residue His15 and residues surrounding residues 86–97 of loop 6 (positions 99–110, Figure 7.2) and residues 27–38 (positions 36–51), while correlations were observed with residues 42–46 of β-strand 1 (positions 55–59). Residues in the vicinity of and including the catalytic Lys38 (position 51) showed anti-correlations near residues 60–65 of loop 4 (positions 73–78) and residues near catalytic residue His128 (position 142). Furthermore, His128 also showed correlations with residues near His15 (position 17) and loop 4, and anti-correlations with residues near Lys38. Residues 28–36 (positions 37–49) showed anti-correlations with residues 98–102 (positions 111–115).

Contrary to members from the other phylogenetic subfamilies, HsR5 of the Angiogenin-like subfamily showed significantly weaker correlations and anti-correlations between residues. The lack of conformational dynamics of HsR5 and the resulting absence of correlated motions is consistent with our observations from relaxation dispersion experiments, which showed the absence of motions on the millisecond timescale in this enzyme (Figure 3.3C). HsR4 of the Others subfamily showed numerous dynamical correlations and anti-correlations between different regions of the protein, suggesting enhanced correlated dynamics throughout the protein, consistent with NMR relaxation dispersion experiments, which showed extensive conformational exchange in this protein (Figure 3.3D). Dynamical correlations of all 23 RNase homologs are shown in Figure 7.3.

3.6 **Discussion**

Increasing evidence from characterization of discrete enzyme systems suggests the important role of conformational motions in the catalytic function in these systems (Gagné and Doucet 2013; Kohen 2015; Doshi *et al.* 2016; Holliday *et al.* 2017). Despite tremendous progress in our understanding of the role of conformational motions in function, evolutionary conservation of dynamical properties on functionally relevant timescales among structural and functional homologs within enzyme families remains largely uncharacterized. Numerous studies using normal mode analysis have shown the conservation of backbone flexibility in homologous proteins sharing a common structural framework (Maguid *et al.* 2008; Marsh and Teichmann 2014; Fuglebakk *et al.* 2015). Experimental and computational approaches further showed the conservation of dynamics in selected proteins within enzyme families (Gagné *et al.* 2012; Marsh and Teichmann 2014; Narayanan *et al.* 2017). However, systematic characterization of superfamily-wide conservation of dynamical properties on the functionally relevant µs-ms timescales is limited.

In this study, we combined ^{15}N-CPMG NMR relaxation dispersion experiments and detailed MD simulations to systematically characterize dynamical properties of diverse pancreatic-type RNases on the µs-ms timescale. Using a set of > 20 representative RNases, chosen based on the availability of three-dimensional structures in the PDB, we probed the conservation of dynamical patterns in the homologous RNases. Phylogenetic analysis led to the grouping of sequences into four clusters, 16 of these sequences forming part of three functionally distinct subfamilies (Figure 3.2C), where individual members within subfamilies also share common

biological function distinct from those of other subfamilies. Little is known about the biological function of the remaining seven members that were grouped into the Others cluster.

Slowest dynamical modes were previously shown to display the largest evolutionary conservation (Maguid *et al.* 2008), with these modes being the most robust to amino acid substitutions (Zheng *et al.* 2007). Through quantitative characterization of the dynamical properties of the slowest modes from MD simulations, we show distinct variations in the dynamical patterns between different members of the RNase superfamily. Conservation of dynamical patterns was observed for sequences clustered into phylogenetic subfamilies (Figure 3.5) that incidentally also shared similar biological functions. This conservation in the dynamical profiles within subfamilies was also observed on the millisecond timescale probed by NMR for select RNases (Figure 3.3). These observations are consistent with previous studies showing that enzymes sharing similar functions also share similar reaction-promoting dynamics (Xia *et al.* 2013; Ramanathan *et al.* 2014; Kerns *et al.* 2015).

Rapid changes through evolution in dynamical properties were previously shown to be linked to functional divergence in proteins (Lai *et al.* 2012). Sequence and structural alignments of RNases used in this study showed variations, corresponding to insertions/deletions, in different regions along the primary sequence of RNases (Figure 3.2). A comparison of dynamical patterns of just these non-consensus amino acid positions showed notable differences between subfamilies (Figure 3.7D). These observations suggest that dynamical differences observed between sequences sharing distinct functional properties could be linked to the type of residues in non-consensus amino acid positions, which may influence the functional identity of these sequences.

The use of chimeric constructs such as $BtRA_{HsR3}$ investigated in this study serves as an important control system. It is widely discussed that the overall enzyme dynamics (particularly at slow timescales) is determined by the structural fold. Comparison of the conformational exchange profiles for the chimeric and wild-type RNases (Figures 3.3 and 3.4) indicates that for enzymes sharing the same fold, significant differences could arise from changes in non-conserved or variable regions. However, it is not sufficient to say that loop regions are the most dynamical regions. The extent and range of motions depends on the nature and type of residues in the loop regions. Indeed, characterization of dynamical correlations from computer simulations of the $BtRA_{HsR3}$ chimera shows the long-range effect of loop 1 swapping on loop 4 and residues in the vicinity of active-site residues His119 and Lys41 (Figure 3.6B).

Each enzyme could provide a unique set of local dynamical motions depending on the sequence (such as seen in the case of loop 1 in the two parents and the chimera, Figure 3.4). Characterization of these differences and its relation to changes in catalytic efficiency requires further investigation. Furthermore, it should be noted that even in the conserved regions of fold, different family members could exhibit distinct dynamics. For example, see Figure 3.3 and related text for differences in the dynamics of the V_2 domain in different subfamilies. These observations open new avenues for probing the effect of conformational exchange associated with loop swapping between members from the different subfamilies. Of notable interest is the effect on the conformational dynamics and catalytic turnover of swapping loop 4 from BtRA on HsR5, which is characterized by a truncated loop 4 and a 10^5-fold lower catalytic efficiency relative to BtRA. While swapping the loop 4 residues from HsR5 on to BtRA and vice versa has been performed before, the effect of this swapping on the conformational exchange and the resulting impact on catalysis have not been characterized (Harper and Vallee 1989; Allemann *et al.* 1991; Jang *et al.* 2009).

HsR5 exhibits at least 10^5-fold weaker ribonucleolytic activity than bovine RNase A against standard RNA substrates (Harper and Vallee 1989), although this ribonucleolytic activity remains essential for angiogenesis. This low enzyme activity has been structurally and mutationally rationalized by the obstructive positioning of a C-terminal residue segment of HsR5 (residues 116–123, primarily Glu117), which is occluding a pyrimidine-binding subsite that is normally solvent exposed in homologous RNases (Leonidas *et al.* 2002). This trademark blockage of the active site in HsR5 demonstrates that a conformational change of this C-terminal tail is required to alleviate the occluded pocket for proper RNA substrate recognition and ribonucleolytic activity to occur. The ^{15}N-CPMG dynamical investigation performed here did not provide satisfactory evidence that this conformational change occurs on the millisecond timescale. Since movement of the C-terminal tail is essential for the ribonucleolytic activity of HsR5, this was not totally unexpected considering that the catalytic rate of this enzyme occurs on a much slower timescale than its RNase homologs. For instance, HsR5 exhibits a k_{cat}/K_M = 1.1 $M^{-1}s^{-1}$ against dinucleotide RNA substrate UpA, relative to k_{cat}/K_M = 4 000 000 $M^{-1}s^{-1}$ in RNase A (Harper and Vallee 1989). Additionally, no less than 19 unique protein partners have been experimentally reported to interact with HsR5 for a wide variety of biological processes involving cell migration, adhesion, invasion, proliferation, and apoptosis (Sheng and Xu 2016). This could suggest that the ribonucleolytic activity of HsR5 is carefully programmed in the cellular context, perhaps only triggered in the presence of the proper protein-binding partner and/or after translocation to the correct cellular compartment. Conformational exchange of the

C-terminal tail could thus be closely modulated or allosterically controlled by an uncharacterized protein partner. Although we did not observe [^{1}H,^{15}N]-HSQC (heteronuclear single-quantum coherence) resonances that could indicate residues experiencing exchange on a slower timescale, future chemical exchange saturation transfer (CEST), ZZ-exchange, and/or hydrogen-deuterium-exchange experiments could provide additional clues on slow timescale motions experienced by HsR5.

In summary, characterization of dynamical properties of diverse RNases showed their conservation among sequences that also share similar biological functions. Interestingly, mutations of subfamily sequences in regions showing large dynamical motions have been shown to alter their biological function (Table 7.2). The similarity in dynamics for members with common biological function requires further investigation, especially for subfamilies where the connection between the biological and chemical function is unclear. The present results provide vital clues that conservation of dynamics as a part of a protein fold may have a broader significance beyond enzyme catalysis.

3.7 Acknowledgments

We thank Tara Sprules of the Québec/Eastern Canada High Field NMR Facility (McGill University) and Sameer Al-Abdul-Wahid (University of Guelph), and Drs. Om Choudhary and Arvind Ramanathan for helpful discussions. This work was supported by NIGMS/NIH under award number R01GM105978 (to N.D., C.S.C., and P.K.A.), Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery grant under award number RGPIN-2016-05557 (to N.D.), NSERC Postgraduate Scholarship-Doctoral (to D.N.B.), Postdoctoral Fellowship from the Fondation Armand-Frappier (to C.N.), and Fonds de Recherche Québec – Santé (FRQS) Research Scholar Junior 2 Career award number 32743 (to N.D.). Computing time from Oak Ridge Leadership-class Computing facility through the ALCC was used for calculations.

4 TROISIÈME ARTICLE

L'article précédent a démontré que la classification des RNases selon les critères de l'évolution permettait d'expliquer certaines différences entre les profils d'échanges conformationnels, particulièrement ceux se produisant à l'échelle de la nanoseconde. Des RNases appartenant à des regroupements différents arborent en effet des profils distincts d'échanges conformationnels, alors que ceux appartenant à un même groupe présentent des profils plus similaires. Toutefois, il importe de rappeler que les RNases de différentes sous-familles ont des activités biologiques distinctes de celles d'autres sous-familles, telles qu'énumérées à la section 1.2.3. En contrepartie, les RNases présentent une conservation de la fonction catalytique malgré la grande variété de leurs fonctions biologiques. Ceci implique que la liaison des ligands est conservée chez toutes les RNases catalytiquement actives peu importe leurs fonctions biologiques, avec une variabilité induite par les différences de préférences envers les séquences de nucléotides. Nous nous sommes donc intéressés aux changements de structure associés à la liaison des ligands. En effet, la liaison de ligands implique des changements dans le paysage conformationnel de l'enzyme, tel qu'expliqué à la section 1.4.2, et nous avons cherché à voir si ces changements pouvaient être conservés d'une enzyme à l'autre si le phénomène de liaison lui-même était conservé. L'article suivant explore cette question : **C Narayanan*, DN Bernard, K Bafna, D Gagné, PK Agarwal et N Doucet (2018) Ligand-Induced Variations in Structural and Dynamical Properties Within an Enzyme Superfamily. Front Mol Biosci 5, 54**.

Cet article a été publié dans la revue Frontiers in Molecular Biosciences le 12 juin 2018. Contrairement aux deux articles précédents, Chitra Narayanan en est l'unique première auteure. Donald Gagné (RNases 4 et 5) et moi (RNases 2 et 3) avons effectué l'ensemble des expériences en laboratoire, c'est-à-dire l'expression et la purification des protéines et les expériences RMN (attributions, titrages, CPMG), et j'ai effectué l'ensemble des analyses de ces expériences RMN, à l'exception du Chemical Shift Projection Analysis (CHESPA), dont les résultats demeurent les plus importants de l'article. Le projet était cosupervisé par Pratul Agarwal et Nicolas Doucet.

4.1 Résumé

La catalyse enzymatique est un processus complexe impliquant plusieurs étapes, incluant la reconnaissance et la liaison des substrats, la conversion chimique elle-même, et la libération des produits. De plus en plus d'études démontrent les rôles fonctionnel et évolutif que jouent les phénomènes d'échanges conformationnels dans l'activité catalytique optimale. La liaison de ligands change le paysage conformationnel des enzymes, causant des réarrangements conformationnels à longue distance. En utilisant comme systèmes modèles différents membres de la superfamille des ribonucléases pancréatiques dont les fonctions sont distinctes, nous avons caractérisé les changements de structure et de conformation associés à la liaison de deux mononucléotides. En analysant les résultats d'expériences de titrages RMN par chemical shift projection analysis (CHESPA) et en les combinant avec des expériences de dispersion de la relaxation, nous démontrons que les membres biologiquement distincts de la superfamille des RNases subissent des perturbations uniques lors de la liaison de ces ligands, qui ne sont pas conservées même chez des membres structurellement similaires. Les réseaux d'acides aminés subissant des variations coordonnées de déplacement chimique lors de la liaison des deux ligands sont uniques à chaque RNase analysée. Nos résultats révèlent la contribution de réarrangements conformationnels aux perturbations de déplacements chimiques observées. Ces observations procurent un aperçu important dans la contribution des différentes spécificités de liaison et des échanges conformationnels aux perturbations associées à la liaison de ligands chez des membres fonctionnellement distincts de la superfamille des RNases pancréatiques.

4.2 Abstract

Enzyme catalysis is a complex process involving several steps along the reaction coordinates, including substrate recognition and binding, chemical transformation, and product release. Evidence continues to emerge linking the functional and evolutionary role of conformational exchange processes in optimal catalytic activity. Ligand binding changes the conformational landscape of enzymes, inducing long-range conformational rearrangements. Using functionally distinct members of the pancreatic ribonuclease superfamily as a model system, we characterized the structural and conformational changes associated with the binding of two mononucleotide ligands. By combining NMR chemical shift titration experiments with the chemical shift projection analysis (CHESPA) and relaxation dispersion experiments, we show that biologically distinct members of the RNase superfamily display discrete chemical shift

perturbations upon ligand binding that are not conserved even in structurally related members. Amino acid networks exhibiting coordinated chemical shift displacements upon binding of the two ligands are unique to each of the RNases analyzed. Our results reveal the contribution of conformational rearrangements to the observed chemical shift perturbations. These observations provide important insights into the contribution of the different ligand binding specificities and effects of conformational exchange on the observed perturbations associated with ligand binding for functionally diverse members of the pancreatic RNase superfamily.

4.3 Introduction

Enzyme catalysis accelerates reaction rates of chemical reactions up to 20 orders of magnitude relative to uncatalyzed reactions (Wolfenden 2006). In the widely understood paradigm, enzymes act by reducing the free energy barrier, thus facilitating the formation of the transition state. The mechanism of enzyme catalysis is complex, including, but not limited to, the following steps: substrate recognition and binding to the active site; the chemical step involving the conversion of substrate(s) to product through the transition state; and release of product(s). Any of these steps along the reaction coordinates could act as the rate-limiting step, thus determining the rate of enzyme turnover (Gutteridge and Thornton 2004; Narayanan *et al.* 2016). More recently, conformational rearrangements, corresponding to time-dependent atomic displacements of residues and/or larger structural elements, have been suggested to influence the catalytic power of enzymes (Pelz *et al.* 2016; Kovermann *et al.* 2017). This conformational flexibility, while maintaining the native three-dimensional structure of the protein, is often essential for optimal enzyme function (Henzler-Wildman *et al.* 2007a). Enzymes were shown to sample distinct conformations, termed sub-states, facilitated by conformational fluctuations that occur over a wide range of timescales (Narayanan *et al.* 2016). While the role of conformational exchange in enzyme catalysis is debated, evidence from experimental and computational approaches have revealed the correlation between enzyme turnover rates and the timescale of conformational motions in numerous enzyme systems, including but not limited to, alcohol dehydrogenase, dihydrofolate reductase, and ribonuclease A (RNase A) (Agarwal *et al.* 2012; Narayanan *et al.* 2016).

Members of the pancreatic ribonuclease superfamily have served as a model system for numerous biophysical experiments, including enzyme mechanism studies (Sorrentino 2010). Bovine RNase A is the prototypical member of this superfamily, whose primary function is the cleavage of the 3',5'-phosphodiester bond in single- and double-stranded RNA substrates.

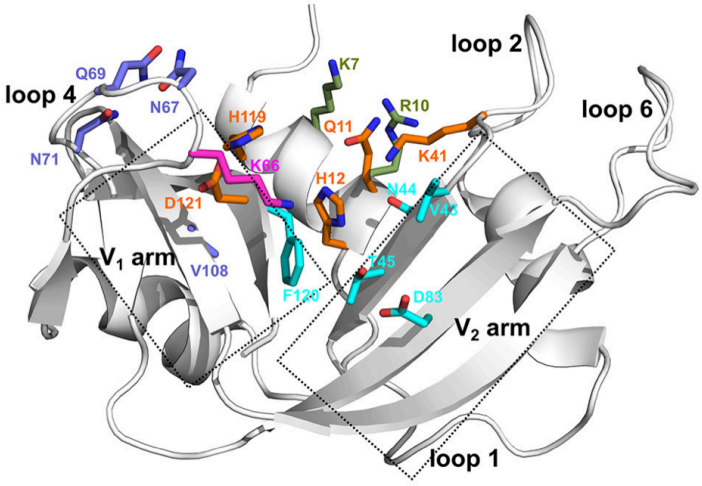


Figure 4.1 Ligand binding subsites identified in bovine RNase A

Residues of the ligand binding subsites, corresponding to the base B_1, B_2, and phosphate P_0, P_1, and P_2 subsites are represented as sticks colored cyan, purple, magenta, orange, and olive, respectively. Catalytic residues His12, Lys41, and His119 form part of the P_1 binding site.

Several phosphate (P_n) and nucleotide base (B_n) binding subsites that interact with the substrate molecules were identified in the active site of bovine RNase A (Figure 4.1), which displays a strong preference for a pyrimidine in the primary base binding site (B_1) and a purine in the secondary base binding site (B_2) (Nogués *et al.* 1995). The rate-limiting step was previously shown to correspond to a conformational change in a distal loop that is associated with the product release step in RNase A (Watt *et al.* 2011; Gagné and Doucet 2013). The functional role of conformational exchange in product release was previously shown to rely on the movement of distal loop regions in RNase A, a hypothesis that we further extended to include functional RNase homologs sharing a conserved structural fold (Cole and Loria 2002; Watt *et al.* 2007; Doucet *et al.* 2009, 2011; Gagné *et al.* 2012; Gagné and Doucet 2013; Narayanan *et al.* 2017, 2018). Mutations of residues in these loop regions were shown to result in reduced rate constants for product release and lower substrate affinity, highlighting the role of these long-range motions in this enzyme (Gagné and Doucet 2013). Eight catalytically active (canonical) and five inactive (non-canonical) RNases were identified in the sequencing of the human genome (Cho *et al.* 2005). In addition to their common ribonucleolytic function, the canonical RNases, henceforth referred to as subtypes, have evolved to perform other biological functions such as host defense, immunosuppressivity, angiogenesis, and anti-pathogenic activity, among others (Sorrentino 2010). Further, the experimentally characterized human RNase subtypes display a wide range of substrate specificities (Boix *et al.* 2013), catalytic

activities (Sorrentino 2010; Gagné and Doucet 2013) and conformational fluctuations on the millisecond timescale (Narayanan *et al.* 2017, 2018). Efforts to relate specific conformational exchange events with ribonucleolytic function in this enzyme family is thus limited by the broader and often RNA-independent biological functions of many homologous RNase superfamily members.

Coevolving amino acid residues that are proximal to each other in the three-dimensional structure may control the biological properties of proteins (Halabi *et al.* 2009). In a recent study, we identified networks of coevolving amino acids that control distinct aspects of biochemical functions in the pancreatic RNase superfamily by combining the sequence-based statistical analysis with experimental observations (Narayanan *et al.* 2017). Our results demonstrated that networks of residues within a larger functional sector are involved in fine-tuning the catalytic activity among the different RNase subtypes, thus dictating the functional diversity among these RNases. Detailed characterization of the dynamical properties of over 20 RNases in the ligand-free states further revealed large differences in their global conformational exchange patterns observed for members within the superfamily (Narayanan *et al.* 2018a). Using a diverse set of RNases grouped into functionally distinct phylogenetic subfamilies, we demonstrated the conservation of conformational exchange profiles between RNases within subfamilies sharing similar biological functions, while significant differences were reported between subfamilies (Narayanan *et al.* 2018a). These observations from characterization of the ligand-free (apo) state of enzymes provided important insights into the selective pressure that may influence the exchange profiles within a superfamily.

Ligand binding influences the conformational landscape of enzymes, leading to structural and dynamical changes. However, these changes remain largely uncharacterized for most RNases. We previously showed the long-range effects of ligand binding on the structural and dynamical behaviors of select RNases (Gagné *et al.* 2012, 2015b). While these studies provided important insights into the effect of ligand binding to these RNases, several important questions remain unanswered, which the present study seeks to address. For instance, how do structural and conformational exchange properties of different enzymes within the superfamily change upon ligand binding? Do members within functionally distinct subfamilies display similar conformational rearrangements upon ligand binding, as observed for the ligand-free states of members of the RNase superfamily? This study aims to compare and characterize the structural and conformational changes associated with ligand binding for select RNases, corresponding to

representative family members with distinct biological functions, and to gain insights into the mechanism of ligand binding for members of the pancreatic RNase superfamily.

In this study, we probe the chemical shift changes associated with the binding of two mononucleotide ligands (3'-UMP and 5'-AMP) to five selected RNases, corresponding to representative members of four distinct phylogenetic subfamilies. The two mononucleotides mimic the hydrolysis products of the model RNA dinucleotide substrate UpA. Our results show that binding of each of the two ligands induces distinctly different effects on selected RNases. Despite their local and global structural similarity, RNases from functionally distinct subfamilies displayed different effects upon binding to each of the two ligands. Further, RNases within a subfamily also displayed different magnitudes of chemical shift changes near the active site, in addition to notable differences in the long-range effects of the ligand binding subsite. We determined the coordinated changes in chemical shift displacements associated with the binding of the two ligands for the selected RNases using the NMR chemical shift projection analysis (CHESPA). Our results show that networks of residues displaying coordinated displacements vary in RNases within and between different subfamilies. Our results further illustrate the contribution of dynamical changes upon ligand binding to the observed chemical shift perturbations. We suggest that, among other factors, the distinct nucleotide binding specificities and conformational rearrangements triggered by ligand binding may be contributing to the unique functional and biological roles of these RNases within the cell, despite their apparent structural similarity.

4.4 Methods

4.4.1 Enzyme Cloning, Expression and Purification

Bovine RNase A and human RNases 2, 3, 4, and 5 were cloned, expressed, and purified according to protocols described previously (Doucet *et al.* 2009; Gagné *et al.* 2012, 2015b). Sequences were codon-optimized for *Escherichia coli* expression and cloned into NdeI/HindIII-digested expression vector pJexpress411 (ATUM, Newark CA). ^{15}N- and [^{15}N,^{13}C]-labeled protein expression and purification was performed using previously described protocols (Doucet *et al.* 2009; Gagné *et al.* 2012), with these modifications: the temperature was lowered to 30˚C following addition of IPTG, the volume of culture media was 1 L, and bacteria were grown overnight before being harvested by centrifugation. Protein concentrations were determined

using extinction coefficients of 9,880 (RNase A), 17,460 (RNases 2, 3), and 11,835 (RNases 4, 5) $M^{-1}cm^{-1}$, respectively, as estimated by ExPASy ProtParam.

4.4.2 Solution NMR Experiments

2D ^{1}H-^{15}N HSQC, 3D-HNCACB and 3D-CBCA(CO)NH assignment experiments were performed using a Varian INOVA 500 MHz (11.7 T) spectrometer at 298 K. NMR data processing and analyses were performed using NMRPipe (Delaglio *et al.* 1995), CcpNmr Analysis (Vranken *et al.* 2005), and Sparky (Goddard and Keneller 2008).

4.4.3 NMR Titration Experiments

All NMR titration experiments were conducted at 298K on ^{15}N-labeled 150–450 µM protein samples in 15 mM sodium acetate at pH 5.0. The pH was carefully monitored throughout the experiments and readjusted with Tris-base or acetic acid if necessary. Ligands 3'-UMP (Chemical Impex Intl Inc., Wood Dale, IL, USA) and 5'-AMP (BioBasic Inc., Markham, ON Canada) were purchased commercially and dissolved in the same buffer as the protein. ^{1}H-^{15}N sensitivity-enhanced HSQC experiments were acquired at 800 MHz (18.8 T) using spectral widths (points) of 2025 (160) and 8000 Hz (1024) in the ω_1 and ω_2 dimensions, respectively. Titration experiments were performed for each ligand with enzyme:ligand molar ratios up to 1:12 (1:12), 1:18 (1:18), 1:24 (1:18), 1:30 (1:18) and 1:18 (1:30) for the 3'-UMP- (5'-AMP-) bound states of bovine RNase A and human RNases 2, 3, 4, and 5, respectively.

The equilibrium dissociation constant (K_d) was calculated by plotting weighted average chemical shift differences ($\Delta\delta_{obs}$) as a function of ligand concentration and fitting the data using the following equation:

Équation V $$\Delta\delta_{obs} = \frac{\Delta\delta_{\max}}{2[P]_t}\left([P]_t + \frac{[L]_t}{[P]_t}[P]_t + K_d - \sqrt{\left([P]_t + \frac{[L]_t}{[P]_t}[P]_t + K_d\right)^2 - 4\frac{[L]_t}{[P]_t}[P]_t^2}\right)$$

The K_d was estimated by simultaneously fitting the data of all residues affected by ligand binding (Williamson 2013). We note that $\Delta\delta_{max}$ is a fitted parameter and not an experimentally measured one. Uncertainties on these values are given by the standard deviations of the fits. Chemical shift perturbations ($\Delta\delta_{obs}$) were calculated as the difference in the weighted average chemical shift of the ligand-bound (5'-AMP and 3'-UMP) and apo states, as shown below.

Équation VI

$$\Delta\delta_{obs} = \sqrt{\frac{\left(\Delta\delta_H\right)^2 + \left(0.2\Delta\delta_N\right)^2}{2}}$$

4.4.4 ^{15}N-Carr-Purcell-Meiboom-Gill (CPMG) NMR Relaxation Experiments

Relaxation dispersion experiments for the apo and 3'-UMP- or 5'-AMP-saturated enzyme complexes were performed using published methods (Doucet *et al.* 2009) and pulse sequences (Loria *et al.* 1999b). Experiments were carried out on 500 MHz (11.7 T) and 800 MHz (18.8 T) Varian (Agilent) NMR spectrometers equipped with a triple-resonance cold probe and pulsed-field gradients. Interleaved two-dimensional spectra were collected in a constant time manner with τ_{cp} repetition delays of 0.625, 0.714 (×2), 1.0, 1.25, 1.67, 2.0, 2.50 (×2), 3.33, 5.0, and 10 ms, within a total relaxation period of 40 ms. NMR spectra were processed using NMRPipe (Delaglio *et al.* 1995), analyzed with Sparky (Goddard and Keneller 2008) and in-house CPMG scripts. Collected data was dual-fitted to the Carver-Richards full relaxation dispersion equation (Manley and Loria 2012).

4.4.5 Definitions for (Un)Coordinated Dynamical Changes

Residues which showed a difference between measured R_2 ($1/\tau_{cp}$) values at fast (τ_{cp} = 0.625 ms) and slow (τ_{cp} = 10 ms) refocusing pulse delays greater than 2 s^{-1} were considered for further analysis, similar to previous studies (Gagné *et al.* 2012, 2015b). We compared these residues which show relaxation dispersion curves with $\Delta R_2 > 2$ s^{-1} for the 3'-UMP- and 5'-AMP-bound states with that of the apo state to identify residues displaying (un)coordinated dynamical changes. Residues that show a gain (or loss) of millisecond exchanges in both the 3'-UMP- and 5'-AMP-bound states relative to the apo state are defined as displaying coordinated changes in motions. Residues displaying a gain (or loss) of dynamics in only one of the two (5'-AMP or 3'-UMP) ligand-bound states relative to the apo state are defined as displaying uncoordinated changes. We perform a qualitative comparison of the residues displaying (un)coordinated changes in dynamics by comparing residues that show loss (or gain) of conformational exchange in either or both ligand-bound states relative to the apo state of the enzymes, respectively. Consequently, the current interpretation reports on qualitative changes in relaxation dispersion profiles between enzyme states and does not presume to quantitatively describe changes in the sign and direction of chemical shifts between two similar relaxation dispersion profiles.

4.4.6 Chemical Shift Projection Analysis (CHESPA)

CHESPA was performed based on the protocol described by Selvaratnam *et al.* (2012). The chemical shift perturbations ($\Delta\delta_{obs}$) of the ligand-bound state relative to the apo state correspond to the shifts for the highest enzyme:ligand molar ratios for each enzyme. For each of the five RNases, residues with a chemical shift variation $\Delta\delta_{obs}$ > 0.05 ppm were selected for further analysis. The two CHESPA parameters, projection angle (cos(θ)) and fractional shift (X), were calculated according to the equation below using the ^{1}H and ^{15}N peak coordinates in their free (apo) form and upon binding to the two ligands, 3'-UMP (vector A) and 5'-AMP (vector B) at saturation conditions.

Équation VII

$$\cos\theta = \frac{\vec{A}\cdot\vec{B}}{\left|\vec{A}\right|\cdot\left|\vec{B}\right|}, \quad X = \frac{\vec{A}\cdot\vec{B}}{\left|\vec{B}\right|^2}$$

4.4.7 Bioinformatics Analyses

Multiple sequence alignment of the bovine and Hominidae RNases 1–8 sequences in fasta format was performed using ClustalΩ (Sievers *et al.* 2011). Phylogenetic analysis of all RNases was performed using the maximum likelihood approach with RaxML v8.0.26 and the WAG amino acid substitution model (Stamatakis 2014). Bootstrapping analysis over 100 iterations was used to assess the reliability of the branching. Figtree, v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/) was used to visualize the phylogenetic tree input in newick format.

4.5 Results

4.5.1 Evolutionary Relationship Between Sequences

Phylogenetic analyses provide important insights into the evolutionary determinants of the structural and functional diversity within an enzyme family. Here, we performed the phylogenetic classification of the eight canonical RNase subtypes from bovine and Hominidae members (Figure 4.2). Phylogenetic clustering led to grouping of these sequences into distinct subfamilies, whereby sequences within subfamilies share similar biological functions, consistent with previous observations (Narayanan *et al.* 2018a). The eosinophil-associated RNase-like (EAR-like) homologs in human were shown to display antiviral and antimicrobial activities, while

angiogenins were named after their role in angiogenesis (Koczera *et al.* 2016). Human RNase 4 was shown to be expressed in host-defense associated tissues, while the primary function of RNase A-like sequences involves the degradation of RNA (Koczera *et al.* 2016). RNases selected for this study share an average pairwise sequence identity of 37% and are identified in blue in Figure 4.2A.

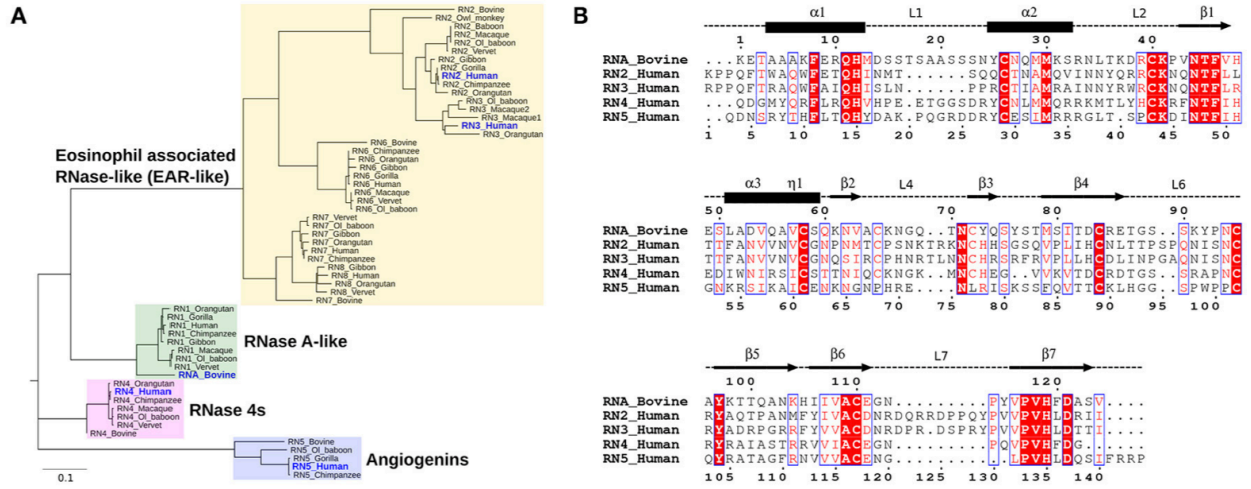


Figure 4.2 Sequence characterization of pancreatic-type RNases

(A) Phylogenetic clustering of Hominidae and bovine RNases 1–8 groups RNases into four subfamilies. The multiple sequence alignment of all RNases used for this analysis is shown in Figure 7.8. (B) Multiple sequence alignment of bovine RNase A (RNA_Bovine), human RNases 2 (RN2_Human), 3 (RN3_Human), 4 (RN4_Human), and 5 (RN5_Human). The primary sequence numbering for RNase A and the consensus sequence, corresponding to sequence numbering including gaps in the alignment, are shown in the top and bottom of the alignment, respectively. Secondary structure of bovine RNase A is traced on top of the sequence alignment using blocks, arrows and dashed lines to represent α-helix, β-strand and loop regions, respectively. Sequence alignments were prepared using EsPript (Robert and Gouet, 2014).

We selected five representative members from the four functionally distinct phylogenetic subfamilies to characterize the effect of ligand binding on the structural and dynamical properties of these functionally distinct RNases. The selected enzymes are bovine RNase A (RNA_Bovine), human RNase 2 (RN2_Human), RNase 3 (RN3_Human), RNase 4 (RN4_Human) and RNase 5 (RN5_Human). Multiple sequence alignment of the selected sequences showed the conservation of residues in the active site (His12, Lys41 and His119, RNase A numbering) and other residues associated with substrate binding and discrimination in RNase A, including Thr45 (B_1 pyrimidine binding subsite), Asn71 (B_2 purine binding subsite) and phosphate binding subsites Gln11, Asp121 (Figure 4.2B). Other ligand binding subsites identified in RNase A, such as Lys7, Arg10, Lys66, and Asp83 (Raines 1998), are not

conserved across other RNases, suggesting a potential effect on the substrate specificity and affinity in these RNases.

4.5.2 Chemical Shift Changes Associated With Ligand Binding

NMR chemical shifts are sensitive reporters of changes in the chemical environment of atoms and can be used to probe structural and conformational changes associated with ligand binding (Wishart *et al.* 1991; Case 1998). This change in chemical shift can be detected as a shift in the peak of affected residues in the ^{1}H-^{15}N HSQC spectrum of a protein. Here, we performed NMR chemical shift titration experiments with increasing concentrations, up to saturation, of two ligands, 3'-UMP and 5'-AMP, to compare and characterize the effect of ligand binding on the

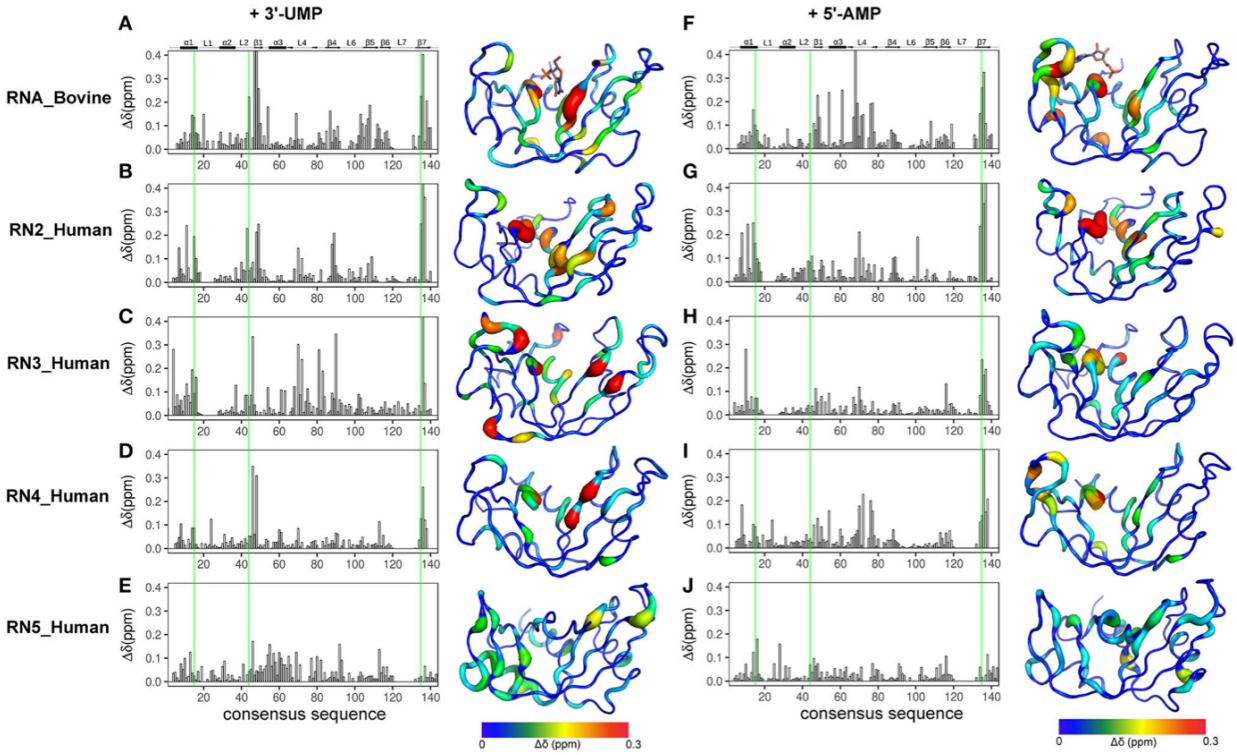


Figure 4.3 Effect of 3'-UMP and 5'-AMP ligand binding on functionally distinct RNases

Compounded chemical shift perturbations ($\Delta\delta_{obs}$ at the highest ligand concentration) relative to the apo form upon binding of two mononucleotides 3'-UMP (A–E) and 5'-AMP (F–J) for bovine RNase A (A,F) and human RNases 2 (B,G), 3 (C–H), 4 (D,I), and 5 (E,J) are plotted as function of consensus sequence. Active-site residues (His12, Lys41, His119, RNase A numbering) are highlighted using green lines. Chemical shift perturbations were calculated by comparing the shifts at the largest enzyme:ligand molar ratios relative to the apo state for each enzyme. $\Delta\delta_{obs}$ are shown using the putty representation on the 3D structures to the right of the plots. Secondary structure of bovine RNase A is traced on top of the sequence alignment using blocks, arrows and dashed lines to represent α-helix, β-strand and loop regions, respectively. The 3'-UMP and 5'-AMP ligand positions, depicted on the structures of RNase A (A,F), were obtained from PDB entries 1O0N and 1Z6S, respectively.

conformational properties of the five RNases described above. The pyrimidine and purine mononucleotides are known to bind, respectively, to the B_1 and B_2 subsites in RNase A. Figure 4.3 shows the compounded chemical shift changes ($\Delta\delta_{obs}$ at the highest ligand concentration) upon binding of 3'-UMP (Figures 4.3A–E) and 5'-AMP (Figures 4.3F–J) for the five RNases as a function of the consensus sequence. The consensus sequence offers easier comparison of proteins with different sequence lengths by including gaps in the multiple sequence alignment. Chemical shift perturbations are also represented using a rainbow color scheme on the three-dimensional structures, shown to the right of the plots. Residues displaying large chemical shift perturbations ($\Delta\delta_{obs}$ > 0.1 ppm at the highest ligand concentration) are highlighted as spheres on the 3D structures of the different RNases in Figure 7.7. A list of residues displaying $\Delta\delta_{obs}$ > 0.1 ppm for all RNases is provided in Table 7.3.

Table 4.1 Binding affinities of functionally distinct RNases for RNA and DNA ligands

	K_d 3'-UMP (μM)	K_d 5'-AMP (μM)	K_d dATATA (μM)[a]
RNA_Bovine	9.7 ± 0.9[b]	124 ± 1[b]	N/A[c]
RN2_Human	455.9 ± 10.6	539.6 ± 31.4	383.9 ± 33.3
RN3_Human	460 ± 100	300 ± 45	302 ± 31
RN4_Human	10320 ± 622	1324 ± 72	72.9 ± 9
RN5_Human	1942 ± 105	2476 ± 77	N/A[c]

[a]Single-stranded penta-deoxyribonucleotide (DNA) substrate analog of adenine and thymine nucleotide bases; [b]Taken from Gagné *et al*. (2012); [c]Not available

Binding of 3'-UMP (Figures 4.3A–E) resulted in notable differences in the residues affected upon binding to the functionally distinct RNases. Large chemical shift perturbations were observed primarily near the B_1 pyrimidine-binding subsite in most RNases, confirming the existence of an RNase A-like pyrimidine subsite in homologous RNases. RNase A (RNA_Bovine) displayed the largest 3'-UMP-induced chemical shift changes for residues near Thr45, His119 and Lys41, which directly interact with this ligand. Additional large perturbations ($\Delta\delta_{obs}$ > 0.1 ppm at the highest ligand concentration) were observed for some residues far from the active site (Figure 7.7), including residues of the β5 strand (consensus sequence positions 104–111) and Leu51 (α3), suggesting long-range conformational changes or motions in these regions upon ligand binding. The significant effect of 3'-UMP is consistent with the high binding affinity determined for RNase A (Table 4.1). RNase 2 (RN2_Human) showed fewer residues displaying large perturbations (15 residues with $\Delta\delta_{obs}$ > 0.1 ppm at the highest ligand concentration) relative to RNase A (21 residues with $\Delta\delta_{obs}$ > 0.1 ppm at the highest ligand concentration), consistent with the weaker binding affinity of the ligand for this enzyme relative to RNase A (Table 4.1). In addition to the chemical shift variations near the B_1 subsite, RNase 2

also showed long-range effects of ligand binding for residues Ser64 and Lys66 in L4 and Cys37 of L2, suggesting potential conformational rearrangements triggered by 3'-UMP binding in these loop regions. RNase 3 (RN3_Human) showed perturbations dispersed throughout the protein with large chemical shift variations in loops L4 (positions 67–76) and L5 (positions 80–83), regions that were minimally perturbed in the other RNases upon 3'-UMP binding. The catalytic residues Lys38 and His128 showed no significant perturbations upon 3'-UMP binding, consistent with the lower dissociation constants determined here (Table 4.1) and reported previously for this enzyme (Gagné *et al.* 2012). The disperse but large ($\Delta\delta_{max}$ > 0.1 ppm at the highest ligand concentration) chemical shift perturbations, in addition to the observation that the displacement of some resonances upon 3'-UMP binding does not follow a straight line, suggests that this enzyme may either be experiencing conformational exchange in distal regions upon ligand binding and/or binding ligand in more than one site (Williamson 2013). RNase 4 (RN4_Human), which experiences important and broadly distributed conformational exchange on the millisecond timescale in its apo form (Narayanan *et al.* 2017, 2018), showed few chemical shift perturbations upon ligand binding. As expected, these perturbations are primarily localized to the B_1 pyrimidine binding site, correlating with a very low affinity for 3'-UMP (Table 4.1). The lack of significant perturbations is in agreement with the very low affinity of RNase 4 for 3'-UMP (K_d = 10,320 ± 622 μM). Interestingly, RNase 4 shows a much higher binding affinity for the oligonucleotide ligand dATATA (K_d = 72.9 ± 9 μM). We note that RNase 4 is the only member for which we observe such distinct binding affinities between single-nucleotide mimics of reaction products and a pentanucleotide DNA substrate analog. RNase 5 (RN5_Human) showed no effect in the B_1 subsite while perturbations were observed far from the active site, suggesting weaker affinity and non-specific binding of the substrate, observations consistent with a previous report (Gagné *et al.* 2015b).

Titration with 5'-AMP resulted in perturbations of residues near the purine binding site (B_2) in all RNases, except RNase 5 (Figure 4.3F–J). Specifically, RNases A, 2 and 4 displayed significant perturbations in L4, residues of which were expected to interact with the purine base (Figure 7.7). RNase 2 showed additional perturbations in the N-terminal helix α1. RNase 3 showed fewer perturbed residues upon 5'-AMP binding relative to 3'-UMP binding, with residues localized primarily to the B_1 subsite. The effects of ligand binding in loop 4 were diminished in RNase 3, with fewer residues and smaller magnitudes of perturbations observed upon ligand binding. RNase 4 showed large perturbations primarily in the L4 and the C-terminal regions. In contrast, RNase 5 showed the smallest perturbations among all RNases tested, with no changes in its significantly truncated loop 4, and very few residues displaying large chemical

shift perturbations ($\Delta\delta_{max}$ > 0.1 ppm at the highest ligand concentration) throughout the protein upon ligand binding (Figure 7.7), suggesting few interactions with the ligand. These observations are in agreement with the low binding affinity determined for this enzyme (Table 4.1), and are consistent with previous observations (Gagné *et al.* 2015b). Overall, our results highlight the distinctive effects of ligand binding on the local chemical environment and/or potential changes in dynamic behavior for these different RNases. These observations illustrate that while all RNases share a similar structural fold, they display distinctly different conformational rearrangements upon binding of the two single-nucleotide RNA product mimics, further confirming their affinity differences and potential functional specialization.

4.5.3 Comparative Chemical Shift Analysis for RNases

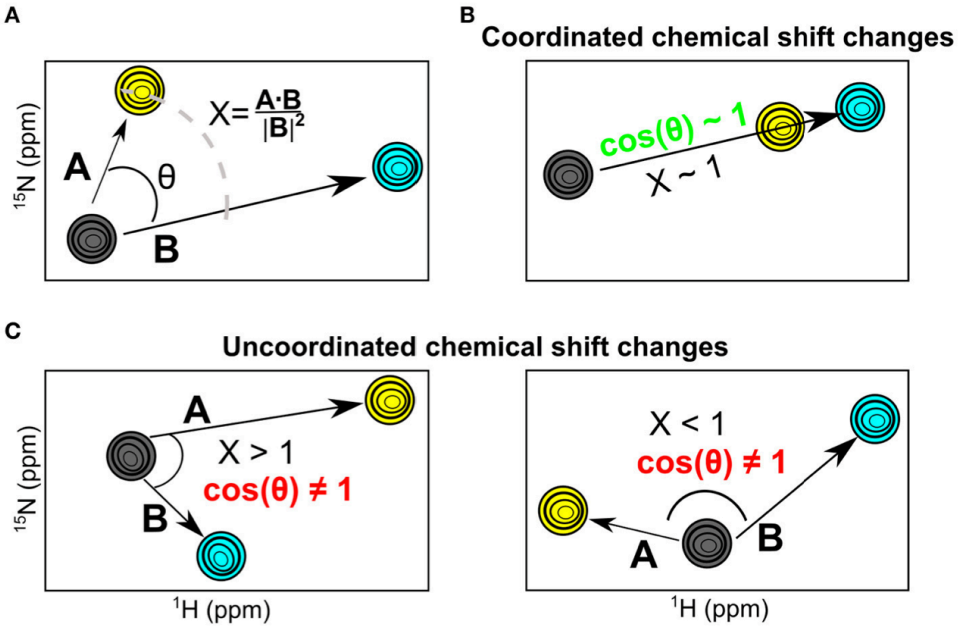


Figure 4.4 The NMR chemical shift projection analysis (CHESPA)

(A) Schematic representation of the CHESPA analysis showing ^{1}H-^{15}N HSQC chemical shift resonances of the apo (gray), 3'-UMP (yellow) and 5'-AMP (cyan) states. Vectors A, B correspond to the compounded chemical shifts calculated for 3'-UMP-bound and 5'-AMP-bound complexes relative to the apo state. The two projection analysis parameters, projection angle and fractional shift, were calculated as described in the section Methods. (B,C) Graphical representation of scenarios representing coordinated (B) and uncoordinated displacements (C) of the ^{1}H-^{15}N HSQC chemical shifts upon ligand binding.

While chemical shift perturbations provide interesting information on local and long-range effects triggered by ligand binding to a protein, they nevertheless have limited comparative

value when juxtaposing the effects of one system relative to another. In contrast, the chemical shift projection analysis (CHESPA) provides a systematic characterization of the effect of ligand binding or other perturbations on the magnitude and direction of the ^{1}H-^{15}N HSQC chemical shift displacements (Selvaratnam *et al.* 2012; Axe and Boehr 2013; Gagné *et al.* 2015b). By characterizing the effect of two ligands that bind to distinct nucleotide subsites in the active site, we previously showed that this powerful technique can provide important insights into the long-range structural and dynamical changes associated with ligand binding (Gagné *et al.* 2015b). Here, we use the chemical shift projection analysis (CHESPA) to probe the coordinated changes associated with 3'-UMP and 5'-AMP binding to the five selected RNases. The schematic representation of the CHESPA approach is shown in Figure 4.4A. The projection angle (θ) is defined as the angle between the chemical shift peak displacements (for residues with $\Delta\delta_{obs} > 0.05$ ppm at the highest ligand concentration) upon binding 3′-UMP (**A** vector) and 5'-AMP (**B** vector) relative to the apo state. The fractional shift (X) corresponds to the magnitude of displacement of **A** relative to **B**. Chemical shift changes are defined as coordinated when $\cos(\theta) \geq 0.9$ and the chemical shift perturbations upon binding of the two ligands is similar, and uncoordinated when $\cos(\theta) < 0.9$ (Figures 4.4B,C).

The projection angle and fractional shift as a function of the consensus sequence determined for each of the five RNases is shown in Figure 4.5. For RNase A (RNA_Bovine), most residues involved in both 3'-UMP and 5'-AMP ligand binding, including Lys66 (P_0), Thr45 (B_1) and the catalytic His12, showed coordinated displacements (Figure 4.5A). Interestingly, active-site residues His119/Lys41, and Lys7, which form the P_1 binding site, alongside Thr17 (loop 1) and residues of the C-terminal β7 strand showed uncoordinated displacements of the chemical shifts, indicating the different effects upon binding of the two ligands (Figure 4.5A). RNase 2 (RN2_Human) showed residues with both coordinated and uncoordinated displacements in the ligand binding sites, with residues of loop L4 displaying uncoordinated displacements (Figure 4.5B). In contrast to other RNases, RNase 3 (RN3_Human) showed predominantly uncoordinated displacements throughout the protein, with coordinated displacements observed only for Gln4 and the active site His15, suggesting a distinctly different effect of ligand binding in this enzyme (Figure 4.5C), consistent with the unusual behavior of 3'-UMP binding described above (Figure 4.3C). RNase 4 (RN4_Human) showed coordinated chemical shift displacements in the active site, while uncoordinated displacements were observed in the C-terminal β7 strand, similar to that of RNase A (Figure 4.5D). Interestingly, most residues displaying coordinated displacements are localized to the V_2 domain of RNase 4. RNase 5 (RN5_Human) showed

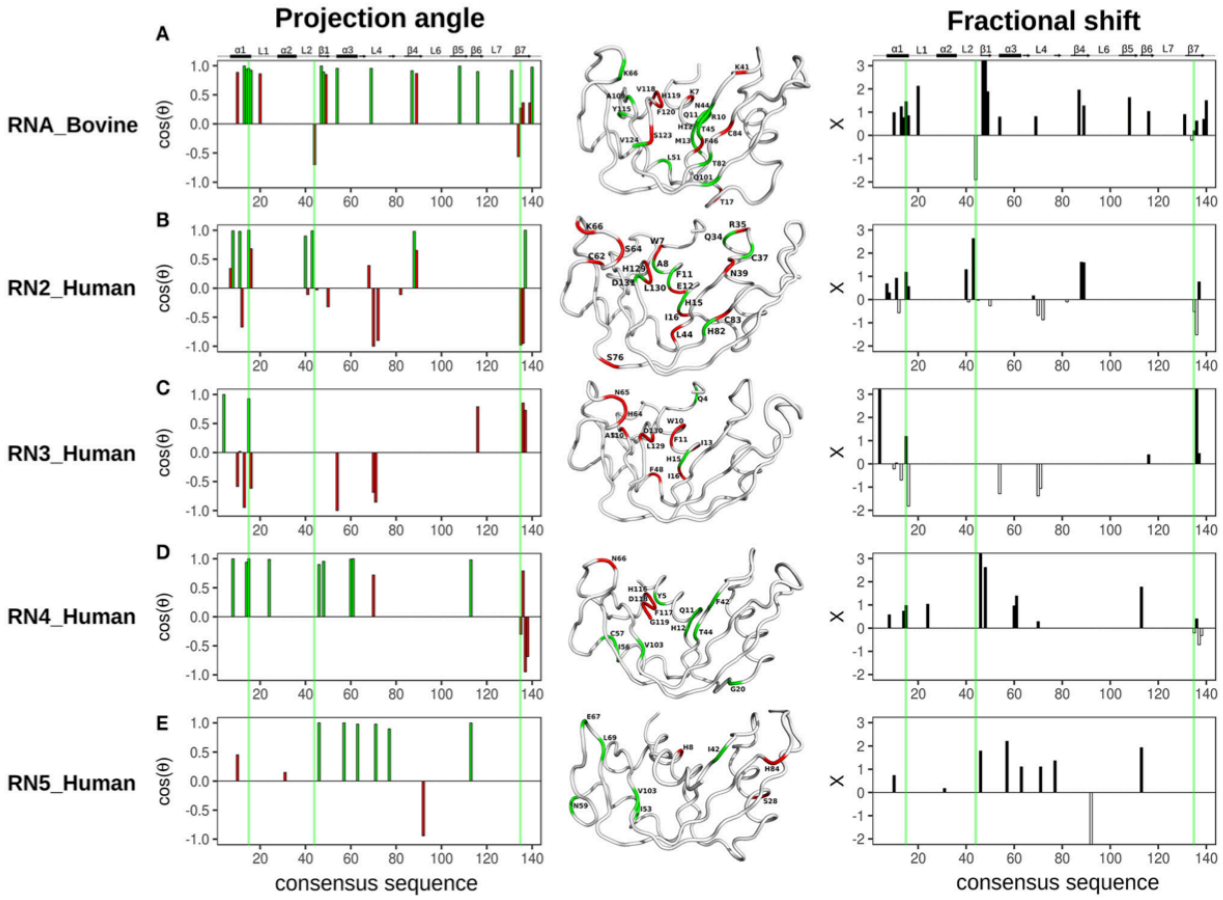


Figure 4.5 **Chemical shift projection analysis of 3'-UMP and 5'-AMP ligand binding to functionally distinct RNases**

The projection angle (cos(θ), left) and the fractional shift (X, right), calculated for residues with $\Delta\delta_{obs}$ > 0.05 ppm at the highest ligand concentration, are plotted as a function of consensus sequence for bovine RNase A and human RNases 2–5 upon binding of 3'-UMP and 5'-AMP single nucleotide RNA ligands. (A–E) Residues showing cos(θ) ~ 0.9 are colored green (coordinated), those with cos(θ) < 0.9 are colored red (uncoordinated). Residues are also identified on the three-dimensional structures of the five RNases using tube representations in the color scheme described above. Fractional shifts as a function of consensus sequence for bovine RNase A and human RNases 2–5, respectively, are shown on the right. Residues displaying positive and negative fractional shifts are displayed using black and white bars, respectively. Active-site residues (His12, Lys41, His119, RNase A numbering) are highlighted using green lines. Secondary structure of bovine RNase A is traced on top of the sequence alignment using blocks, arrows and dashed lines to represent α-helix, β-strand and loop regions, respectively.

coordinated chemical shift displacements predominantly in the V_1 domain, consistent with previous observations (Gagné *et al.* 2015b). Uncoordinated displacements were observed for residues His8, Ser28 and His84 (Figure 4.5E). Residues displaying coordinated chemical shift displacements also showed fractional shift ≥ 1, suggesting similar or larger changes in magnitude upon binding of 3'-UMP relative to 5'-AMP in all enzymes. Overall, our results

suggest that while most of the amino acid residues that comprise the ligand binding pocket are identical in the primary and 3D structures of the selected RNases (Figure 4.2B), subtle variations in the active-site environment can lead to significantly different long-range effects throughout the structures upon ligand binding.

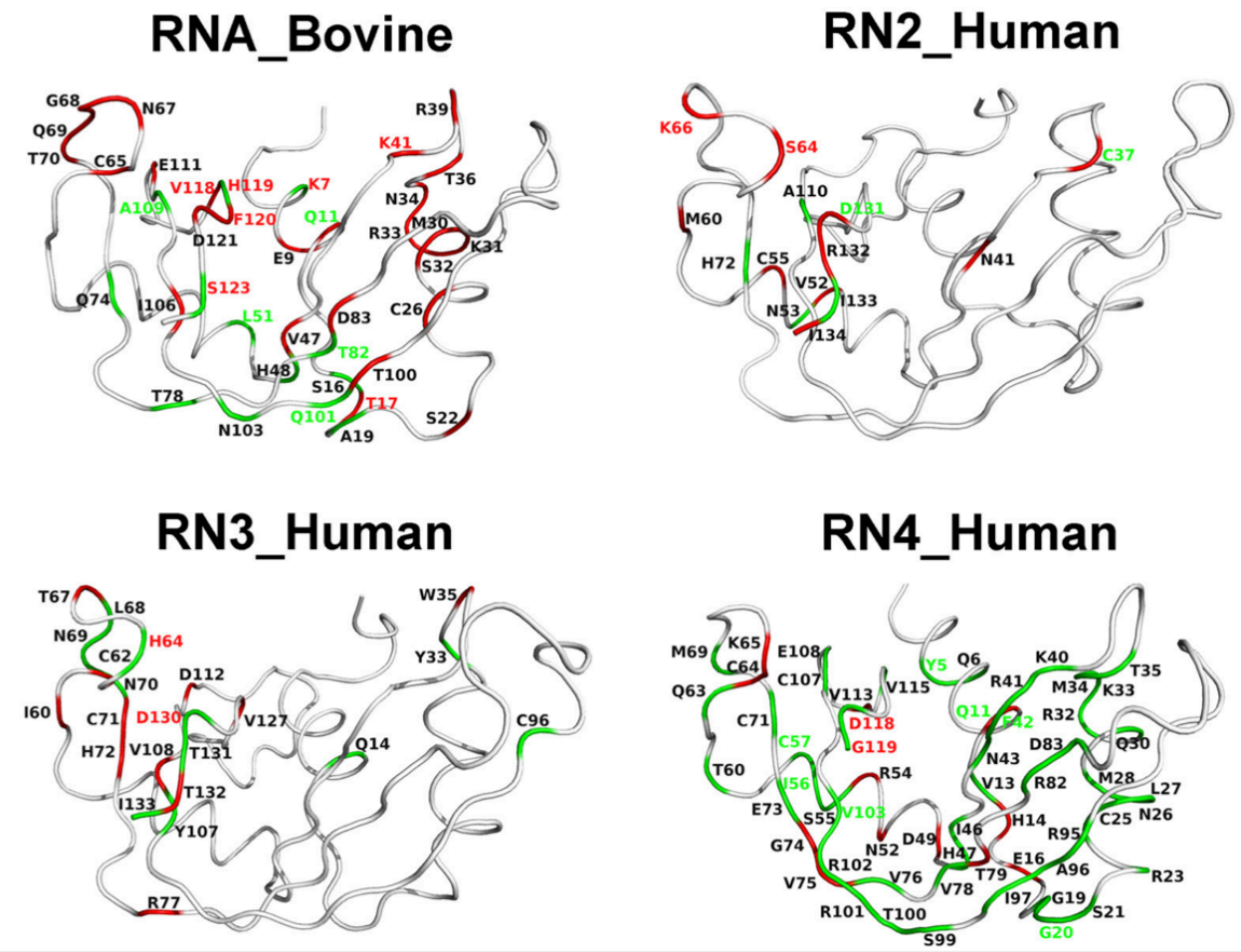


Figure 4.6 **Coordinated conformational exchange and chemical shift perturbations in functionally distinct RNases upon ligand binding**

Millisecond timescale dynamics was probed using ^{15}N-CPMG relaxation dispersion experiments following binding of 3'-UMP and 5'-AMP to bovine RNase A (RNA_Bovine) and human RNases 2 (RN2_Human), 3 (RN3_Human), and 4 (RN4_Human). RNase 5 (RN5_Human) is not shown, as no measurable relaxation dispersion effects were observed upon ligand binding to this enzyme on this timescale. Residues displaying changes in conformational exchange upon binding (relative to the apo state) for both 3'-UMP and 5'-AMP are shown in green, while residues that show dynamical changes upon binding of only one of the two ligands are shown in red (see section Methods). Dynamical residues that also display coordinated or uncoordinated displacements in the chemical shift projection analysis (Figure 4.5) are highlighted using green and red residue labels, respectively.

Chemical shift perturbations report on changes upon ligand binding which may arise as a consequence of local/global structural rearrangements and/or changes in conformational

motions in proteins. To gain insights into the potential role of dynamical changes in the observed chemical shift perturbations upon ligand binding, we used NMR ^{15}N-CPMG relaxation dispersion experiments to identify residues that exhibit similar (or different) dynamical behavior upon binding of each of the two mononucleotide ligands (Figure 4.6). We note that RNase 5 (RN5_Human), which shows no measurable relaxation dispersion effects in the free and ligand-bound states, is not shown in Figure 4.6. Dynamical change is defined as the gain or loss of millisecond motions of any given residue, which displays a relaxation dispersion curve and $\Delta R_2 > 2$ s^{-1} (Figure 7.9), upon ligand binding relative to the apo state (see section Methods). Residues that display dynamical changes upon binding of both 3'-UMP and 5'-AMP ligands are shown in green, while residues which show dynamical changes upon binding of only one of the ligands are colored red. Further, residues displaying conformational exchange on the millisecond timescale that also show coordinated (uncoordinated) displacements in the chemical shift projection analysis (CHESPA) are identified using green (red) residue labels. Our results show that residues experiencing conformational exchange form significantly distinct dynamic clusters from one RNase member to the other, again supporting a link between millisecond timescale dynamics and divergent functional and biological specialization between these evolutionary distinct subfamily members. Also, while many residues display dynamical changes upon binding of the two ligands, only a subset of these residues exhibit significant chemical shift perturbations. Interestingly, most residues distal to the active site (defined as residues that are farther than 6 Å from the catalytic triad) that showed (un)coordinated CHESPA displacements (i.e., residues with red/green residue labels) also display (un)coordinated dynamical changes (red/green color of the cartoon loop) upon binding of the two ligands in the four RNases. For example, residues Thr17, Leu51, Thr82, and Gln101 in RNase A (RNA_Bovine); Lys66 of RNase 2 (RN2_Human); Tyr5, Gly20, Ile56, Cys57, and Val103 of RNase4, which are distal from the active site, show (un)coordinated perturbations in chemical shifts and similar (un)coordinated dynamical changes upon binding the two ligands. This trend is not observed for some residues in the vicinity (< 6 Å) of the catalytic triad, such as Lys7, Gln11 in RNase A, Cys37, Asp131 in RNase 2; Asp130 in RNase 3; and Glu11, Asp118, and Gly119 in RNase 4. Exceptions to this trend include Ser123 of RNA_Bovine and His64 of RN3_Human, that display opposite trends to those described above for residues distal from the catalytic triad. We also note that the projection analysis is sensitive to changes in the ^{1}H and ^{15}N chemical shift changes while the dynamical changes are sensitive to the ^{15}N chemical shift changes. Consequently, our analysis provides a qualitative comparison of the conformational exchange of residues between the apo and ligand-bound states. Nevertheless, these observations suggest that dynamical

changes on functionally relevant timescales may be contributing to the coordinated displacements of distal residue networks in RNases. While further experiments are required to confirm this hypothesis, the qualitative comparison of the changes in the conformational exchange of residues presented in this study indicate that these events might correlate with long-range allosteric effects controlling the biological function of specific RNase subfamily members.

4.6 **Discussion**

The role of conformational dynamics for optimal enzyme catalysis has emerged for a variety of enzyme systems (Narayanan *et al.* 2016). In their ligand-free states, several enzymes were shown to sample conformations observed upon ligand binding, suggesting that enzymes are primed for their catalytic function (Henzler-Wildman *et al.* 2007b; Holliday *et al.* 2017; Kovermann *et al.* 2017). Ligand binding induces chemical shift changes in the protein, which may arise from changes associated with direct interaction with the ligand or conformational changes (dynamic and allosteric) induced by ligand binding (Williamson 2013). While ligand-induced effects on structural and dynamical changes have been characterized for discrete enzymatic systems (Dhulesia *et al.* 2008; Gagné *et al.* 2015a; Doshi *et al.* 2016; Goricanec *et al.* 2016), these remain largely uncharacterized for most members within a superfamily. As a result, comparison of the effects of ligand binding between different members within a family that display distinct ligand specificities and catalytic efficiencies is lacking.

In this study, we combined chemical shift titration experiments with the chemical shift projection analysis and relaxation dispersion experiments to identify networks of amino acids displaying coordinated displacements upon binding of two mononucleotide ligands. The chosen ligands mimic cleavage products of a UpA dinucleotide RNA substrate and are known to bind to different nucleotide binding sites in bovine RNase A (Figure 4.1) (Nogués *et al.* 1995). Using select RNases corresponding to representative members of distinct phylogenetic subfamilies, we characterized the effect of binding and conformational exchange induced by these ligands to gain insights into the mechanism and functional biological differences observed in these structurally homologous RNases. Our results show that ligand binding induces different local and long-range effects on the various RNases, even for enzymes within the same subfamily, suggesting distinct conformational rearrangements upon ligand binding that may not be conserved in closely related RNases.

The binding affinity for a ligand, which relies on the subtle balance between substrate selectivity and avoiding being trapped with the bound ligand, among other factors, influences the catalytic efficiency of enzymes (Kovermann *et al.* 2017). Chemical shift perturbations dispersed far from the ligand binding site may be attributed to conformational rearrangements in the protein, assuming that a ligand binds to a single binding site (Williamson 2013). Our results show that binding of the two mononucleotide ligands to the selected RNases triggers changes in distal regions in some of these enzymes. For example, binding of 3'-UMP induced effects in loops 4 and 5 and other regions distal from the active site in RNase 3, in addition to the non-linear chemical shift change upon addition of increasing concentrations of 3'-UMP, consistent with results from previous studies (Gagné *et al.* 2012). These observations suggest either conformational changes induced by ligand binding and/or binding to more than one site (Williamson 2013). In contrast, RNase 4 showed very few perturbations, with residues displaying any significant changes localized to the pyrimidine binding site (Figure 4.3). These observations are consistent with the low affinity observed for this enzyme (K_d = 10,320 ± 622 μM for 3'-UMP). Interestingly, the low affinity for this 3'-UMP product analog does not correlate with the high binding affinity of a pentanucleotide single-stranded DNA substrate mimic (dATATA) to RNase 4 (Table 4.1). Considering the relatively low binding affinities of single-nucleotide product analogs, this observation suggests the existence of favorable synergistic subsite binding energies in the context of longer oligonucleotide substrates, and/or minimally that the 5-methyl moiety on the thymine pyrimidine DNA base provides favorable interactions with the enzyme that are absent in the homologous uracyl RNA nucleotide.

Amino acid residues belonging to the same allosteric network were previously shown to display coordinated chemical shift changes upon ligand binding or other perturbations (Selvaratnam *et al.* 2012). Using the CHESPA approach, we previously identified a network of amino acid residues involved in maintaining the structural stability of human angiogenin (RNase 5), highlighting long-range coordinated perturbations upon ligand binding. Here, we identified networks of amino acid residues that display coordinated chemical shift displacements upon binding of the two ligands in select RNases (Figure 4.5). Our results show that not all RNases display the same coordinated amino acid perturbations previously shown in RNase 5. Further, for RNases displaying coordinated displacements, the amino acid networks exhibiting such displacements are different. For example, while residue networks exhibiting coordinated displacements (defined in Figure 4.4B) in RNase A are localized primarily near the P_2 and B_1 binding sites (Figure 4.1), other RNases display significant differences in the localization of residues exhibiting coordinated changes upon binding of the two ligands. This lack of similarity

in amino acid networks of coordinated displacements, even within closely related sequences such as RNases 2 and 3, suggests that the different RNases are uniquely adapted to their respective functions. We speculate that this may arise due to a variety of factors, including distinct nucleotide specificities. Further experiments probing the effect of binding a variety of nucleotides will provide insights into the contribution of nucleotide specificity to the observed effects.

In a recent study published from our group, we showed that the conservation of dynamical properties among RNase homologs correlates with their evolutionary conservation and shared biological functions (Narayanan *et al.* 2018a). This previous work provided important insights into the role of conserved dynamical properties beyond enzyme catalysis. However, that study was performed on ligand-free states of enzymes and thus did not outline potential allosteric paths or conserved dynamic networks specific to subfamily members in the presence of ligands. Binding of a ligand can induce conformational rearrangements that may contribute to the observed chemical shift perturbations. In this work, we aimed to gain insights into the contribution of the conformational motions on the observed chemical shift perturbations by comparing CHESPA results with the coordinated changes in conformational exchange probed by ^{15}N-CPMG relaxation dispersion experiments on the micro-millisecond timescale. The current study represents an additional step toward achieving the overarching goal of correlating the differences in dynamic/allosteric responses of distinct RNases with their diverse chemical and biological functions. Addressing this question on a broader functional and evolutionary scale requires the development of protein-specific allosteric modulators and/or mutagenesis studies that perturb the allosteric behavior of a targeted enzyme by directly affecting its ribonucleolytic activity or perturbing the other biological function in an independent fashion.

The present work illustrates a direct link between conformational perturbations and chemical shifts induced by ligand binding among structurally similar RNase homologs. Our results show that while not all residues displaying conformational exchange on the catalytically relevant millisecond timescale are reflected in the observed chemical shift perturbations, most residues displaying perturbations also exhibit coordinated dynamical changes (Figure 4.6). These observations highlight the contributions to observed chemical shift changes that arise from conformational rearrangements occurring upon ligand binding. Overall, these observations suggest that ligand-induced effects may be influenced, among other factors, by a combination of the different binding affinities and conformational dynamics of the different RNases. Further

experiments are necessary to gain insights into the individual contribution of each factor to the observed effects on ligand binding in these functionally distinct RNases.

4.7 Acknowledgements

We thank Tara Sprules of the Québec/Eastern Canada High Field NMR Facility (McGill University) and Sameer Al-Abdul-Wahid (University of Guelph) for their technical assistance. This work was supported by a NIGMS/NIH grant under award number R01GM105978 (to ND and PKA), a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant under award number RGPIN-2016-05557 (to ND), an NSERC Postgraduate Doctoral Scholarship (to DNB), a Postdoctoral Fellowship from the Fondation Armand-Frappier (to CN), and a Fonds de Recherche Québec—Santé (FRQS) Research Scholar Junior 2 Career Award (number 32743, to ND).

5 QUATRIÈME ARTICLE

L'article précédent a démontré que même les changements de structure et de conformations associés à la catalyse enzymatique (fonction strictement conservée chez les RNases actives) n'étaient pas conservés chez les différentes RNases humaines. Ceci pourrait signifier que la fonction catalytique puisse ne pas être considérée seule dans une enzyme, mais que l'enzyme doit être prise dans son ensemble afin d'expliquer ses comportements. Encore une fois, la différence dans les fonctions biologiques des RNases humaines s'impose comme première cause possible pour les différences observées lors des travaux précédents. La solution logique est donc d'aller explorer une zone plus restreinte de l'espace des séquences, en s'intéressant à des protéines possédant des fonctions biologiques similaires. Est-ce que des enzymes ayant une structure, une fonction catalytique et des fonctions biologiques similaires auront des mouvements atomiques similaires ? L'article présentement intitulé « **Divergence of conformational motions along the evolutionary pathway in the eosinophil ribonuclease subfamily** » décrit les recherches associées à ce questionnement.

Cet article écrit par **David N Bernard, Chitra Narayanan, Tim Hempel, Khushboo Bafna, Purva Prashant Bhojane, Myriam Létourneau, Elizabeth E Howell, Frank Noé, Pratul K Agarwal et Nicolas Doucet** est en cours de préparation dans le but d'une soumission au journal Structure. J'en suis le premier auteur, et l'article a été rédigé par Chitra Narayanan, Tim Hempel et moi-même, à l'exception des sections sur la cytotoxicité et sur la cinétique enzymatique qui ont été rédigées par Myriam Létourneau et Khushboo Bafna, respectivement. J'ai effectué toutes les expressions et purifications de protéines, les tests antibactériens contre *E. coli* et *S. aureus*, les modélisations par homologie, les analyses de similarité cosinus, ainsi que toutes les expériences RMN à l'exception de quelques résultats de CPMG effectués par Donald mais que j'ai réanalysés. Ceci a donc impliqué l'attribution de quatre protéines dont j'ai déposé la liste des résonances sur la BMRB, ainsi que la correction d'une attribution précédemment déposée mais qui s'est avérée erronée (Gagné *et al*. 2018). Chitra Narayanan a effectué les simulations bioinformatiques de dynamique moléculaire, les calculs de $RMSF_{10}$, ainsi que l'analyse phylogénétique. Tim Hempel a calculé et analysé les nombreuses simulations par modèles de Markov. Myriam Létourneau s'est occupée de l'exécution et de l'analyse des expériences de cytotoxicité sur les cellules HeLa. Khushboo Bafna et Purva Prashant Bhojane ont mesuré l'activité catalytique de chacune des RNases et analysé les données associées. Le projet a été supervisé par Elizabeth Howell, Frank Noé, Pratul Agarwal et Nicolas Doucet.

5.1 Résumé

Des avancements récents en ingénierie des protéines ont révélé l'importance des échanges conformationnels dans l'apparition de nouvelles fonctions. Par contre, la connaissance des échanges conformationnels n'englobe qu'un nombre fini de protéines distinctes aux structures et fonctions distinctes. Afin de mieux comprendre les phénomènes d'échanges structurels, nous avons effectué la caractérisation extensive de ces événements au sein de la superfamille des ribonucléases A. Les RNases fonctionnellement distinctes démontrent une diversité dans leur flexibilité, mais peu d'études se sont intéressées à des protéines ayant une fonction similaire. Nous avons cherché à comprendre la flexibilité conformationnelle à l'intérieur d'une seule sous-famille de RNases, les RNases d'éosinophiles. Nous avons caractérisé les fonctions biologiques de différents membres de la sous-famille grâce à des expériences d'activité enzymatique, d'action antibactérienne et de cytotoxicité. En parallèle, nous avons caractérisé leurs échanges conformationnels par RMN (CPMG, CEST) et par méthodes bioinformatiques (dynamique moléculaire, modèles de Markov). Malgré une conservation des fonctions, nous démontrons que ces enzymes peuvent avoir des profils d'échanges conformationnels similaires ou différents, ce qui implique une divergence plutôt qu'une conservation de la flexibilité. Comparer des enzymes progressivement plus éloignées les unes des autres devrait révéler comment elles acquièrent de nouvelles fonctions et/ou modifient leur comportement au niveau moléculaire à travers l'évolution.

5.2 Abstract

Recent protein engineering advancements revealed the importance of conformational exchange in new function emergence. However, knowledge of conformational exchange encompasses a finite number of functionally and structurally distinct proteins. To better understand structural exchange phenomena, we extensively characterized these events in the ribonuclease A superfamily. Functionally distinct RNases display diversity in their flexibility, but few studies explored proteins exhibiting similar functions. We sought to understand conformational flexibility within a single RNase subfamily, eosinophil RNases. We characterized the biological function of various subfamily members using enzymatic, antibacterial and cytotoxicity assays. In parallel, we characterized their conformational exchange using NMR (CPMG, CEST) and bioinformatic analyses (MD, MSM). Despite functional conservation, we show that these enzymes can either display similar or distinct conformational exchange profiles, implying divergence rather than

conservation of flexibility. Comparing progressively more distant enzymes should unravel how they evolve new functions and/or alter their behavior on a molecular level.

5.3 Introduction

Evidence for the importance of conformational flexibility in protein function has, over the years, led to the evolution of the paradigm of structure-function correlations to an ensemble representation of protein structures (Frauenfelder *et al*. 1988; Henzler-Wildman & Kern 2007). Conformational dynamics may occur over a wide range of timescales, from fast side-chain motions to slower loop and domain motions (Doucet et Pelletier 2011; Bhabha *et al*. 2015; Maria-Solano *et al*. 2018). These movements facilitate the sampling of conformational *substates*, which correspond to energy minima in the conformational landscape of proteins (Agarwal *et al*. 2016; Wei *et al*. 2016). Indeed, conformational dynamics on the timescale of catalytic turnover has been shown to be critical for optimal biological function in a variety of enzymes such as *E. coli* dihydrofolate reductase (Bhabha *et al*. 2011), cyclophilin A (Fraser *et al*. 2009), adenylate kinase (Kovermann *et al*. 2017), HIV-1 protease (Narayanan *et al*. 2016), DhlA haloalkane dehydrogenase (Kokkonen *et al*. 2018) and bovine RNase A (Doucet *et al*. 2011; Gagné *et al*. 2015b). Further, perturbation of the observed conformational exchange, through mutations in functionally important regions, greatly affects catalytic turnover in many of these systems (Kovrigin and Loria 2006; Fraser *et al*. 2009; Doucet *et al*. 2011; Bhabha *et al*. 2011). However, in most cases, understanding of such substate-transitioning movements is limited to discrete – and often unrelated – proteins in the sequence space, with the aforementioned examples being the most prominent ones.

One of the major challenges in identifying the role of conformational motions in function is to decouple the individual contributions from structure and dynamics. One way to address this is to characterize the conformational landscape of enzymes sharing a conserved structure, prompting the need to investigate these behaviors in closely related proteins. How are conformational dynamics conserved between related enzymes, and how does this correlate with this diversity of catalytic and biological functions? Members of the pancreatic ribonuclease (RNase) superfamily, which display a conserved structure (Figure 5.1) and catalyze the endonucleolytic cleavage of RNA (Raines 1998; Sorrentino 2010), represent ideal model systems for characterizing the role of conformational motions in function. Canonical RNases in this superfamily have experienced a high amino acid substitution rate (Cho *et al*. 2005) that allowed the various members to develop a remarkable diversity of biological roles (Rosenberg

et al. 1995; Zhang and Rosenberg 2002; Sorrentino 2010; Goo and Cho 2013; Premzl 2013), which include host defense (Gupta *et al*. 2013), angiogenesis (Leonidas *et al*. 1999), neuroprotection (Li *et al*. 2013), vascular homeostasis (Lu *et al*. 2018), and digestion (Sorrentino 2010), in addition to the core catalytic function. Recent studies have illustrated the conservation of amino acid networks in the RNase superfamily whose millisecond dynamics was correlated with enzyme catalysis (Narayanan *et al*. 2017). Further, conservation of global dynamical properties was observed for RNases grouped into phylogenetic subfamilies, where members within subfamilies share similar biological functions (Narayanan *et al*. 2018a).

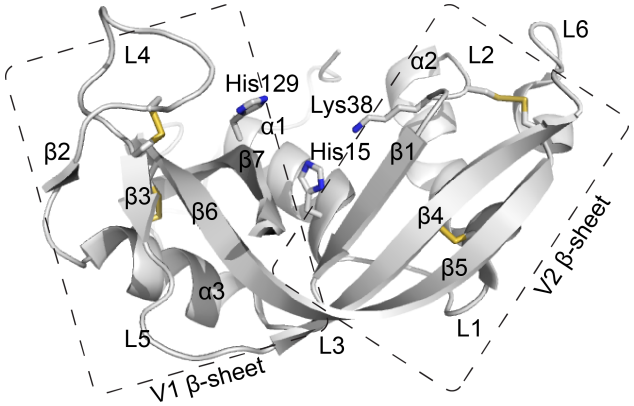


Figure 5.1 Structure of an RNase

Cartoon representation of human eosinophil-derived neurotoxin (HsR2, PDB ID 1gqv). Secondary structure elements as well as the catalytic residues, represented as sticks, are annotated on the structure.

While the phylogenetic clustering of RNases resulted in the classification of RNases with distinct biological functions (Narayanan *et al*. 2018a), there is significant diversity in the specifics of RNase catalytic and biological functions within subfamilies. For instance, members of the eosinophil associated RNase subfamily comprises RNases 2 and 3, which arose as a result of a gene duplication event that occurred when Old and New World monkeys diverged from each other (Rosenberg 2008). Human RNase 2, also known as eosinophil-derived neurotoxin (EDN), displays potent antiviral (Domachowske *et al*. 1998c; Bedoya *et al*. 2006) and chemotactic activities (Rosenberg 2015), while human RNase 3, also known as eosinophil cationic protein (ECP), displays cytotoxic (Navarro *et al*. 2008), helminthotoxic (Singh and Batra 2011) and antibacterial activities (Lehrer *et al*. 1989; Rosenberg 1995; Salazar *et al*. 2014). RNase 2 displays distinct millisecond conformational exchange characteristics and was shown to be nearly 90 times more catalytically active than RNase 3 (Rosenberg and Dyer 1997). Using

closely related enzymes of this subfamily as model systems, here we perform a systematic characterization of the conformational flexibility, catalytic and biological functions of various eosinophil RNases. We integrated ^{15}N-CPMG, ^{15}N-CEST NMR experiments and molecular dynamics simulations to quantify the conformational exchange patterns of the enzymes in apo form spanning nanosecond (ns) to second (s) timescales. We measured the kinetic parameters, antibacterial activity against Gram-positive and Gram-negative bacteria, and cytotoxicity assays against HeLa cells to characterize and compare the catalytic and biological functions of these enzymes. Our results illustrate a diversity in the conformational exchange profiles that is greater than the diversity in catalytic and biochemical activities for RNases within this phylogenetic subfamily. These observations suggest a rapid divergence of conformational exchange within the eosinophil RNase subfamily that may fine-tune the biochemical and biological functions of these enzymes along their evolutionary trajectory.

5.4 Materials and Methods

5.4.1 Phylogenetic Analysis.

RNases 1-8, comprising hominid, bovine and mouse sequences, were selected for the analysis. A list of all sequences is provided in the supporting information (Supplementary table 7.4). Multiple sequence alignment of the selected sequences was performed using Clustal Ω (Sievers *et al*. 2011). The phylogenetic tree was inferred using the maximum likelihood approach implemented in PhyML (Guindon *et al*. 2010) and the JTT substitution model. Reliability of the internal branches was estimated using aLRT test (SH-like) (Anisimova and Gascuel 2006). The phylogenetic tree was visualized using Figtree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

5.4.2 Protein Expression, Refolding, and Purification

Unlabeled RNase A was purchased from BioBasic (Markham, ON, Canada). Human RNase 3 (HsR3) gene was obtained as described previously (Gagné *et al*. 2012). *E. coli* codon-optimized genes of *Macaca fascicularis* RNase 3 (MfR3, UniProtKB entry P47779), *Pongo pygmaeus* RNase 3 (PpR3, P47781), *Pongo abelii* RNase 3 (PaR3, H2NKI8) and *Aotus trivirgatus* RNase 2 (AtR2, O18937) were synthesized by DNA2.0/ATUM (Newark, CA, USA) in pJexpress 414 vectors. Unlabeled, ^{15}N- or {^{13}C,^{15}N}-labeled samples of the proteins were produced as described previously (Boix *et al*. 1999b), with the following differences. After induction with IPTG, bacteria cultures were grown overnight at 30°C, and the refolding step was performed at

pH 9.1 for HsR2 and AtR2. Identical purification fractions were identified by circular dichroism. Protein concentration was determined using extinction coefficients of 17460, 17460, 14940, 11960, 13450 and 13450 $M^{-1}cm^{-1}$ for HsR3, HsR2, MfR3, PpR3, PaR3 and AtR2, respectively, as estimated by ProtParam.

5.4.3 Steady State Kinetics

Steady state kinetic data were obtained on a PerkinElmer λ35 UV-vis spectrophotometer (Waltham, MA, USA). Activity of RNases was measured in 50 mM sodium acetate, 1 mM EDTA at pH 5.5 and 25°C. Dinucleotide substrate UpA was obtained from Dharmacon at GE Healthcare (Lafayette, CO, USA). The concentration of UpA was measured using the extinction coefficient 24600 $M^{-1}cm^{-1}$ supplied from Dharmacon. Change in UpA absorbance was measured in 1 or 2 mm path length cuvettes at 286 nm for 10 min, enzyme and substrate concentrations used are indicated in Supplementary Table 7.5, and all measurements were performed in duplicate. Water was incubated with 1% DEPC for 1 h and autoclaved. All equipments were cleaned with DEPC-treated water before starting the experiment to inhibit the environmental contaminant RNases from interfering with the kinetic assay. Buffer was made with DEPC treated water as well. The data were fitted to the Michaelis-Menten equation using SigmaPlot.

5.4.4 Antibacterial Assays

Escherichia coli DH5α strain (Gram-negative) and *Staphylococcus aureus* Newman strain (Gram-positive) bacteria were grown to mid-exponential phase and resuspended in 10 mM sodium phosphate buffer at pH 7.4. These bacteria were then incubated at 37°C for three hours in presence of 0.5, 2.0 or 5.0 μM of RNase A, HsR3, MfR3, PpR3, PaR3, AtR2, or HsR2, each performed in triplicate (Lehrer *et al*. 1989; Carreras *et al*. 2003; Torrent *et al*. 2009). RNase A was used as a non-bactericidal enzyme while HsR3 was used as a positive control (Carreras *et al*. 2003; Rosenberg 2008). Bacterial suspensions were then diluted and plated as 4 x 20 μL droplets on LB-agar plates as per the Miles-Misra technique (Miles *et al*. 1938), and allowed to grow overnight at 37°C. Colony-forming units (CFU) were then counted the following day.

5.4.5 Cytotoxicity Assays

Culture of the human HeLa cell line was achieved using DMEM supplemented with 10% Foetal Bovine Serum (FBS) as culture media. HeLa cells were seeded in 96-well plates at a density of 50 000 cells/well and incubated overnight at 37°C. Cells were then serum deprived for 2 h prior

to RNase treatment performed in serum free culture media with a final volume of 100 µL per well. Cells were treated with RNases at concentrations ranging from 0 to 10 µM for 48 hours. Cell viability was then evaluated through the use of the Cell Titer-Blue Assay from Promega (Madison, WI, USA) following the supplier's protocol. Data were recorded on a Tecan Infinite M1000 Pro microplate reader (Männedorf, Switzerland) set at 560 nm and 590 nm for excitation and emission, respectively. Results were expressed as percentage of control (non-treated cells) and represent the mean±SEM of at least 3 independent experiments performed in triplicate.

5.4.6 Solution NMR Experiments

NMR spectra were recorded at 298 K on 100-to-400 µM protein samples containing 15 mM sodium acetate, pH 5.0, and 10% $^{2}H_2O$. All spectra were performed on a Varian INOVA spectrometer (Palo Alto, CA, USA) at a working frequency of 500 or 800 MHz (MfR3 and PpR3) or on a Bruker Avance III spectrometer (Billerica, MA, USA) at a working frequency of 600 MHz (equipped with a 5 mm TCI cryoprobe) or 800 MHz (equipped with a 5 mm TXI probe) (PaR3 and AtR2). 2D {^{1}H,^{15}N}-HSQC, 3D-HNCACB, 3D-CBCA(CO)NH assignment experiments were performed for all nonhuman proteins. Relaxation-compensated ^{15}N-Carr-Purcell-Meiboom-Gill (^{15}N-CPMG) relaxation dispersion experiments (Loria *et al*. 1999b) were recorded in an interleaved fashion with τ_{cp} delays of 0.625, 0.714 (x2), 1.25, 1.667, 2.0, 2.5 (x2), 3.333, 5.0 and 10.0 ms, using a total relaxation period of 40.0 ms, as described earlier (Doucet *et al*. 2009; Gagné *et al*. 2012; Gagné *et al*. 2015b). ^{15}N-Chemical Exchange Saturation Transfer (^{15}N-CEST) experiments (Vallurupalli *et al*. 2012) were performed using 15 and 25 Hz B_1 irradiation fields. Series of 114 to 124 2D datasets were acquired, corresponding to ^{15}N offsets interspaced by 0.25 ppm (~20 Hz) that span the whole resonance array of each analyzed protein and for both B_1 fields. All spectra were processed using NMRPipe (Delaglio *et al*. 1995), and peak picking was performed using CcpNmr Analysis (Vranken *et al*. 2005) and Sparky (Goddard and Keneller 2008). ^{15}N-CPMG-derived $R_{2,eff}$ values were fitted to the full single-quantum CPMG equation (Carver and Richards 1972; Kempf and Loria 2004) using GraphPad Prism. ^{15}N-CEST intensity profiles were analyzed using the Chemex scripts (https://github.com/gbouvignies/chemex; Vallurupalli *et al*. 2012).

5.4.7 Homology modeling

Structures of hominid RNases PpR3, PaR3, MfR3, and AtR2 were determined through homology modeling using MODELLER v9.13 with human RNases 2 (PDB ID 1GQV) and RNase

3 (PDB ID 1QMT) structures as templates. The four hominid RNase sequences share high sequence identities with human RNases 2 and 3 (Table 5.1). Models were generated using the automodel option and structures with the lowest DOPE scores for each enzyme were selected for each of the four enzymes.

Table 5.1 Sequence identity and similarity between enzymes of interest

	BtRA	HsR2	AtR2	HsR3	MfR3	PpR3	PaR3
BtRA		35%/46%	34%/46%	28%/45%	27%/41%	28%/44%	30%/42%
HsR2	35%/46%		69%/76%	65%/75%	64%/73%	70%/78%	76%/82%
AtR2	34%/46%	69%/76%		68%/75%	62%/72%	69%/78%	68%/75%
HsR3	28%/45%	65%/75%	68%/75%		86%/92%	85%/91%	74%/85%
MfR3	27%/41%	64%/73%	62%/72%	86%/92%		81%/90%	68%/81%
PpR3	28%/44%	70%/78%	69%/78%	85%/91%	81%/90%		84%/89%
PaR3	30%/42%	76%/82%	68%/75%	74%/85%	68%/81%	84%/89%	

5.4.8 Computational simulations

Crystal structures of human RNases 2 and 3, PDB IDs 1GQV and 1QMT, respectively, were used as starting structures. For other RNases with no available crystal structures, homology models were used as the initial coordinates (see above). We validated our models by comparing NMR ^{15}N chemical shifts with computationally derived ones using the CS2BACKBONE module of PLUMED 2.4.4 (Tribello *et al*. 2014) as shown in Supplemental Figure 7.14. All simulations were performed using the graphical processing unit (GPU) enabled version of AMBER (Case *et al*. 2005) v14 (pmemd program) and the ff14SB force field. Each enzyme was placed in a rectangular box with SPC/E water and neutralized through the addition of counter ions. The minimum distance between the enzyme and box edge was set to 10 Å. System equilibration, comprising a series of energy minimization and equilibration steps, was performed at 300 K using the protocol described previously (Agarwal 2004; Narayanan *et al*. 2018a). Simulations were performed using timesteps of 2 fs under constant energy conditions (NVE) with periodic boundary conditions. Long-range electrostatic interactions were computed using the particle mesh Ewald (PME) method and a cut-off of 8 Å. All simulations were performed for a total of 1 μs for each enzyme.

5.4.9 Computational analyses

Root mean square fluctuations of the top ten modes ($RMSF_{10}$) was calculated using the trajectory (conformations) obtained for each enzyme using the *cpptraj* program

(https://github.com/Amber-MD/cpptraj). Translational and rotational diffusion of individual structures was removed by superposing all structures within a trajectory to a reference.

Markov state Model (MSM) (Schütte *et al*. 1999; Swope *et al*. 2004; Noé *et al*. 2007; Prinz *et al*. 2011) analysis was performed to estimate the conformational exchange timescales using PyEMMA (Scherer *et al*. 2015). For each residue, we computed backbone and side-chain torsion angles and used them separately as input features for per-residue MSMs. We thereby assume that the degrees of freedom that govern the motions of the single amino acids are defined by their constant local environment such that larger scale protein motions are negligible for the timescales derived here. The continuous angle trajectories per residue were discretized with 150 cluster centers (*k*-means clustering) to obtain time series of discrete states. Model validation was performed by comparing results for different numbers of discrete states (50, 100, 150) and by ensuring implied timescale convergence. Hidden Markov Models (HMMs) that fulfilled these measures were manually selected to estimate the implied timescales presented below. Model estimation was performed at a lag time of 10 ns.

5.4.10 Cosine similarity comparisons

Cosine similarity comparisons were performed using the above sequence alignments for the dimensionality, residue-dependent parameters extracted from NMR experiments or MSM analyses as the vector components, and the following formula:

Équation VIII

$$\cos\theta = \frac{\vec{A}\cdot\vec{B}}{\left\|\vec{A}\right\|\left\|\vec{B}\right\|} = \frac{\sum_{i=0}^{N} A_i B_i}{\sqrt{\sum_{i=0}^{N} A_i^2}\sqrt{\sum_{i=0}^{N} B_i^2}}$$

This technique uses a vector representation of each protein, with a dimensionality equal to the number of positions in the consensus sequence (134 when comparing only eosinophil RNases, 140 when accounting for RNase A). The consensus sequence represents an indexing corresponding to the multiple sequence alignment of RNases that includes gaps representing insertions/deletions in sequences, and facilitates easier comparison of sequences of different lengths. The value of any residue-dependent parameter can be used as the vector components. Calculating the cosine of the angle θ between these vectors ($\cos(\theta)$) through a dot product thus gives an easily obtainable quantitative way to compare proteins, with values close to 1 (i.e. parallel vectors) corresponding to similar proteins. In order to clarify these comparisons, the

effects of location and exchange rate have been bisected. For pairwise comparison of the exchanging residue location, we used a binary encoding (1 or 0) for exchanging and non-exchanging (or alignment gap) residues as they are identified by CPMG and CEST experiments as well as MSM analyses. Residue-specific k_{ex} values obtained either from CPMG or CEST and MSM implied timescales were then compared directly to compare timescale magnitudes.

5.5 Results

5.5.1 Phylogenetic analysis

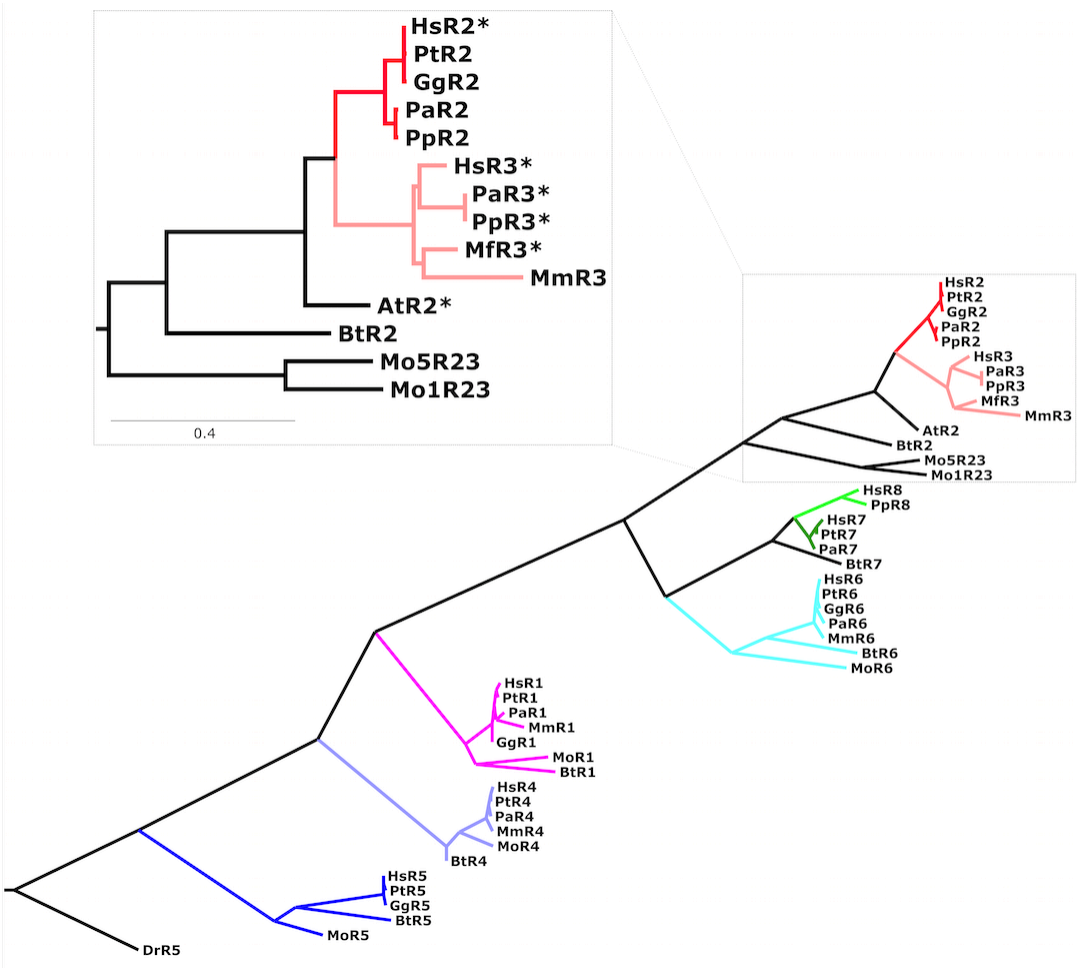


Figure 5.2 Phylogenetic classification of pancreatic-type RNases

Multiple sequence alignment of select members of the RNase superfamily depicted in Supplemental Figure 7.10 was used for phylogenetic clustering. Branching of the eight human RNases is identified using distinct colors. Inset shows the branch corresponding to the Eosinophil RNase subfamily. The six RNases characterized in this work, HsR2, HsR3, AtR2, PpR3, PaR3 and MfR3, are indicated using an asterisk in the inset (*).

To characterize the evolutionary relationships between the homologous sequences within the RNase superfamily for functional and dynamical property comparison, we performed the phylogenetic clustering of a number of RNase sequences from different organisms. Our results show the classification of the canonical RNases into distinct subfamilies. As shown in Figure 5.2, and corroborated by earlier studies (Rosenberg and Domachowske 2001; Pizzo and D'Alessio 2007; Premzl 2013; Goo and Cho 2013; Narayanan *et al*. 2018a), the homologous RNase sequences are clustered by RNase subfamily rather than by species, indicating that the proteins diverged before the species did. RNases 2 and 3 diverged only in Old World monkeys, and thus are relatively new genes (Zhang and Rosenberg 2002). The owl monkey RNase (AtR2) belongs neither to RNases 2 nor to RNases 3, it is rather an undifferentiated eosinophil RNase. RNase 3 sequences from hominids display high sequence similarity, yet the elevated rate at which eosinophil RNases have evolved (Rosenberg *et al*. 1995) has made them dissimilar enough to serve as effective models to probe the effect of dynamical variation on their function.

5.5.2 Biochemical and biological activities of the eosinophil RNases

The premise for the comparison of hominid eosinophil RNases is that their biological activities must be similar, and similar to that of their human counterpart. To test this, we characterized the catalytic and biological (antibacterial and cytotoxic) activities of the selected RNases. We quantified the enzymatic activity of the eosinophil RNases using the smallest dinucleotide substrate UpA for the endonucleolytic cleavage reaction (Table 5.2). The kinetic parameters of other human RNases were quantified for reference. While these enzymes are able to accommodate larger substrates (Gagné and Doucet 2013), which have a better affinity for the enzymes due to the additional interactions formed between the two molecules, the choice of the dinucleotide substrate was to facilitate comparison between the different enzymes and limit the cleavable bonds to a single one per substrate molecule. Our results showed that all enzymes were able to cleave UpA, as expected for the canonical members of this superfamily (Sorrentino 2010). All RNases 3 displayed a k_{cat} of ~1 s^{-1}, and a K_M of ~1-5 mM, resulting in a catalytic efficiency (k_{cat}/K_M) that is three to four orders of magnitude weaker than that of the prototypical RNase A (BtRA). HsR2 showed the highest k_{cat} (~88 s^{-1}) among eosinophil RNases and a catalytic turnover only 20 times lower than BtRA. AtR2 displayed a lower k_{cat}/K_M than all of the other eosinophil RNases, as was previously reported (Zhang and Rosenberg 2002).

Table 5.2 Comparison of the kinetic and biological properties of RNases

Enzyme	pI	k_{cat}	k_{cat} relative to RNase A	K_M	K_M relative to RNase A	k_{cat}/K_M	k_{cat}/K_M relative to RNase A	Antibacterial activity	Cytotoxicity
		s^{-1}		μM		$M^{-1}s^{-1}$			
BtRA	8.64	471±26	1.000	130±35	1.00	$3.62*10^{6}$	1.000	-	-
HsR2	9.20	87.83±3.90	0.186	520±54	4.00	$1.69*10^{5}$	0.047	+/-	+
AtR2	9.20	0.52±0.10	0.0011	4960±1450	38.15	105	$2.90*10^{-5}$	-	+
HsR3	10.47	0.97±0.06	0.0021	1230±220	9.46	789	$2.18*10^{-4}$	++	++
MfR3	10.60	1.000±0.142	0.0021	2650±69	20.38	377	$1.04*10^{-4}$	++	+
PpR3	10.01	1.28±0.12	0.0027	1380±310	10.62	928	$2.56*10^{-4}$	+	+
PaR3	9.53	Not saturated		Not saturated		416	$1.15*10^{-4}$	+	+

While catalytic activity is common to all canonical RNases, RNases 3 are defined by their antibacterial activity, which was shown to be lacking even in RNase 2 (Rosenberg and Dyer 1995). We verified this biological role of eosinophil RNases against both Gram-negative *Escherichia coli* and Gram-positive *Staphylococcus aureus* (see Figure 5.3). Our results showed that all Old World monkey eosinophil RNases 3 (HsR3, MfR3, PpR3, PaR3), which share an elevated pI, caused a dose-dependent reduction in the survival rate of both bacteria, with a stronger effect against *E. coli* than against *S. aureus*. HsR3 and MfR3 displayed the most potent

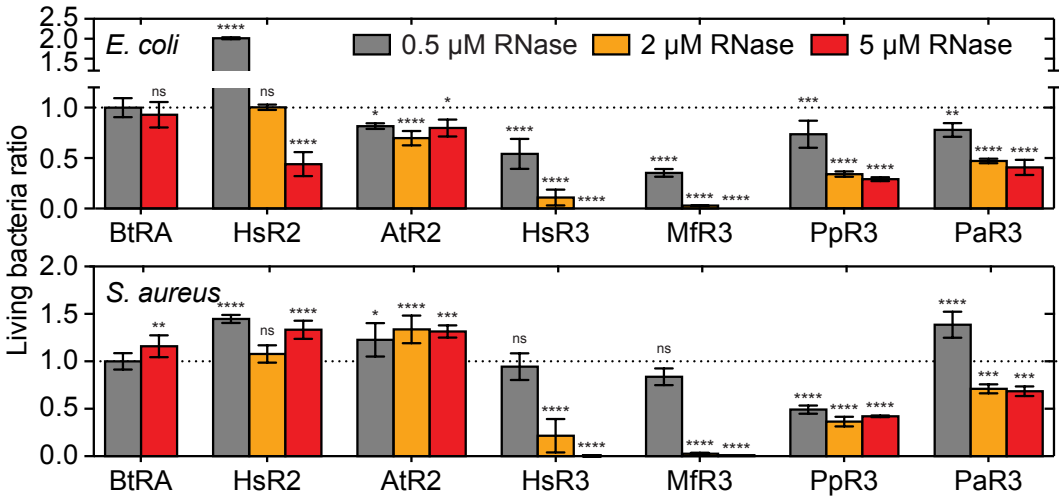


Figure 5.3 Antibacterial activity of hominid eosinophil RNases

Normalized counts of colony-forming units of exponentially growing gram-negative *E. coli* or gram-positive *S. aureus* incubated for three hours in presence of various amounts of eosinophil RNases. Bovine RNase A was used as a negative control, and HsR3 was used as a positive control. Each condition was monitored at least in triplicate. Significance was assessed using OneWay ANOVA with * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$ and **** for $p < 0.0001$.

toxicity, with a survival rate lower than 1.1% of the negative control in presence of 5 µM of RNase. The other RNases 3 were less potent, but still efficient, with 28.9% (42.1%) and 40.8% (68.4%) of *E. coli* (*S. aureus*) surviving a 5 µM concentration of PpR3 and PaR3, respectively. Surprisingly, HsR2 displayed antibacterial activity against *E. coli*, leading to a survival rate of 45.2% at 5 µM concentration. We note that the maximal concentration used was greater than the one reported by Rosenberg (Rosenberg 1995), and no antibacterial effect was observed at 2 µM, consistent with their observations. We did however observe a significant increase in the bacteria survival rate (~200% of the negative control) for *E. coli* incubated with 0.5 µM of HsR2. No antibacterial activity was detected for HsR2 against *S. aureus*. Finally, AtR2 did not significantly affect the survival rate of either bacteria, consistent with previous observations (Zhang and Rosenberg 2002).

To characterize the cytotoxic effects of RNases, we performed cytotoxicity assays against HeLa cells using increasing concentrations of these enzymes (Figure 5.4). To the best of our knowledge, this is the first study reporting the direct cytotoxic effects of these nonhuman RNases against eukaryotic cells. All tested eosinophil RNases decreased HeLa cell survival in a concentration-dependent manner. As expected, RNase A did not alter HeLa cells ability to metabolize the Cell Titer Blue assay reagent while HsR3 significantly reduced cell viability at 50 µM. These results are in accordance with previous observations demonstrating the ability of HsR3 to induce a significant reduction in HeLa cell viability at 20 µM and 40 µM (Navarro *et al*. 2008). Our results showed a significant effect even at 10 µM concentration (Figure 5.4).

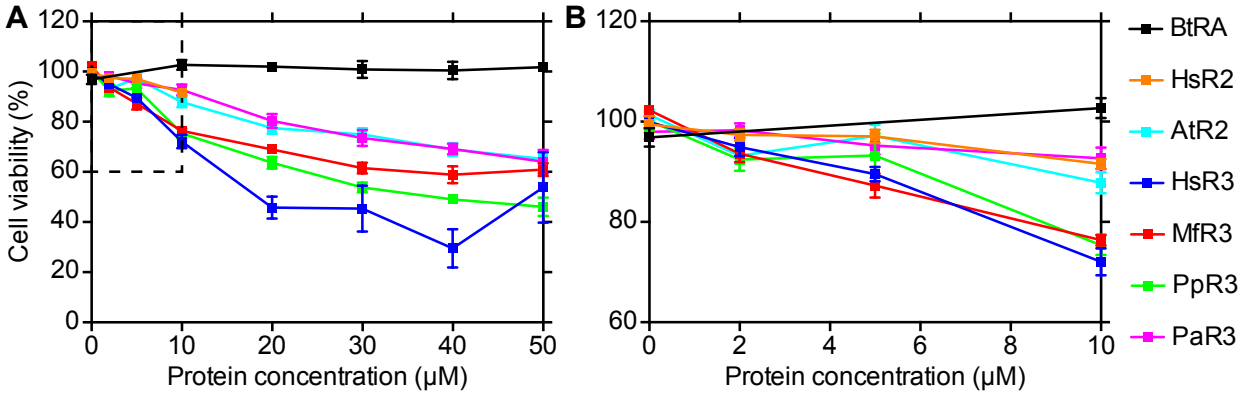


Figure 5.4 Cytotoxic activity of hominid eosinophil RNases

A) Viability of HeLa cells as a function of increasing concentrations of hominid RNases as assessed with the Cell Titer-Blue Cell Viability Assay from Promega. Bovine RNase A was used as a noncytotoxic control. Results are expressed as a percentage of control (non-treated cells) and correspond to mean ± SEM. A close-up can be seen in B).

Interestingly, while AtR2 did not exhibit antibacterial activity, it displayed cytotoxic activity to an extent comparable to that of other eosinophil RNases. The cytotoxic potency observed herein correlates with respective protein pI values, with the highest cytotoxic effects for RNases exhibiting pI values above 10 (HsR3, MfR3 and PpR3), an intermediate response with 9 < pI < 10 (HsR2, PaR3 and AtR2) and no ability to kill HeLa cells, even at 50 µM, for RNase A (pI value below 9). These observations are further supported by recent studies evaluating the cytotoxicity of various mouse eosinophil cationic ribonucleases (mEAR), showing that mEAR2 and mEAR5, which display high pI values, were the most cytotoxic among all tested mEARs (Attery and Batra 2016).

5.5.3 NMR assignments of the four monkey eosinophil RNases

To determine the atomistic behavior of the various eosinophil RNases, we assigned protein backbone atoms using {^{1}H,^{15}N}-HSQC, HNCACB and CBCA(CO)NH NMR spectra. All HSQC spectra (see Supplemental Figures 7.11 and 7.12) showed well-dispersed resonances, indicating well-folded proteins. Assigned non-proline backbone amides for each protein amounted to 120 out of 121 (MfR3), 121 out of 122 (PpR3), 115 out of 121 (PaR3), and 118 out of 121 (AtR2) resonances. All chemical shift assignments, including Cα and Cβ resonances, were deposited in the Biological Magnetic Resonance Bank, under accession numbers 27542, 27544, 27545 and 27546, respectively, for the above proteins. We note that residues from the putative loop 5 (Ser74 to Val78) could not be assigned in PaR3, due to linebroadening of those resonances, thus suggesting an intermediate exchange regime for this loop. This is further supported by the observed weakness in the amplitude of upstream residues His72 and His73 from PaR3, along with that of similarly positioned residues Cys71, His72, His73 and Ser74 of AtR2, or Gly75, Ser76 and Gln77 from HsR2 (Gagné *et al*. 2017; Gagné *et al*. 2018), both of which share a single residue difference with the homologous loop 5 of PaR3.

5.5.4 Comparison of conformational exchange timescales of eosinophil RNases

The effect of conformational exchange of residues in a distal loop on the rate-limiting product release step was demonstrated for bovine RNase A (Cole and Loria 2002; Watt *et al*. 2007). While such a direct link between conformational exchange and catalytic turnover has not been confirmed in other RNases, the presence of concerted motions on functionally relevant timescales was reported for several RNases, including HsR2 and HsR3 (Gagné *et al*. 2012; Gagné and Doucet 2013; Gagné *et al*. 2017; Gagné *et al*. 2018; Narayanan *et al*. 2018a). Here,

we characterized conformational exchange properties of RNases over the ns to s timescales. We performed ^{15}N-CPMG and ^{15}N-CEST experiments to characterize conformational exchange on the timescale of the catalytic turnover (k_{cat}) of the studied eosinophil RNases (see Table 5.2). The CEST experiments, which measure exchange events between ~2.5 ms and ~50 ms (k_{ex} ~20 to ~400 s^{-1}) (Vallurupalli *et al*. 2012; Vallurupalli *et al*. 2017), are used to complement the CPMG relaxation dispersion, which measure exchanges occurring over the ~300 µs to ~10 ms time frame (k_{ex} ~100 to ~3000 s^{-1}) (Kempf and Loria 2004). MD simulations were used to probe the ns-µs timescale motions. We also used Markov State Model (MSM) analysis (Scherer *et al*. 2015) to obtain equilibrium-weighted estimates of the timescale of conformational exchange processes in the simulation ensembles.

Figure 5.5 shows the three-dimensional distribution of conformational motions on the ns-s timescales with residues displaying conformational exchange depicted as spheres. Figure 5.5A illustrates the difference between the two types of analyses performed on the various datasets, i.e. considering the exchanging residue location alone or combining that information with the exchange rate or timescale. A summary of residues displaying conformational exchange is presented in Supplementary Tables 7.5, 7.6 and 7.7. Residues undergoing conformational exchange on the CPMG timescale were located primarily in the V_1 domain in all eosinophil RNases and include residues of loop 4 and β-strands 2, 3, 6 and 7 (Figure 5.5B). Additional exchanging residues were also observed in α-helix 3 of all eosinophil RNases except HsR3, in loop 5 of PpR3, PaR3 and AtR2, and in loop 7 of AtR2. A comparison of the exchange rates (Figure 5.5B), represented using a color gradient, showed similar rates for HsR3, MfR3, PpR3 and AtR2, with a majority of exchanging residues displaying k_{ex} values ranging between 200 and 500 s^{-1} in these proteins. A few isolated residues experienced faster exchange rates (yellow spheres) in MfR3 and PpR3. Exchanging residues in HsR2 displayed faster exchange rates, ranging between 1000 and 1400 s^{-1}, consistent with the higher k_{cat} of this enzyme compared to other eosinophil RNases. PaR3 exchange rates spanned a larger range, between 500 up to 1900 s^{-1}, with no apparent clustering of residues sharing similar k_{ex}. Most of the residues identified as undergoing conformational exchange using CEST (Figure 5.5C) were also observed as undergoing conformational exchange on the slower end of the CPMG timescale (Figure 5.5B), consistent with the overlapping range of timescales in these experiments. This network of interconnected residues are clustered in the V_1 domain of AtR2, HsR3, MfR3, and PpR3. The lack of residues experiencing conformational exchange on the CEST timescale in HsR2 and PaR3 is consistent with the faster k_{ex} values determined from CPMG experiments.

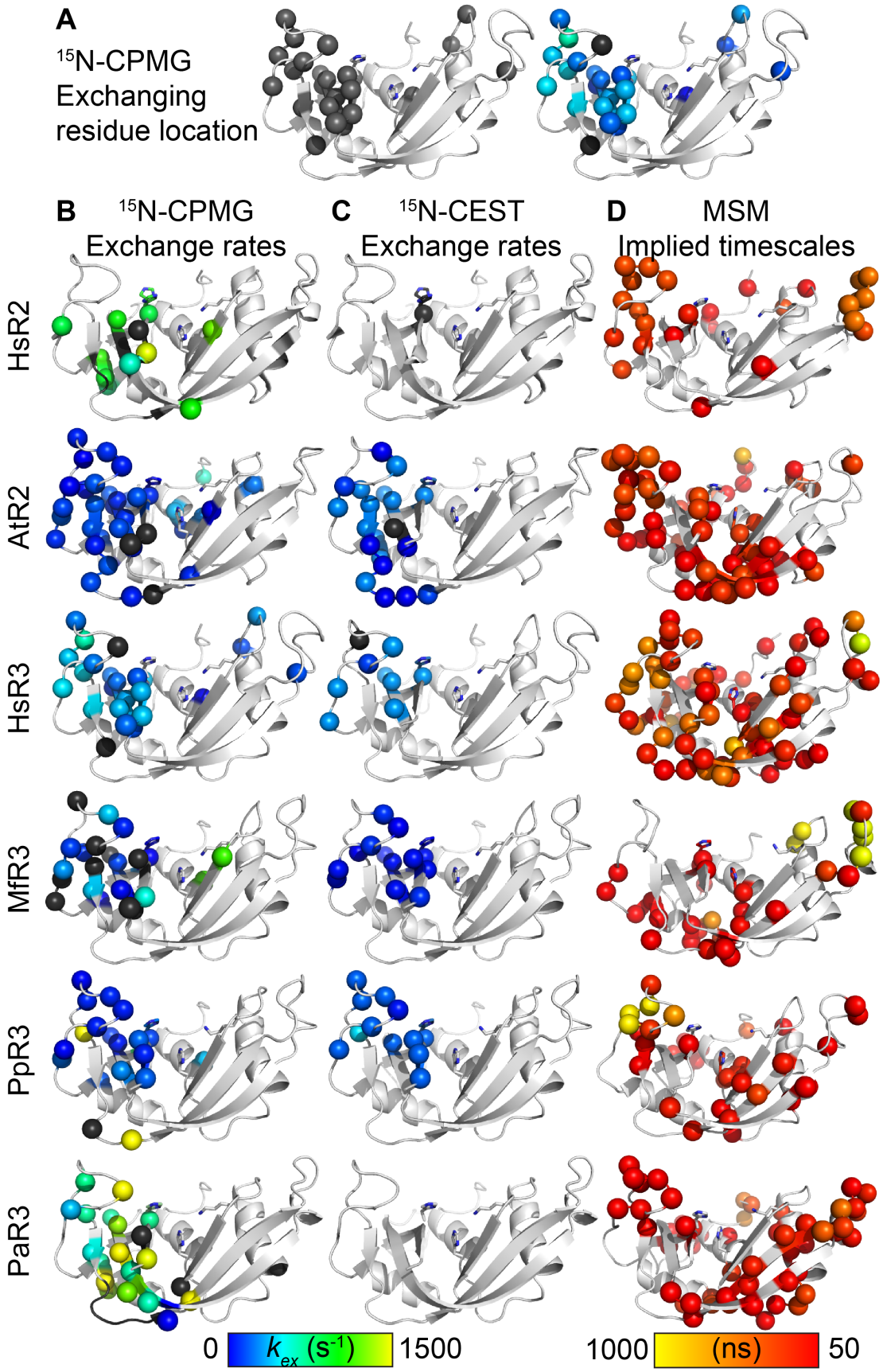

A
15N-CPMG
Exchanging
residue location
B 15N-CPMG
Exchange rates
C 15N-CEST
Exchange rates
D MSM
Implied timescales
HsR2
AtR2
HsR3
MfR3
PpR3
PaR3
0
kex (s-1)
1500
1000
(ns)
50

Figure 5.5 Conformational exchange in hominid eosinophil RNases

Subpanel A depicts the two different analysis methods used to characterize conformational exchange within each RNase, using HsR3 ^{15}N-CPMG data as an example. The left panel depicts exchanging residues as grey spheres to highlight their location, whereas the right panel includes the exchange rate. Residues undergoing conformational exchange were identified using dual fits of ^{15}N-CPMG relaxation dispersion NMR data (B - first column), ^{15}N-CEST experiments (C - second column), and Markov State Model analysis implied timescales (D - third column) and are represented as spheres. Catalytic residues are depicted as sticks. In B and C, exchanging residues are colored according to the exchange rate k_{ex}, whereas residues are colored according to the implied timescale of the exchange process in D. The magnitude of k_{ex} values and implied timescales are represented using the color gradients presented on the bottom. Residues whose fits give a k_{ex} with a standard deviation larger than the actual value are depicted as black spheres. Black loops correspond to unassigned residues due to linebroadening in the NMR spectra. Only residues with an implied timescale slower than 50 ns (see material and methods) are depicted as spheres in the third column. Representative curves for each technique are displayed in Supplementary Figure 7.15 and 7.16.

We characterized the ns-μs timescale motions by determining MSM timescales or decorrelation rates of backbone and side-chain torsion dynamics from the MD ensembles for the different eosinophil RNases, in an effort to tease out comparably slow, metastable dynamics rather than fast, thermal fluctuations. Figure 5.5D shows residues experiencing conformational exchange on the ns-μs timescale as spheres on the three-dimensional structure of the six eosinophil RNases. Exchanging residues vary in number and location across the different RNases. HsR3 and HsR2 displayed the largest (65) and the smallest (26) number of exchanging residues, respectively. Residues of loop 4 experience conformational exchange in all RNases except MfR3, which showed no exchange in this region on the faster timescale. The maximum likelihood estimates of loop 4 motion timescales were observed to span from 60-100 ns for PaR3 (fastest) to 100-750 ns for PpR3 (slowest). Conformational exchange was also observed to varying extents for residues of loop 6 across the different RNases, most noticeably in MfR3 with many residues exchanging on a timescale neighboring ~850 ns.

We further determined the structural changes associated with exchange observed from the MSM analysis. Figure 5.6 shows a noteworthy process for His82, a residue in the V_2 β-sheet known to be essential for substrate binding in the B_1 subsite of HsR2 and HsR3. Our results show that rotamer dynamics is conserved for all eosinophil RNases, as this applies to the structurally equivalent Tyr82 in all other eosinophil RNases (AtR2, PaR3, PpR3 and MfR3). Additionally, we identified an exchange phenomenon on the 0.1-to-1 μs range in a cryptic pocket between α1, α2 and β1 at conserved residue Phe43. However, this event was only observed for AtR2 and HsR3, pointing towards a pattern that was not evolutionary conserved. We note that this analysis is restricted to characterizing single residue dynamics in the ns-μs timescale in only one of potentially many metastable states due to limited sampling. While our

sampling is too sparse to predict slower timescale motions or provide high confidence timescales (Figure 7.13), our analyses indicate convergence of MSM implied timescales for all RNases.

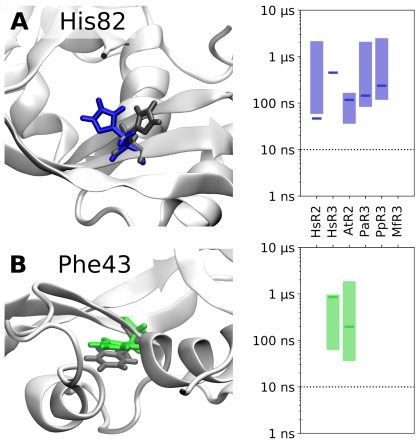


Figure 5.6 Examples of residue dynamics experienced on the 100 ns timescale

These processes are depicted as representative structures of MSM metastable states. A) Two rotamers of His82 (HsR2, HsR3 (shown structure), PpR3, MfR3) or Tyr82 (AtR2, PaR3) (located on β4) are the configurations between which the MSM exchange process occurs. It can be observed in all investigated RNases. Sampling was not sufficient to estimate implied timescales in the case of MfR3. B) Phe43 (on β1, embedded between α1 and α2) is conserved in all RNases but shows conformational exchange on the 100 ns regime only for AtR2 (shown structure) and HsR3.

5.5.5 Quantitative characterization of conformational exchange properties of eosinophil RNases

A comparison of the location of residues undergoing conformational exchange revealed qualitative similarities in the dynamical patterns of eosinophil RNases on the catalytically relevant timescales (Figure 5.5). To quantitatively characterize this similarity in the conformational exchange profiles, we calculated the pairwise cosine similarity between all RNases (Figure 5.7). The conformational exchange profile for bovine RNase A (BtRA), determined previously from CPMG experiments (Gagné *et al*. 2015b), was used for comparison with eosinophil RNases. Characterization of the pairwise cosine similarity between the different RNases showed higher similarity between HsR3, MfR3, PpR3 and AtR2, with the cos(θ) ranging from 0.583 to 0.730 on the CPMG timescale (Figure 5.7A, left panel). PaR3 displayed a cos(θ) of ~0.45 with other eosinophil RNases, even though most of its exchanging residues are located in the V_1 β-sheet like other eosinophil RNases. HsR2 and BtRA displayed the lowest cosine similarity scores with other RNases, highlighting the distinctly different exchange profiles of

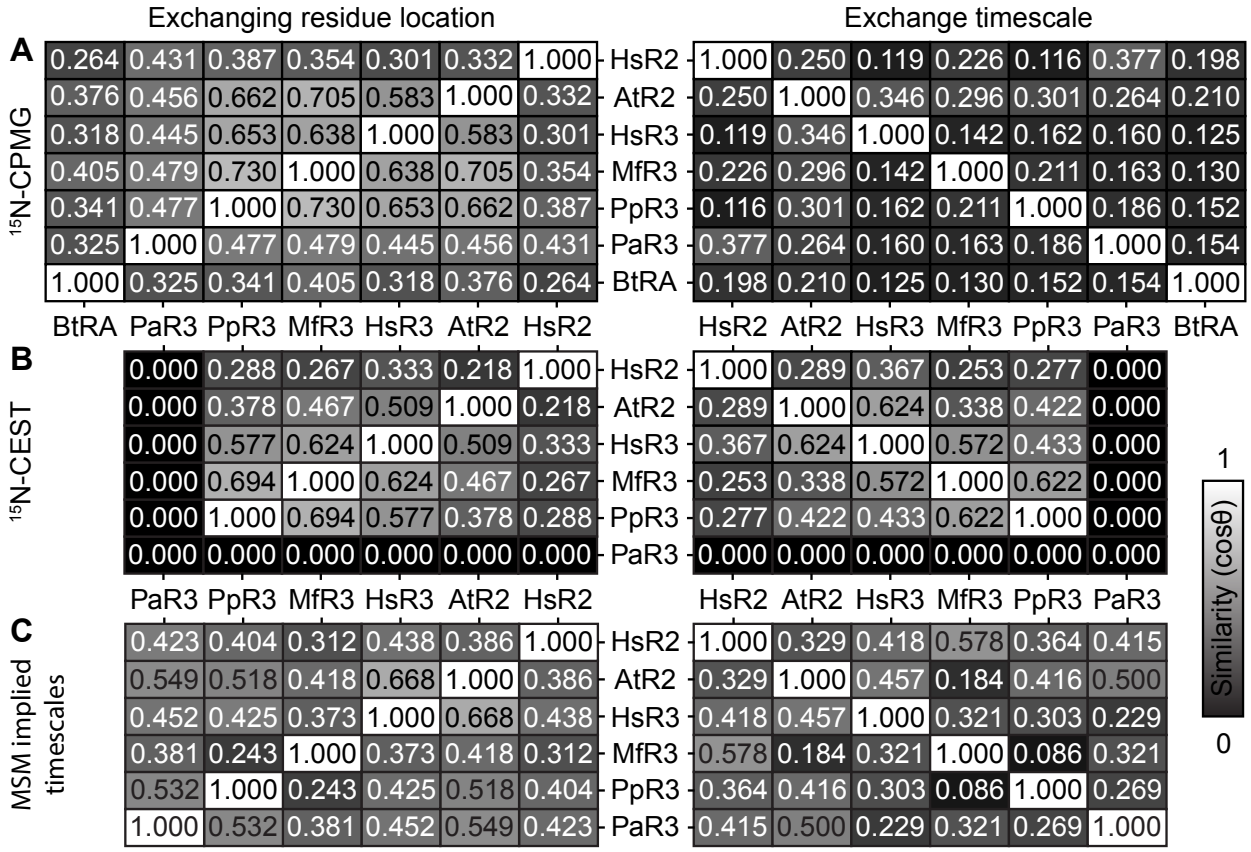


Figure 5.7 Pairwise comparison of eosinophil RNase exchange profiles

Pairwise cosine similarity values for location (left panels) and exchange timescale (right panels) calculated based on comparison of residues along the consensus sequence, color-coded based on strong (white) to weak (dark) similarity.

these enzymes relative to other RNases. A comparison of the cosine similarities based on the magnitude of the exchange rates (k_{ex}) showed smaller values, approaching 0, highlighting a lack of similarity of the exchange rates between RNases (Figure 5.7A, right panel). Among RNases that displayed exchange on the CEST timescale (Figure 5.7B, left panel), highest similarities (between ~0.55 and ~0.7) were observed between HsR3, MfR3 and PpR3. HsR2 showed lowest correlations with other RNases. No correlations were observed involving PaR3, which showed no exchange on the CEST timescale. A comparison of the exchange timescales (Figure 5.7B, right panel) showed similar trends as that observed for residue locations, albeit with small variations in the similarity scores. The cosine similarity values are smaller for most pairwise comparisons, with similarities > 0.5 observed between only three enzyme pairs (AtR2 with HsR3, HsR3 with MfR3, and MfR3 with PpR3), suggesting a diversity in the conformational exchange profiles on this timescale. On the ns-μs timescale probed by MSM analyses,

similarities between the location of the metastable dynamics (Figure 5.7C, left panel) vary between ~0.25 and ~0.67. A comparison of the magnitude of the timescales lowers the similarity values down to at most 0.578, going as low as 0.086 for the MfR3-PpR3 pair (Figure 5.7C, right panel). This corresponds once again to a lack of similarity for the exchange on this timescale. Overall, more enzymes tend towards dissimilarity with their homologs than the reverse, and this holds true at all timescales.

5.5.6 Characterizing the amplitude of atomic fluctuations on the ns-μs timescale

We characterized the amplitude of conformational fluctuations on the faster timescale by computing the root mean square fluctuations for the top ten modes ($RMSF_{10}$) from molecular dynamics simulation trajectories as a function of the consensus sequence. A comparison of the $RMSF_{10}$s of C_α atoms for the six eosinophil RNases (Figure 5.8A) showed qualitative similarity in the dynamical patterns for all members of this subfamily on the faster timescale, consistent with previous observations (Narayanan *et al*. 2018a). Large fluctuations were observed in loops 4 (consensus positions 57-70), 6 (consensus positions 88-96) and 7 (consensus positions 114-124), in addition to the N- and C-terminal regions. Interestingly, AtR2, the undifferentiated member of this subfamily, showed additional fluctuations in the loop 1 region (consensus

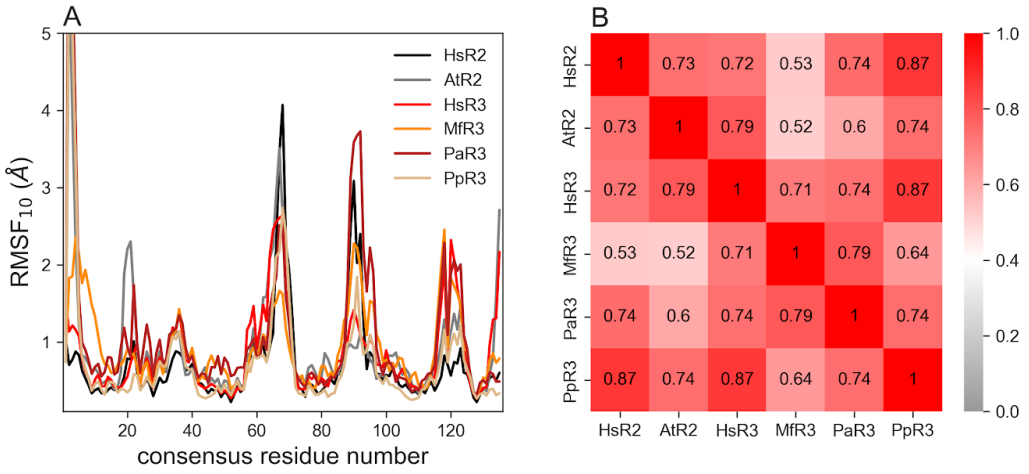


Figure 5.8 Amplitude of atomistic fluctuations on the ns-μs timescale

A) Root mean square fluctuations as a function of the consensus sequence for members of the eosinophil associated RNases. The calculated RMSFs represent the Cα displacements of the top ten quasi-harmonic modes of eosinophil RNases. The consensus residue numbering represents the indexing that includes gaps, corresponding to insertions/deletions, in sequences. B) Pairwise Pearson's correlation coefficients for the eosinophil RNases calculated based on comparison of $RMSF_{10}$ values for each of the positions without an insertion/deletion in any of the six RNases.

positions 17-22). To quantitatively characterize the similarity between the eosinophil RNase sequences, we calculated the pairwise Pearson's correlation coefficients between the different members (Figure 5.8B). Our results showed average correlations between eosinophil RNases of 0.7±0.1, similar to previous observations (Narayanan *et al*. 2018a). Large correlations were observed between most eosinophil RNases, with the exception of MfR3, which showed the smallest correlations with other RNases.

5.6 **Discussion**

The importance of conformational exchange in enzyme function has been illustrated for a variety of model systems (Fraser *et al*. 2009; Bhabha *et al*. 2011; Doucet *et al*. 2011; Gagné *et al*. 2015b; Narayanan *et al*. 2016; Kovermann *et al*. 2017). While these studies demonstrated the role of conformational motions on the function of discrete enzymes, they pertain to proteins that are distant in the sequence space, and thus do not account for the diversity of conformational motions observed within protein families. Recent studies have attempted to address this by probing the conformational properties of homologous sequences within protein families (Wales and Engen 2006; Fang *et al*. 2014; Narayanan *et al*. 2018a). Enzymes clustered into phylogenetic subfamilies that share similar biological functions were shown to display conservation of dynamical properties (Narayanan *et al*. 2018a). However, these observations still do not account for the notable differences in the function observed between members within phylogenetic subfamilies. Identification of factors driving this difference in function will provide insights into the evolutionary trajectory of these enzymes. To characterize the role of conformational exchange on evolutionary divergence, here we determined the catalytic, biological and dynamical properties of six members of the eosinophil RNase subfamily using a combination of biochemical, biophysical and computational approaches.

Eosinophil RNases correspond to one of the subfamilies of pancreatic RNases and include human RNases 2 and 3, which evolved as a result of a gene duplication event after the divergence of Old World monkeys (Rosenberg and Dyer 1995). A comparison of the catalytic activities shows that human RNase 2 (HsR2) displays significantly higher affinity for the UpA substrate and is > 80 times more catalytically active than all other eosinophil RNases (Table 5.2). Previous studies reported the distinctly different antibacterial activities of human and owl monkey eosinophil RNases (Zhang and Rosenberg 2002). Our results further revealed a correlation between antibacterial activities and protein charge, with RNases with pI > 10 displaying the largest antibacterial activities while RNases with pI between 9 and 10 showed

lower activities (Figure 5.3). Further experiments, using mutagenesis, are necessary to confirm this charge dependence of the antibacterial activity. All eosinophil RNases display cytotoxic activities against HeLa cells (Figure 5.4), in line with previous observations for human RNases 2 and 3 (Navarro *et al*. 2008). These results indicate that the cytotoxic activities do not display the charge-dependent gradient in activity observed from antibacterial assays of eosinophil RNases.

We characterized the timescale and amplitude of conformational motions over the ns-s timescales by combining observations from MD simulations, ^{15}N-CPMG and ^{15}N-CEST experiments. Our results show that eosinophil RNases display similarity in the 3D location of residues undergoing conformational exchange on the functionally relevant μs-ms timescale (Figure 5.5A and B), consistent with previous reports (Gagné and Doucet 2013; Narayanan *et al*. 2018a). A comparison of conformational exchange timescales illustrates the distinctly different exchange profiles for HsR2 and PaR3. Residues of AtR2 and all RNases 3, except PaR3, that displayed slower conformational exchange in CPMG measurements were complemented by similar observations from CEST measurements on the ms-s timescale (Figure 5.5C). Quantitative characterization of the similarity in the conformational exchange patterns between the eosinophil RNases showed higher cosine similarity scores for AtR2 and all RNases 3 except PaR3 while HsR2 displayed the largest variations with other RNases (Figure 5.7). Similar trends with diminished similarity scores were observed from CEST experiments, with the exception of PaR3 which showed no exchange on this timescale. Characterization of the faster ns-s timescale motions using MSM analysis showed distinct exchange profiles for the different eosinophil RNases. A comparison of the amplitude of conformational motions showed higher similarities between eosinophil RNases, consistent with previous observations (Narayanan *et al*. 2018a).

Our results show that higher pairwise cosine similarity scores between RNases (Figure 5.7) do not correlate with the proximity in the evolutionary tree of eosinophil RNases (Figure 5.2). This is best exemplified by AtR2, the undifferentiated eosinophil RNase, which shares a higher cosine similarity on all timescales to RNases 3 than PaR3, a post-differentiation enzyme. In addition, cosine similarity scores between HsR2 and other eosinophil RNases were just as low as that between these enzymes and BtRA, which diverged early in the evolutionary tree (Figure 5.2). These results hint at a rapid divergence of conformational exchange patterns, in contrast to a strict conservation, along the evolutionary pathway. Such divergence has been described previously, albeit on much more evolutionarily distant enzymes (Bhabha *et al*. 2013). Our observations may also suggest a potential limitation of the analysis, which compares the

similarity across each consensus residue position, and is not sensitive to any similarities that may arise from the comparison of through-space amino acid networks.

The extent to which eosinophil RNases are subjected to conformational exchange may be linked to the rapid molecular evolution of members within this subfamily (Rosenberg *et al*. 1995). Indeed, proteins undergoing extensive conformational exchange have been shown to also possess high evolvability, defined by Tokuriki and Tawfik as “the ability to rapidly adopt (within a few sequence changes) new functions within existing folds” (Tokuriki and Tawfik 2009). Following this definition, evolvability is hard to evaluate without recourse to extensive mutagenesis, but another feature of highly evolvable enzymes was shown to be the ability to perform more than one reaction with a weak efficiency compared to specialized enzymes of the same fold catalyzing the same reactions (Campbell *et al*. 2016). Numerous examples of such highly flexible, generalist enzymes have been reported, primarily through the characterization of the mechanisms allowing successful protein engineering (Tokuriki *et al*. 2012; Clouthier *et al*. 2012; Dellus-Gur *et al*. 2015; Campbell *et al*. 2016; Risso *et al*. 2017; Campbell *et al*. 2018; Gobeil *et al*. 2019). Our observations for AtR2 shows a larger number of exchanging residues at all timescales than all other studied eosinophil RNases while displaying lower chemical and biological activities. These two characteristics are associated with enzyme evolvability, suggesting that these motions may be necessary to preserve such a property. Subsequent evolution from an undifferentiated protein that might be similar to AtR2 could have led to the “freezing out” of some unproductive motions to give rise to the conserved clusters found in hominid RNases 3. Similarly, hominid RNases 2 may have evolved independently to acquire their antiviral activity, and the larger number of “frozen out” unproductive motions might have led to the increase in activity compared to the other eosinophil RNases. The immune system being in a constant arms race with rapidly evolving pathogens such as viruses, bacteria, and other parasites, it must be able to adapt for a species to be able to carry on over the course of evolution (Pál *et al*. 2006). High evolvability is thus a desirable feature for host defense proteins such as eosinophil RNases.

In conclusion, individual side-chain nanosecond fluctuations as well as clusters of residues exchanging on slower timescales seem to be conserved within eosinophil RNases sharing chemical and biological functions, but the specifics of the large-scale movements on all timescales investigated tend towards rapid divergence rather than conservation. Our results show how enzymes from the eosinophil RNase subfamily retain flexibility throughout evolution, and how these movements change from one member to the next, thus providing a possible

mechanism by which they remain adaptable towards targets which are themselves changing rapidly. We thus demonstrate how rapidly conformational fluctuations can differ between closely related proteins.

5.7 Acknowledgements

The authors thank Dr. Tara Sprules from the Quebec/Eastern Canada High Field NMR Facility (McGill University), Dr. Sameer Al-Abdul-Wahid from the University of Guelph NMR Centre, and Pr. Ewen Lescop from the Institut de Chimie des Substances Naturelles (CNRS, France) for their technical assistance and theoretical NMR discussions. We also thank Marie-Christine Groleau and Pr. Éric Déziel from INRS for their precious help with the antibacterial assays. TH acknowledges financial support from Deutsche Forschungsgemeinschaft (SFB/TRR 186, Project A12) and wishes to thank Simon Olsson (FU Berlin, Germany) for his introduction to NMR.

6 CONCLUSION

6.1 Sommaire et discussion

Le paradigme selon lequel la structure d'une protéine détermine sa fonction a permis d'importants progrès en protéomique au cours des dernières décennies (Berg *et al*. 2002). Ce paradigme est présentement en train de s'adapter à la flexibilité de la structure des protéines, puisque les conformations alternatives permettent à la protéine d'effectuer différentes sous-fonctions. Par conséquent, les transitions entre ces conformations deviennent nécessaires à la compréhension de la fonction des protéines. En effet, les échanges conformationnels sont critiques à la fonction de nombreux systèmes enzymatiques, par exemple la dihydrofolate réductase de *E. coli* (Bhabha *et al*. 2011) la cyclophiline A (Fraser *et al*. 2009), l'adénylate kinase (Kovermann *et al*. 2017), la protéase du VIH (Narayanan *et al*. 2016), l'haloalcane déhalogénase DhlA (Kokkonen *et al*. 2018) et la ribonucléase A bovine (Doucet *et al*. 2011; Gagné *et al*. 2015b). De plus, dans la majorité de ces enzymes, la perturbation des échanges conformationnels a un effet direct sur le renouvellement catalytique. Par contre, la compréhension des échanges conformationnels ainsi que du paysage conformationnel dans son ensemble aurait le potentiel de faciliter grandement la compréhension, la modulation, et surtout l'ingénierie des fonctions protéiques (Serrano-Hervás *et al*. 2018; Maria-Solano *et al*. 2018). Malheureusement, comprendre les échanges conformationnels s'avère complexe, considérant à quel point la séquence, la structure, la fonction et la flexibilité conformationnelle sont intriquées. Même dans les cas où certains de ces facteurs ont pu être étudiés indépendamment (Wales et Engen 2006; Gagné et Doucet 2013; Fang *et al*. 2014; Narayanan *et al*. 2018a), ni la conservation de structure tertiaire ni la conservation de séquence n'était suffisante pour conserver la flexibilité conformationnelle, compliquant davantage l'interprétation de ces comportements protéiques. Donc comment serait-il possible de découvrir les secrets des échanges conformationnels se produisant dans les protéines? Nous avons décidé d'approcher ce problème depuis l'angle de la superfamille des ribonucléases A, un ensemble d'enzymes partageant un repliement commun mais possédant une grande variété de fonctions biologiques, dans une tentative d'étudier indépendamment les variables « structure » et « fonctions. »

Afin d'introduire le sujet du projet de recherche, l'article de revue « Roles of conformational motions in enzyme function : selected methodologies and case studies » (Narayanan, Bernard et Doucet 2016) présentait plusieurs techniques utilisées afin d'analyser les mouvements à

l'échelle moléculaire chez les protéines, à la fois expérimentales et computationnelles, dont certaines ont été utilisées dans les articles subséquents. Deux exemples d'enzymes pour lesquelles de tels changements de conformation sont importants y sont présentés en détails par la suite, faisant le lien entre observations biophysiques et comportements biologiques. Les modèles présentés, soient la protéase du VIH et l'ADN polymérase β, sont deux enzymes pour lesquelles les changements de conformation sont bien documentés et elles démontrent l'importance de les considérer lors de la caractérisation de la fonction biologique d'une protéine. Toutefois, cet article de revue n'a pas présenté l'intégralité des techniques disponibles permettant d'étudier les changements de conformation. Seulement parmi les expériences RMN, on retrouve en plus les expériences de relaxation $R_{1\rho}$, d'échange ZZ, de CEST (utilisé dans l'article de la section 5) et d'échange H-D, qui étudient respectivement les échanges conformationnels aux échelles de temps des µs, des ms, des ms-s et des min-h (Palmer 2014). Plusieurs méthodes n'ayant pas été couvertes par l'article du chapitre 2 ont par contre été décrites dans un article de revue que notre groupe a soumis au Journal of Biological Chemistry, tel que mentionné dans la section 7.5.

Avec l'article « Conservation of dynamics associated with biological function in an enzyme superfamily » (Narayanan, Bernard *et al*. 2018a) commence l'analyse des échanges conformationnels au sein de la superfamille des RNases. Les mouvements intrinsèques de 23 RNases ont été caractérisés par une combinaison de simulations bioinformatiques et d'expériences RMN, et ils ont été comparés à la lumière d'une analyse phylogénétique. Le rassemblement de ces 23 protéines en quatre regroupements phylogénétiques a fait ressortir des ressemblances qualitatives entre les mouvements d'enzymes d'un même regroupement, particulièrement en ce qui a trait à la compartimentalisation relevée précédemment et aux fluctuations des chaînes latérales des acides aminés individuels.

Par la suite, l'article « Ligand-induced variations in structural and dynamical properties within an enzyme superfamily » (Narayanan, Bernard *et al*. 2018b) s'est intéressé aux changements de conformation concomitants à la liaison de ligands, et à comparer la façon dont les RNases humaines réagissaient à ces événements. Pour ce faire, des analyses bioinformatiques ont été effectuées sur des données de titrages RMN de cinq RNases avec deux mononucléotides, en plus d'expériences RMN pour caractériser les changements conformationnels en absence et en présence de ces ligands. Comme la fonction catalytique est strictement conservée chez les RNases actives, il était logique de penser que la liaison des ligands pourrait provoquer des réponses similaires au niveau des changements de conformation associés. Or, l'effet de la

liaison des deux mononucléotides sur le déplacement chimique des différents groupements ^{15}N-^{1}H de même que sur leur flexibilité conformationnelle s'est avéré unique à chaque RNase, y compris pour deux RNases appartenant à un même regroupement identifié dans l'article précédent. Un aspect qui aurait également pu être étudié dans cet article mais auquel nous n'avons pas pensé sur le coup aurait été d'étudier expérimentalement les changements conformationnels impliquant les chaînes latérales catalytiques, tout particulièrement les histidines, en présence et en absence de ligand (Kemme *et al*. 2019). En effet, leur chaîne latérale possède des groupements ^{15}N–^{1}H qui auraient pu être étudiés par CPMG ou par analyse « model-free, » et plusieurs des structures cristallographiques de RNases rapportées dans la PDB font état de deux positions possibles pour la chaîne latérale de l'histidine catalytique de l'extrémité C-terminale (Swaminathan *et al*. 2002; Holloway *et al*. 2005; Kazakou *et al*. 2008).

Les résultats précédents ont montré que la conservation de structure et d'activité catalytique n'était pas suffisante pour expliquer la disparité dans les mouvements que subissent les RNases. Selon l'hypothèse originale stipulant que la conservation de la structure et de la fonction pourrait être corrélée avec la conservation de la flexibilité moléculaire, la grande distance évolutive de même que la différence de fonctions biologiques pourraient expliquer cette disparité. C'est pourquoi nous nous sommes intéressés à un sous-groupe restreint de RNases appartenant à une seule sous-famille dans l'article présentement en préparation « Divergence of conformational motions along the evolutionary pathway in the eosinophil ribonuclease subfamily » (Bernard *et al*. en préparation). Suite à la caractérisation des activités biologiques de divers membres de la sous-famille des RNases d'éosinophiles, qui a confirmé leur conservation des activités catalytique, antibactérienne et cytotoxique, des expériences RMN et des simulations bioinformatiques ont analysé les mouvements intrinsèques des RNases étudiées. Une comparaison quantitative des échanges conformationnels que subissent les RNases d'éosinophiles à l'échelle de la µs-ms a démontré la divergence rapide des profils des mouvements globaux malgré la proximité phylogénétique entre les enzymes. Toutefois, cette divergence ne semble pas être aléatoire, comme le montre la conservation de la compartimentalisation des acides aminés subissant ces échanges de conformation, qui elle est corrélée à la conservation de la fonction biologique.

De manière intéressante, la divergence constatée dans les mouvements globaux ne semble pas être aléatoire. En effet, la compartimentalisation des résidus subissant des échanges conformationnels à l'échelle des µs-ms et des ms-s semble être conservée au sein d'une sous-

famille malgré de nombreuses différences dans la séquence des protéines. En guise de comparaison, la mutation ponctuelle A109G ou alors le remplacement de la boucle 1 par celle de la RNase 3 humaine ont toutes deux complètement modifié le profil d'échange conformationnel et les spécificités de liaison de la RNase A (Doucet *et al*. 2009; Gagné *et al*. 2012; Gagné *et al*. 2015b). Inversement, la dihydrofolate réductase (DHFR) de *E. coli* et celle de l'humain catalysent toutes deux la réduction enzymatique du dihydrofolate en tétrahydrofolate, mais alors que le réarrangement de la boucle de la Met20 à l'échelle de la µs-ms est critique à la catalyse chez l'enzyme de *E. coli* (Boehr *et al*. 2006; Bhabha *et al*. 2011), son homologue humain obtient le même résultat sans procéder à ce mouvement (Davies *et al*. 1990). Les résultats de la présente thèse, particulièrement ceux des deux derniers articles, combinés à ces informations, permettent d'avancer que malgré une tolérance relativement permissive, ce ne sont pas toutes les modifications du profil d'échange conformationnel qui préserveront la fonction protéique.

L'étude des changements conformationnels chez les RNases d'éosinophiles n'a pas permis d'établir la fonction des conformations transitoires auprès de la fonction biologique de ces enzymes. Aucune corrélation n'a pu être établie entre l'échelle de temps de ces mouvements et celle de la catalyse enzymatique (auquel cas PaR3 aurait présenté une différence significative avec les autres RNases 3), ni avec l'une ou l'autre des fonctions biologiques (auquel cas les mouvements de AtR2 n'auraient pas été sur la même échelle de temps que les autres RNases 3, et ceux de PaR3 l'auraient été). Ceci dit, il est intéressant de noter que les mouvements chez les RNases d'éosinophiles se concentrent tous dans le domaine V_1. Il a déjà été établi que l'action antibactérienne des RNases 3 passe par l'ancrage de l'enzyme dans la membrane bactérienne et que cet ancrage s'effectue par des résidus du domaine V_2 (Carreras *et al*. 2003). Il a été proposé que la RNase 3 agisse contre les bactéries via un mécanisme « carpet-like » (Salazar *et al*. 2014), soit le même mécanisme que celui par lequel agissent de nombreux peptides antimicrobiens (Lee *et al*. 2018). Cela signifie donc que le domaine V_1 formerait une protubérance à la surface de la bactérie. Des articles ont montré l'activation des mastocytes par les éosinophiles de même que l'implication de la RNase 3 dans cette activité, bien que le mécanisme exact ne soit pas connu (Lu *et al*. 2018). Il pourrait être possible que les mouvements que subit le domaine V_1 de la RNase 3 permettent la sélection conformationnelle d'une autre molécule (ligand nucléotidique ou partenaire d'une interaction protéine-protéine) et que cette reconnaissance soit l'une des étapes permettant le « cross-talk » entre éosinophiles et mastocytes (Lu *et al*. 2018).

6.2 **Conclusion et perspectives**

L'hypothèse de départ stipulait que les changements conformationnels importants pour la (les) fonction(s) d'enzymes auront été conservés par l'évolution si la structure et la (les) fonction(s) sont conservées. Or, les résultats des trois derniers articles démontrent qu'il en est autrement. La caractéristique la mieux conservée chez les RNases canoniques n'est pas liée à des changements de conformation conservés (Narayanan, Bernard *et al*. 2018b), et la conservation des activités biologiques ne suffit pas non plus à préserver la flexibilité conformationnelle (Bernard *et al*. en préparation). La compartimentalisation des mouvements aux échelles de temps plus lentes (µs-ms et ms-s) semble être conservée (Narayanan, Bernard *et al*. 2018a; Bernard *et al*. en préparation), minimalement auprès de la sous-famille des RNases d'éosinophiles, mais les détails de ces mouvements au niveau des acides aminés individuels varient d'une protéine à l'autre (Bernard *et al*. en préparation). En particulier, l'article présenté à la section 5 a démontré que des enzymes d'une même sous-famille, évolutivement rapprochées, et partageant des fonctions biologiques similaires, subissaient des fluctuations conformationnelles qui diffèrent parfois davantage que ce que la distance évolutive aurait pu laisser présager (Bernard *et al*. en préparation). En prenant en compte que ces RNases 3 conservent structure et fonctions, il est raisonnable d'affirmer que les enzymes étudiées dans l'article appartiennent à un « fitness landscape » commun (Romero et Arnold 2009), et que les variations dans la séquence sont assimilables à des mutations globalement neutres entre les séquences des différentes enzymes étudiées (voir section 1.2.3, et les figures 7.2, 7.8 et 7.10). Dans cette optique, deux explications peuvent être avancées, toutes deux supportées par la compartimentalisation des mouvements aux échelles de la µs-ms et de la ms-s et par la disparité des mouvements précis. Selon la première, la pression évolutive pourrait avoir favorisé la conservation de la structure et des fonctions, mais s'il existe plus d'un mécanisme moléculaire permettant l'accomplissement de ces fonctions (par exemple plus d'une trajectoire possible dans le paysage conformationnel entre deux conformations importantes), peut-être l'évolution a-t-elle favorisé un mécanisme unique à chaque enzyme via la fixation d'acides aminés n'affectant pas directement la fonction ni la structure. Cette variabilité dans ces acides aminés impliqués uniquement dans la trajectoire à travers le payage conformationnel pourrait donc être considérée comme une banque de mutations neutres que la protéine pourrait facilement tolérer. Ceci dit, des effets d'épistasie seraient à considérer car ils pourraient rendre ces mutations individuellement défavorables (Sailer et Harms 2017). L'autre explication possible impliquerait plutôt que la pression évolutive ait favorisé la conservation de la structure et des

fonctions, mais que les secteurs (voir section 1.3.1) non concernés par les éléments conservés par ces pressions évolutives aient pu subir des mutations neutres au fil des générations, permettant à chaque orthologue d'explorer des paysages conformationnels différents – mais qui tous permettent de conserver les fonctions primaires de l'enzyme – et qu'ainsi d'autres fonctions puissent apparaître via le développement de puits additionnels dans le paysage conformationnel. En d'autres termes, la première option implique que l'évolution pourrait avoir favorisé la sélection d'un paysage conformationnel particulier parmi plusieurs possibilités (sélectionner le paysage conformationnel de la figure 6.1A plutôt que celui représenté en B, par exemple), alors que la deuxième implique plutôt des fluctuations autour d'un paysage conformationnel original (des mutations neutres transformeraient le paysage conformationnel de celui représenté à la figure 6.1A en celui représenté en B). Les deux cas impliquent une variabilité du paysage conformationnel qui pourrait être à l'origine du lien entre la flexibilité et l'évolvabilité souligné dans le quatrième article, et qui expliquerait pourquoi l'ingénierie des protéines est plus aisée sur les protéines ayant une plus grande flexibilité conformationnelle (Tokuriki et Tawfik 2009; Dellus-Gur *et al*. 2015; Campbell *et al*. 2016; Risso *et al*. 2017).

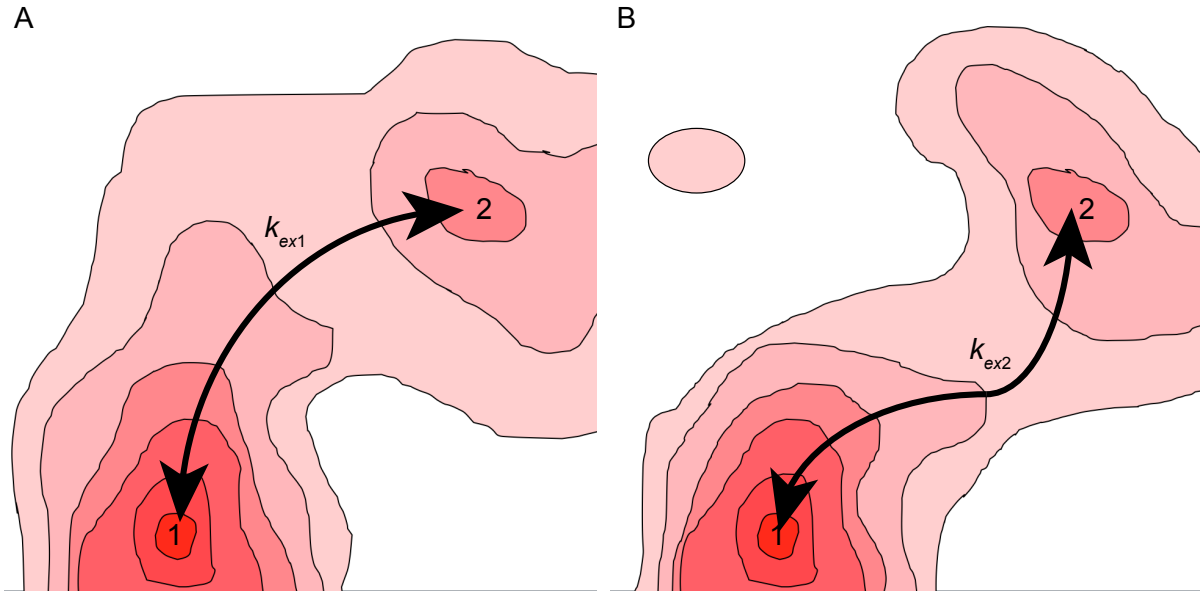


Figure 6.1 Paysages conformationnels différents entre états identiques

Deux exemples de paysages conformationnels à deux dimensions (donc à deux degrés de liberté) entre les conformations 1 et 2, représentées par des puits identiques. Le gradient de couleur correspond à la valeur de ΔG, plus le rouge est foncé, plus le ΔG est négatif. La flèche correspond à la progression du changement de structure que subit la protéine pour passer de l'une à l'autre. Puisque les deux trajectoires sont différentes, les paramètres du phénomène d'échange lui-même, par exemple la constante d'échange k_{ex}, ou les populations relatives des deux états, seront eux aussi différents.

Vérifier ces hypothèses impliquerait de déterminer l'effet sur le paysage conformationnel de chacune des mutations permettant de passer d'une séquence à une autre. Dans un cas idéal, chacune des mutations devrait être neutre au niveau de la stabilité et de la fonction de la protéine. Si la première hypothèse est juste, des mutations neutres devraient conserver les minima correspondant aux conformations principales, seules les barrières les séparant changeraient. Si la deuxième hypothèse est juste, certaines mutations neutres pourraient causer l'apparition ou la disparition de minima.

Dans la poursuite de l'étude de l'évolution des changements conformationnels, il pourrait également être intéressant de reconstruire la séquence d'ancêtres communs (voir section 1.3.2) ayant différentes distances évolutives avec les protéines actuelles. Leur caractérisation permettrait de déterminer quand une fonction particulière serait apparue. Étudier les échanges conformationnels chez ces protéines ancestrales ainsi que chez les intermédiaires entre ces ancêtres et les protéines actuelles permettrait de comprendre comment ceux-ci sont altérés suite à l'apparition d'une fonction.

Pour conclure, les fluctuations d'acides aminés individuels à l'échelle de la nanoseconde ainsi que la compartimentalisation des résidus subissant des échanges à des échelles de temps plus lentes semblent être conservés au sein des regroupements phylogénétiques de RNases. Toutefois, le détail des réarrangements structurels plus globaux tendent vers la divergence rapide plutôt que vers la conservation à toutes les échelles de temps chez les RNases d'éosinophiles. Les résultats présentés dans cette thèse montrent comment les membres de la sous-famille des RNases d'éosinophiles conservent leur flexibilité à travers l'évolution, et comment ces mouvements varient entre les membres de la sous-famille, ce qui fournit un possible mécanisme par lequel ces protéines demeurent adaptables envers des cibles qui elles-mêmes se transforment rapidement.

6.3 **Références**

Abrams C & Bussi G (2014) Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration. Entropy 16, 163–199.

Agarwal PK, Doucet N, Chennubhotla C, Ramanathan A & Narayanan C (2016) Conformational substates and populations in enzyme catalysis. Meth Enzymol 578, 273-297.

Agarwal PK, Schultz C, Kalivretenos A, Ghosh B & Broedel SE Jr (2012) Engineering a hyper-catalytic enzyme by photoactivated conformation modulation. J Phys Chem Lett 3, 1142-1146.

Agarwal PK (2006) Enzymes: An integrated view of structure, dynamics and function. Microb Cell Fact 5, 2.

Agarwal PK (2005) Role of protein dynamics in reaction rate enhancement by enzymes. J Am Chem Soc 127, 15248–15256.

Agarwal PK, Geist A & Gorin A (2004) Protein dynamics and enzymatic catalysis: Investigating the peptidyl-prolyl cis-trans isomerization activity of cyclophilin A. Biochemistry 43, 10605–10618.

Agarwal PK (2004) Cis/trans isomerization in HIV-1 capsid protein catalyzed by cyclophilin A: insights from computational and theoretical studies. Proteins 56, 449-463.

Akiva E, Copp JN, Tokuriki N & Babbitt PC (2017) Evolutionary and molecular foundations of multiple contemporary functions of the nitroreductase superfamily. Proc Natl Acad Sci USA 114, E9549-E9558.

Allemann RK, Presnell SR & Benner SA (1991) A hybrid of bovine pancreatic ribonuclease and human angiogenin: an external loop as a module controlling substrate specificity? Protein Eng 4, 831–835.

Andersen OJ, Grouleff J, Needham P, Walker RC & Jensen F (2015) Toward an enhanced sampling molecular dynamics method for studying ligand-induced conformational changes in proteins. J Phys Chem B 119, 14594–14603.

Anfinsen CB (1959) Part I. Crystalline pancreatic ribonuclease. Introductory remarks. Ann N Y Acad Sci 81, 513-514.

Anisimova M & Gascuel O (2006) Approximate likelihood-ratio test for branches: a fast, accurate and powerful alternative. Syst Biol 55, 539-552.

Arora K, Beard WA, Wilson SH & Schlick T (2005) Mismatch-induced conformational distortions in polymerase β support an induced-fit mechanism for fidelity. Biochemistry 44, 13328–13341.

Attery A & Batra JK (2016) Mouse eosinophil associated ribonucleases: Mechanism of cytotoxic, antibacterial and antiparasitic activities. Int J Biol Macromol 94, 445-450.

Axe JM & Boehr DD (2013) Long-range interactions in the alpha subunit of tryptophan synthase help to coordinate ligand binding, catalysis & substrate channeling. J Mol Biol 425, 1527-1545.

Bakhtina M, Lee S, Wang Y, Dunlap C, Lamarche B & Tsai MD (2005) Use of viscogens, dNTPαS, and rhodium(III) as probes in stopped-flow experiments to obtain new evidence for the mechanism of catalysis by DNA polymerase β. Biochemistry 44, 5177–5187.

Balbo PB, Wang EC & Tsai MD (2011) Kinetic mechanism of active site assembly and chemical catalysis of DNA polymerase β. Biochemistry 50, 9865–9875.

Barnes DE & Lindahl T (2004) Repair and genetic consequences of endogenous DNA base damage in mammalian cells. Annu Rev Genet 38, 445–476.

Batra VK, Beard WA, Shock DD, Pedersen LC & Wilson SH (2008) Structures of DNA polymerase β with active-site mismatches suggest a transient abasic site intermediate during misincorporation. Mol Cell 30, 315–324.

Batra VK, Beard WA, Shock DD, Krahn JM, Pedersen LC & Wilson SH (2006) Magnesium-induced assembly of a complete DNA polymerase catalytic complex. Structure 14, 757–766.

Batot G, Michalska K, Ekberg G, Irimpan EM, Joachimiak G, Jedrzejczak R, Babnigg G, Hayes CS, Joachimiak A & Goulding CW (2017) The CDI toxin of *Yersinia kristensenii* is a novel bacterial member of the RNase A superfamily. Nucleic Acids Res 45, 5013–5025.

Beard WA, Shock DD, Batra VK, Prasad R & Wilson SH (2014) Substrate-induced DNA polymerase β activation. J Biol Chem 289, 31411–31422.

Beard WA & Wilson SH (2006) Structure and mechanism of DNA polymerase β. Chem Rev 106, 361–382.

Beard WA, Shock DD, Yang XP, DeLauder SF & Wilson SH (2002) Loss of DNA polymerase β stacking interactions with templating purines, but not pyrimidines, alters catalytic efficiency and fidelity. J Biol Chem 277, 8235–8242.

Becknell B, Eichler TE, Beceiro S, Li B, Easterling RS, Carpenter AR, James CL, McHugh KM, Hains DS *et al*. (2015) Ribonucleases 6 and 7 have antimicrobial function in the human and murine urinary tract. Kidney Int 87, 151-161.

Bedoya VI, Boasso A, Hardy AW, Rybak S, Shearer GM & Rugeles MT (2006) AIDS Res Hum Retroviruses 22, 897-907.

Berg JM, Tymoczko JL & Stryker L (2002) Biochemistry 5th edition, W H Freeman and Company, New York, 1100 p.

Berlow RB, Swain M, Dalal S, Sweasy JB & Loria JP (2012) Substrate-dependent millisecond domain motions in DNA polymerase β. J Mol Biol 419, 171–182.

Bernard DN, Narayanan C, Hempel T, Bafna K, Bhojane PB, Létourneau M, Howell EE, Noé F, Agarwal PK & Doucet N (en préparation) Divergence of conformational motions along the evolutionary pathway in the eosinophil ribonuclease subfamily.

Bhabha G, Biel JT & Fraser JS (2015) Keep on moving: Discovering and perturbing the conformational dynamics of enzymes. Acc Chem Res 48, 423–430.

Bhabha G, Ekiert DC, Jennewein M, Zmasek CM, Tuttle LM, Kroon G, Dyson HJ, Godzik A, Wilson IA & Wright PE (2013) Divergent evolution of protein conformational dynamics in dihydrofolate reductase. Nat Struct Mol Biol 20, 1243-1249.

Bhabha G, Lee J, Ekiert DC, Gam J, Wilson IA, Dyson HJ, Benkovic SJ & Wright PE (2011) A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. Science 332, 234-238.

Boehr DD, Liu X & Yang X (2014) Targeting structural dynamics of the RNA-dependent RNA polymerase for anti-viral strategies. Curr Opin Virol 9, 194–200.

Boehr DD (2012) Promiscuity in protein-RNA interactions: Conformational ensembles facilitate molecular recognition in the spliceosome. Bioessays 34, 174–180.

Boehr DD, McElheny D, Dyson HJ & Wright PE (2010) Millisecond timescale fluctuations in dihydrofolate reductase are exquisitely sensitive to the bound ligands. Proc Natl Acad Sci USA 107, 1373–1378.

Boehr DD, McElheny D, Dyson HJ & Wright PE (2006) The dynamic energy landscape of dihydrofolate reductase catalysis. Science 313, 1638-1642.

Boix E, Blanco JA, Nogués MV & Moussaoui M (2013) Nucleotide binding architecture for secreted cytotoxic endoribonucleases. Biochimie 95, 1087-1097.

Boix E, Salazar VA, Torrent M, Pulido D, Nogués MV & Moussaoui M (2012) Structural determinants of the eosinophil cationic protein antimicrobial activity. Biol Chem 393, 801–815.

Boix E (2001) Eosinophil cationic protein. Method Enzymol 341, 287–305.

Boix E, Leonidas DD, Nikolovski Z, Nogués MV, Cuchillo CM & Acharya KR (1999a) Crystal structure of eosinophil cationic protein at 2.4 A resolution. Biochemistry 38, 16794-16801.

Boix E, Nikolovski Z, Moiseyev GP, Rosenberg HF, Cuchillo CM & Nogués MV (1999b) Kinetic and product distribution analysis of human eosinophil cationic protein indicates a subsite arrangement that favors exonuclease-type activity. J Biol Chem 274, 15605-15614.

Bowman GR, Voelz VA & Pande VS (2011) Atomistic folding simulations of the five-helix bundle protein λ6–85. J Am Chem Soc 133, 664–667.

Bowman GR, Beauchamp KA, Boxer G & Pande VS (2009) Progress and challenges in the automated construction of Markov state models for full protein systems. J Chem Phys 131, 124101.

Bucher D, Grant BJ, Markwick PR & McCammon JA (2011) Accessing a hidden conformation of the maltose binding protein using accelerated molecular dynamics. PLoS Comput Biol 7, e1002034.

Cabana J, Holleran B, Leduc R, Escher E, Guillemette G & Lavigne P (2015) Identification of distinct conformationsof the angiotensin-II type 1 receptor associated with the $G_{q/11}$ protein pathway and the β-arrestin pathway using molecular dynamics simulations. J Biol Chem 290, 15835-15854.

Camilloni C, Cavalli A & Vendruscolo M (2013) Assessment of the use of NMR chemical shifts as replica-averaged structural restraints in molecular dynamics simulations to characterize the dynamics of proteins. J Phys Chem B 117, 1838–1843.

Camilloni C, Robustelli P, De Simone A, Cavalli A & Vendruscolo M (2012a) Characterization of the conformational equilibrium between the two major substates of RNase A using NMR chemical shifts. J Am Chem Soc 134, 3968–3971.

Camilloni C, Schaal D, Schweimer K, Schwarzinger S & De Simone A (2012b) Energy landscape of the prion protein helix 1 probed by metadynamics and NMR. Biophys J 102, 158–167.

Campbell EC, Correy GJ, Mabbitt PD, Buckle AM, Tokuriki N & Jackson CJ (2018) Laboratory evolution of protein conformational dynamics. Curr Opin Struct Biol 50, 49-57.

Campbell E, Kaltenbach M, Correy GJ, Carr PD, Porebski B, Livingstone EK, Afriat-Jurnou L, Buckle AM, Weik M, Hollfelder F, Tokuriki N & Jackson CJ (2016) The role of protein dynamics in the evolution of new enzyme function. Nat Chem Biol 12, 944-950.

Carreras E, Boix E, Rosenberg HF, Cuchillo CM & Nogués MV (2003) Both aromatic and cationic residues contribute to the membrane-lytic and bactericidal activity of eosinophil cationic protein. Biochemistry 42, 6636–6644.

Carter JD, Gonzales EG, Huang X, Smith AN, de Vera IM, D'Amore PW, Rocca JR, Goodenow MM, Dunn BM & Fanucci GE (2014) Effects of pre and post therapy drug-pressure selected mutations on HIV-1 protease conformational sampling. FEBS Lett 588, 3123–3128.

Carver JP & Richards RE (1972) A general two-site solution for the chemical exchange produced dependence of T_2 upon the Carr-Purcell pulse separation. J Magn Res 6, 89-105.

Case DA, Babin V, Berryman J, Betz RM, Cai Q, Cerutti DS, Cheatham TE, Darden A, Duke RE, Gohlke H *et al*. (2014) AMBER 14 (University of California).

Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B & Woods RJ (2005) The Amber biomolecular simulation programs. J Comput Chem 26, 1668–1688.

Case DA (1998) The use of chemical shifts and their anisotropies in biomolecular structure determination. Curr Opin Struct Biol 8, 624-630.

Chan CC, Moser JM, Dyer KD, Percopo CM & Rosenberg HF (2012) Genetic diversity of human RNase 8. BMC Genomics 13, 40.

Chen S, Ho M-H, Bullock RM, DuBois DL, Dupuis M, Rousseau R & Raugei S (2014) Computing free energy landscapes: application to Ni-based electrocatalysts with pendant amines for H_2 production and oxidation. ACS Catal 4, 229-242.

Cho S, Beintema JJ & Zhang J (2005) The ribonuclease A superfamily of mammals and birds: identifying new members and tracing evolutionary histories. Genomics 85, 208-220.

Chodera JD & Noé F (2014) Markov state models of biomolecular conformational dynamics. Curr Opin Struct Biol 25, 135–144.

Clouthier CM, Morin S, Gobeil SMC, Doucet N, Blanchet J, Nguyen E, Gagné SM and Pelletier JN (2012) Chimeric β-lactamases: global conservation of parental function and fast time-scale dynamics with increased slow motions. PLoS ONE 7, e52283.

Codoñer FM, Alfonso-Loeches S & Fares MA (2010) Mutational dynamics of murine angiogenin duplicates. BMC Evol Biol 10, 310.

Cole R & Loria JP (2002) Evidence for flexibility in the function of ribonuclease A. Biochemistry 41, 6072–6081.

Crabtree B, Holloway DE, Baker MD, Acharya KR & Subramanian V (2007) Biological and structural features of murine angiogenin-4, an angiogenic protein. Biochemistry 46, 2431-2443.

Crestfield AM, Stein WH & Moore S (1963) Alkylation and identification of the histidine residues t the active site of ribonuclease. J Biol Chem 238, 2413-2420.

Cuchillo CM, Nogués MV & Raines RT (2011) Bovine pancreatic ribonuclease: fifty years of the first enzymatic reaction mechanism. Biochemistry 50, 7835-7841.

Cuthbert BJ, Burley KH & Goulding CW (2018) Introducing the new bacterial branch of the RNase A superfamily. RNA Biol 15, 9-12.

Da LT, Wang D & Huang X (2012) Dynamics of pyrophosphate ion release and its coupled trigger loop motion from closed to open state in RNA polymerase II. J Am Chem Soc 134, 2399–2406.

Davies JF, Delcamp TJ, Prendergast NJ, Ashford VA, Freisheim JH & Kraut J (1990) Crystal structures of recombinant human dihydrofolate reductase complexed with folate and 5-deazafolate. Biochemistry 29, 9467-9479.

De Simone A, Montalvao RW, Dobson CM & Vendruscolo M (2013) Characterization of the interdomain motions in hen lysozyme using residual dipolar couplings as replica-averaged structural restraints in molecular dynamics simulations. Biochemistry 52, 6480–6486.

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J & Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6, 277–293.

Dellus-Gur E, Elias M, Caselli E, Prati F, Salverda MLM, de Visser JAGM, Fraser JS & Tawfik DS (2015) Negative epistasis and evolvability in TEM-1 β-lactamase – The thin line between an enzyme's conformational freedom and disorder. J Mol Biol 427, 2396-2409.

Deng NJ, Zheng W, Gallicchio E & Levy RM (2011) Insights into the dynamics of HIV-1 protease: A kinetic network model constructed from atomistic simulations. J Am Chem Soc 133, 9387–9394.

Dhulesia A, Gsponer J & Vendruscolo M (2008) Mapping of two networks of residues that exhibit structural and dynamical changes upon binding in a PDZ domain protein. J Am Chem Soc 130, 8931-8939.

Dill KA & Bromberg S (2011) Molecular driving forces. Statistical thermodynamics in biology, chemistry, physics and nanoscience, second edition. Garland Science, Taylor and Francis group, Abingdon, 784 p.

Diss G, Gagnon-Arsenault I, Dion-Côté A-M, Vignaud H, Ascencio DI, Berger CM & Landry CR (2017) Gene duplication can impart fragility, not robustness, in the yeast protein interaction network. Science 355, 630-634.

Domachowske JB, Dyer KD, Adams AG, Leto TL & Rosenberg HF (1998a) Eosinophil cationic protein/RNase 3 is another RNase A-family ribonuclease with direct antiviral activity. Nucleic Acids Res 26, 3358–3363.

Domachowske JB, Dyer KD, Bonville CA & Rosenberg HF (1998b) Recombinant human eosinophil-derived neurotoxin/RNase 2 functions as an effective antiviral agent against respiratory syncytial virus. J Infect Dis 177, 1458–1464.

Domachowske JB, Bonville CA, Dyer KD & Rosenberg HF (1998c) Evolution of antiviral activity in the ribonuclease A gene superfamily: evidence for a specific interaction between eosinophil-derived neurotoxin (EDN/RNase 2) and respiratory syncytial virus. Nucleic Acids Res 26, 5327-5332.

Donigan KA, Sun KW, Nemec AA, Murphy DL, Cong X, Northrup V, Zelterman D & Sweasy JB (2012) Human POLB gene is mutated in high percentage of colorectal tumors. J Biol Chem 287, 23830–23839.

Doshi U, Holliday MJ, Eisenmesser EZ & Hamelberg D (2016) Dynamical network of residue-residue contacts reveals coupled allosteric effects in recognition, catalysis, and mutation. Proc Natl Acad Sci USA 113, 4735–4740.

Doshi U & Hamelberg D (2015) Towards fast, rigorous and efficient conformational sampling of biomolecules: Advances in accelerated molecular dynamics. Biochim Biophys Acta 1850, 878–888.

Doshi U, McGowan LC, Ladani ST & Hamelberg D (2012) Resolving the complex role of enzyme conformational dynamics in catalytic function. Proc Natl Acad Sci USA 109, 5699–5704.

Doucet N (2011) Can enzyme engineering benefit from the modulation of protein motions? Lessons learned from NMR relaxation dispersion experiments. Protein Pept Lett 18, 336–343.

Doucet N, Khirich G, Kovrigin EL & Loria JP (2011) Alteration of hydrogen bonding in the vicinity of histidine 48 disrupts millisecond motions in RNase A. Biochemistry 50, 1723-1730.

Doucet N & Pelletier JN (2011) Gaining insights into enzyme function through correlation with protein motions. Protein Engineering Handbook, 1 & Volume 2, 187-204.

Doucet N, Watt ED & Loria JP (2009) The flexibility of a distant loop modulates active site motion and product release in ribonuclease A. Biochemistry 48, 7160-7168.

Eisenmesser EZ, Millet O, Labeikovsky W, Korzhnev DM, Wolf-Watz M, Bosco DA, Skalicky JJ, Kay LE & Kern D (2005) Intrinsic dynamics of an enzyme underlies catalysis. Nature 438, 117–121.

Fang J, Nevin P, Kairys V, Venclovas C, Engen JR & Beuning PJ (2014) Conformational analysis of processivity clamps in solution demonstrates that tertiary structure does not correlate with protein dynamics. Structure 22, 572-581.

Feyertag F & Alvarez-Ponce D (2017) Disulfide bonds enable accelerated protein evolution. Mol Biol Evol 34, 1833-1837.

Findlay D, Herries DG, Mathias AP, Rabin BR & Ross CA (1961) The active site and mechanism of bovine pancreatic ribonuclease. Nature 190, 781-784.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A *et al*. (2016) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44, D279–D285.

Fisette O, Lagüe P, Gagné S & Morin S (2012) Synergistic applications of MD and NMR for the study of biological systems. J Biomed Biotechnol 2012, 254208.

Fraser JS, Clarkson MW, Degnan SC, Erion R, Kern D & Alber T (2009) Hidden alternative structures of proline isomerase essential for catalysis. Nature 462, 669-673.

Frauenfelder H, Fenimore PW, Chen G & McMahon BH (2006) Protein folding is slaved to solvent motions. Proc Natl Acad Sci USA 103, 15469–15472.

Frauenfelder H, Parak F & Young RD (1988) Conformational substates in proteins. Annu Rev Biophys Biophys Chem 17, 451-479.

Freudenthal BD, Beard WA, Shock DD & Wilson SH (2013) Observing a DNA polymerase choose right from wrong. Cell 154, 157–168.

Freudenthal BD, Beard WA & Wilson SH (2012) Structures of dNTP intermediate states during DNA polymerase active site assembly. Structure 20, 1829–1837.

Fuglebakk E, Tiwari SP & Reuter N (2015) Comparing the intrinsic dynamics of multiple protein structures using elastic network models. Biochim Biophys Acta 1850, 911–922.

Gabaldón T (2007) Evolution of proteins and proteomes: a phylogenetics approach. Evol Bioinform Online 1, 51-61.

Gagné D, Narayanan C, Bafna K, Charest LA, Agarwal PK & Doucet N (2018) Correction to: Sequence-specific backbone resonance assignments and microsecond timescale molecular dynamics simulation of human eosinophil-derived neurotoxin. Biomol NMR Assign 12, 365-367.

Gagné D, Narayanan C, Bafna K, Charest LA, Agarwal PK & Doucet N (2017) Sequence-specific backbone resonance assignments and microsecond timescale molecular dynamics simulation of human eosinophil-derived neurotoxin. Biomol NMR Assign 11, 143–149.

Gagné D (2015) Conservation et modulation de secteurs dynamiques fonctionnels entre ribonucléases homologues. INRS – Institut Armand-Frappier, Laval. 164p.

Gagné D, Narayanan C & Doucet N (2015a) Network of long-range concerted chemical shift displacements upon ligand binding to human angiogenin. Protein Sci 24, 525–533.

Gagné D, French RL, Narayanan C, Simonović M, Agarwal PK & Doucet N (2015b) Perturbation of the conformational dynamics of an active-site loop alters enzyme activity. Structure 23, 2256-2266.

Gagné D & Doucet N (2014) Sequence-specific backbone (1)H, (13)C, and (15)N resonance assignments of human ribonuclease 4. Biomol NMR Assign 9, 181–185.

Gagné D & Doucet N (2013) Structural and functional importance of local and global conformational fluctuations in the RNase A superfamily. FEBS J 280, 5596-5607.

Gagné D, Charest LA, Morin S, Kovrigin EL & Doucet N (2012) Conservation of flexible residue clusters among structural and functional enzyme homologues. J Biol Chem 287, 44289–44300.

Garcia-Viloca M, Gao J, Karplus M & Truhlar DG (2004) How enzymes work: Analysis by modern rate theory and computer simulations. Science 303, 186–195.

Ghosh AK, Osswald HL & Prato G (2016) Recent progress in the development of HIV-1 protease inhibitors for the treatment of HIV/AIDS. J Med Chem 59, 5172-208.

Gleich GJ & Adolphson CR (1986) The eosinophilic leukocyte: structure and function. Adv Immunol 39, 177-253.

Gobeil SMC, Ebert MCCJC, Park J, Gagné D, Doucet N, Berghuis AM, Pleiss J & Pelletier JN (2019) The structural dynamics of engineered β-lactamases vary broadly on three timescales yet sustain native function. Sci Rep 9, 6656.

Gobeil SMC, Clouthier CM, Park J, Gagné D, Berghuis AM, Doucet N & Pelletier JN (2014) Maintenance of native-like protein dynamics may not be required for engineering functional proteins. Chem Biol 21, 1330–1340.

Goddard T & Keneller DG (2008) SPARKY 3.0. San Francisco, CA: University of California. Available online at: https://www.cgl.ucsf.edu/home/ sparky/

Goldfarb NE, Ohanessian M, Biswas S, McGee TD Jr, Mahon BP, Ostrov DA, Garcia J, Tang Y, McKenna R, Roitberg A *et al*. (2015) Defective hydrophobic sliding mechanism and active site expansion in HIV-1 protease drug resistant variant Gly48Thr/Leu89Met: Mechanisms for the loss of saquinavir binding potency. Biochemistry 54, 422–433.

Goo SM & Cho S (2013) The expansion and functional diversification of the mammalian ribonuclease A superfamily epitomizes the efficiency of multigene families at generating biological novelty. Genome Biol Evol 5, 2124-2140.

Goodey NM & Benkovic SJ (2008) Allosteric regulation and catalysis emerge via a common route. Nat Chem Biol 4, 474–482.

Goricanec D, Stehle R, Egloff P, Grigoriu S, Plückthun A, Wagner G, *et al.* (2016) Conformational dynamics of a G-protein alpha subunit is tightly regulated by nucleotide binding. Proc Natl Acad Sci USA 113, E3629-E3638.

Gotor-Fernández V, Busto E & Gotor V (2006) *Candida antarctica* lipase B: an ideal biocatalyst for the preparation of nitrogenated organic compounds. Adv Synth Catal 348, 797-812.

Graf L, Craik CS, Patthy A, Roczniak S, Fletterick RJ & Rutter WJ (1987) Selective alteration of substrate specificity by replacement of aspartic acid-189 with lysine in the binding pocket of trypsin. Biochemistry 26, 2616-2623.

Grzesiek S, Stahl SJ, Wingfield PT & Bax A (1996) The CD4 determinant for downregulation by HIV-1 Nef directly binds to Nef. Mapping of the Nef binding surface by NMR. Biochemistry 35, 10256–10261.

Grutsch S, Bruschweiler S & Tollinger M (2016) NMR methods to study dynamic allostery. PLoS Comput Biol 12, e1004620.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W & Gascuel O (2010) New algorithms and methods to estimate maximum-likelyhood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59, 307-321.

Gupta A, Jamal S, Goyal S, Jain R, Wahi D & Grover A (2015) Structural studies on molecular mechanisms of nelfinavir resistance caused by non-active site mutation V77I in HIV-1 protease. BMC Bioinform 16, S10.

Gupta SK, Haigh BJ, Griffin FJ & Wheeler TT (2013) The mammalian secreted RNases: mechanisms of action in host defence. Innate Immun 19, 86–97.

Gutte B & Merrifield RB (1971) The synthesis of ribonuclease A. J Biol Chem 246, 1922-1941.

Gutteridge A & Thornton J (2004) Conformational change in substrate binding, catalysis and product release: an open and shut case? FEBS Lett 567, 67-73.

Halabi N, Rivoire O, Leibler S & Ranganathan R (2009) Protein sectors: evolutionary units of three-dimensional structure. Cell 138, 774-786.

Hamelberg D, Mongan J & McCammon JA (2004) Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. J Chem Phys 120, 11919–11929.

Hammes GG, Chang YC & Oas TG (2009) Conformational selection or induced fit: A flux description of reaction mechanism. Proc Natl Acad Sci USA 106:, 13737–13741.

Han B, Liu Y, Ginzinger SW & Wishart DS (2011) SHIFTX2: significantly improved protein chemical shift prediction. J Biomol NMR 50, 43–57.

Hanoian P, Liu CT, Hammes-Schiffer S & Benkovic S (2015) Perspectives on electrostatics and conformational motions in enzyme catalysis. Acc Chem Res 48, 482–489.

Harder J & Schroder J-M (2002) RNase 7, a novel immune defense antimicrobial protein of healthy human skin. J Biol Chem 277 46779-46784.

Harper JW & Vallee BL (1989) A covalent angiogenin/ribonuclease hybrid with a fourth disulfide bond generated by regional mutagenesis. Biochemistry 28, 1875–1884.

Henzler-Wildman K & Kern D (2007) Dynamic personalities of proteins. Nature 450, 964-972.

Henzler-Wildman KA, Thai V, Lei M, Ott M, Wolf-Watz M, Fenn T, Pozharski E, Wilson MA, Petsko GA & Kern D (2007a) Intrinsic motions along an enzymatic reaction trajectory. Nature 450, 838-844.

Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M & Kern D (2007b) A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. Nature 450, 913-916.

Hoang TT & Raines RT (2017) Molecular basis for the autonomous promotion of cell proliferation by angiogenin. Nucleic Acids Res 45, 818-831.

Hofsteenge J, Vicentini A & Zelenko O (1998) Ribonuclease 4, an evolutionarily highly conserved member of the superfamily. Cell Mol Life Sci 54, 804-810.

Hogan SP, Rosenberg HF, Moqbel R, Phipps S, Foster PS, Lacy P, Kay AB & Rothenberg ME (2008) Eosinophils: biological properties and role in health and disease. Clin Exp Allergy 38, 709-750.

Holliday MJ, Camilloni C, Armstrong GS, Vendruscolo M & Eisenmesser EZ (2017) Networks of dynamic allostery regulate enzyme function. Structure 25, 276–286.

Holloway DE, Chavali GB, Hares MC, Subramanian V & Acharya KR (2005) Structure of murine angiogenin: features of the substrate- and cell-binding regions and prospects for inhibitor-binding studies. Acta Crystallogr D Biol Crystallogr 61, 1568-1578.

Hooper LV, Stappenbeck TS, Hong CV & Gordon JI (2003) Angiogenins: a new class of microbicidal proteins involved in innate immunity. Nat Immunol 4, 269-273.

Huang X, Britto MD, Kear-Scott JL, Boone CD, Rocca JR, Simmerling C, McKenna R, Bieri M, Gooley PR, Dunn BM *et al*. (2014) The role of select subtype polymorphisms on HIV-1 protease conformational sampling and dynamics. J Biol Chem 289, 17203–17214.

Huang X, de Vera IM, Veloro AM, Blackburn ME, Kear JL, Carter JD, Rocca JR, Simmerling C, Dunn BM & Fanucci GE (2012) Inhibitor-induced conformational shifts and ligand-exchange dynamics for HIV-1 protease measured by pulsed EPR and NMR spectroscopy. J Phys Chem B 116, 14235–14244.

Huang Y-C, Lin Y-M, Chang T-W, Wu S-J, Lee Y-S, Chang MD-T, Chen C, Wu S-H & Liao Y-D (2007) The flexible and clustered lysine residues of human ribonuclease 7 are critical for membrane permeability and antimicrobial activity. J Biol Chem 282, 4626-4633.

Ishima R & Louis JM (2008) A diverse view of protein dynamics from NMR studies of HIV-1 protease flaps. Proteins 70, 1408–1415.

Ishima R, Torchia DA, Lynch SM, Gronenborn AM & Louis JM (2003) Solution structure of the mature HIV-1 protease monomer: Insight into the tertiary fold and stability of a precursor. J Biol Chem 278, 43311–43319.

Iyer S, Holloway DE & Acharya KR (2013) Crystal structures of murine angiogenin-2 and -3 – probing 'structure – function' relationships amongst angiogenin homologues. FEBS J 280, 302-318.

Jang SH, Song HD, Kang DK, Chang SI, Kim MK, Cho KH, Scheraga HA & Shin HC (2009) Role of the surface loop on the structure and biological activity of angiogenin. BMB Rep 42, 829–833.

Jeschke G & Polyhach Y (2007) Distance measurements on spin-labeled biomacromolecules by pulsed electron paramagnetic resonance. Phys Chem Chem Phys 9, 1895–1910.

Joy JB, Liang RH, McCloskey RM, Nguyen T, Poon AFY (2006) Ancestral reconstruction. PLoS Comput Biol 12, e1004763.

Kamerlin SC & Warshel A (2010) At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? Proteins 78, 1339–1375.

Kar G, Keskin O, Gursoy A & Nussinov R (2010) Allostery and population shift in drug discovery. Curr Opin Pharmacol 10, 715–722.

Kartha G, Bello J & Harker D (1967) Tertiary structure of ribonuclease. Nature 213, 862-865.

Kay LE (2005) NMR studies of protein structure and dynamics. J Magn Reson 173, 193–207

Kazakou K, Holloway DE, Prior SH, Subramanian V & Acharya KR (2008) Ribonuclease A homologues of the zebrafish: polymorphism, crystal structures of two representatives and their evolutionary implications. J Mol Biol 380, 206-222.

Kelemen BR, Klink TA, Behlke MA, Eubanks SR, Leland PA & Raines RT (1999) Hypersensitive substrate for ribonucleases. Nucleic Acids Res 27, 3696–3701.

Kemme CA, Luu RH, Chen C, Pletka CC, Pettitt BM & Iwahara J (2019) Mobility of histidine side chains analyzed with ^{15}N NMR relaxation and cross-correlation data: insights into zinc-finger-DNA interactions. J Phys Chem B 123, 3706-3710.

Kempf JG & Loria JP (2004) Measurement of intermediate exchange phenomena. Methods Mol Biol 278, 185-231.

Kerns SJ, Agafonov RV, Cho YJ, Pontiggia F, Otten R, Pachov DV, Kutter S, Phung LA, Murphy PN, Thai V *et al*. (2015) The energy landscape of adenylate kinase during catalysis. Nat Struct Mol Biol 22, 124–131.

Kirmizialtin S, Nguyen V, Johnson KA & Elber R (2012) How conformational dynamics of DNA polymerase select correct substrates: Experiments and simulations. Structure 20, 618–627.

Kleckner IR & Foster MP (2011) An introduction to NMR-based approaches for measuring protein dynamics. Biochim Biophys Acta 1814, 942–968.

Koczera P, Martin L, Marx G & Schuerholz T (2016) The ribonuclease a superfamily in humans: canonical RNases as the buttress of innate immunity. Int J Mol Sci 17, E1278.

Kohen A (2015) Role of dynamics in enzyme catalysis: Substantial versus semantic controversies. Acc Chem Res 48, 466–473.

Kokkonen P, Bednar D, Dockalova V, Prokop Z & Damborski J (2018) Conformational changes allow processing of bulky substrates by a haloalkane dehalogenase with a small and buried active site. J Biol Chem 293, 11505-11512.

Kövér KE, Bruix M, Santoro J, Batta G, Laurents DV & Rico M (2008) The solution structure and dynamics of human pancreatic ribonuclease determined by NMR spectroscopy provide insight into its remarkable biological activities and inhibition. J Mol Biol 379, 953-965.

Kovermann M, Grundström C, Sauer-Eriksson AE, Sauer UH & Wolf-Watz M (2017) Structural basis for ligand binding to an enzyme

by a conformational selection pathway. Proc Natl Acad Sci USA 114, 6298-6303

Kovrigin EL & Loria JP (2006) Enzyme dynamics along the reaction coordinate: critical role of a conserved residue. Biochemistry 45, 2636-2647.

Krieg J, Hartmann S, Vicentini A, Glasner W, Hess D & Hofsteenge J (1998) Recognition signal for C-mannosylation of Trp-7 in RNase 2 consists of sequence Trp-x-x-Trp. Mol Biol Cell 9, 301–309.

Kunze J, Todoroff N, Schneider P, Rodrigues T, Geppert T, Reisen F, Schreuder H, Saas J, Hessler G, Baringhaus KH *et al*. (2014) Targeting dynamic pockets of HIV-1 protease by structure-based computational screening for allosteric inhibitors. J Chem Inf Model 54, 987–991.

Lai J, Jin J, Kubelka J & Liberles DA (2012) A phylogenetic analysis of normal modes evolution in enzymes and its relationship to enzyme function. J Mol Biol 422, 442–459.

Lee EY, Wong GCL & Ferguson AL (2018) Machine learning-enabled discovery and design of membrane-active peptides. Bioorg Med Chem 26, 2708-2718.

Lee J & Goodey NM (2011) Catalytic contributions from remote regions of enzyme structure. Chem Rev 111, 7595–7624.

Lehrer RI, Szklarek D, Barton A, Ganz T, Hamann KJ & Gleich GJ (1989) Antibacterial properties of eosinophil major basic protein and eosinophil cationic protein. J Immunol 142, 4428-4434.

Leonidas DD, Shapiro R, Subbarao GV, Russo A & Acharya KR (2002) Crystallographic studies on the role of the C-terminal segment of human angiogenin in defining enzymatic potency. Biochemistry 41, 2552–2562.

Leonidas DD, Shapiro R, Allen SC, Subbarao GV, Veluraja K & Acharya KR (1999) Refined crystal structures of native human angiogenin and two active-site variants: implications for the unique functional properties of an enzyme involved in neovascularisation during tumour growth. J Mol Biol 285, 1209-1233.

Leonis G, Steinbrecher T & Papadopoulos MG (2013) A contribution to the drug resistance mechanism of darunavir, amprenavir, indinavir, and saquinavir complexes with HIV-1 protease due to flap mutation I50V: A systematic MM-PBSA and thermodynamic integration study. J Chem Inf Model 53, 2141–2153.

Létourneau D, Bédard M, Cabana J, Lefebvre A, LeHoux J-G & Lavigne P (2016) STARD6 on steroids: solution structure, multiple timescale backbone dynamics and ligand binding mechanism. Sci Rep 6, 28486.

Li S, Sheng J, Hu JK, Yu W, Kishikawa H, Hu MG, Shima K, Wu D, Xu Z, Xin W, Sims KB, Landers JE, Brown RH & Hu G-F (2013) Ribonuclease 4 protects neuron degeneration by promoting angiogenesis, neurogenesis, and neuronal survival under stress. Angiogenesis 16, 387-404.

Linder M, Hermansson A, Liebeschuetz J & Brinck T (2011) Computational design of a lipase for catalysis of a Diels-Alder reaction. J Mol Model 17, 833-849.

Lipari G & Szabo A (1982a) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. J Am Chem Soc 104, 4546–4559.

Lipari G & Szabo A (1982b) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. J Am Chem Soc 104, 4559–4570.

Liu D, Trodler P, Eiben S, Koschorreck K, Müller M, Pleiss J, Maurer SC, Branneby C, Schmid RD & Hauer B (2010) Rational design of *Pseudozyma antarctica* lipase B yielding a general esterification catalyst. Chembiochem 11, 789-795.

Liu J, Wang X-P, Cho S, Lim BK, Irwin DM, Ryder OA, Zhang Y-P & Yu L (2014) Evolutionary and functional novelty of pancreatic ribonuclease: a study of Musteloidea (order Carnivora). Sci Rep 4, 5070.

Liu J, Zhang Y, Lei X & Zhang Z (2008) Natural selection of protein structural and functional properties: a single nucleotide polymorphism perspective. Genome Biol 9, R69.

Loeb LA & Monnat RJ Jr (2008) DNA polymerases and human disease. Nat Rev Genet 9, 594–604.

Loes AN, Bridgham JT & Harms MJ (2018) Coevolution of the toll-like receptor 4 complex with calgranulins and lipopolysaccharide. Front Immunol 9, 304.

Loria JP, Rance M & Palmer AG 3rd (1999a) A TROSY CPMG sequence for characterizing chemical exchange in large proteins. J Biomol NMR 15, 151–155.

Loria JP, Rance M & Palmer AG 3rd (1999b) A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR spectroscopy. J Am Chem Soc 121, 2331–2332.

Lu L, Li J, Moussaoui M & Boix E (2018) Immune modulation by human secreted RNases at the extracellular space. Front Immunol 9, 1012.

Luk LY, Loveridge EJ & Allemann RK (2015) Protein motions and dynamic effects in enzyme catalysis. Phys Chem Chem Phys 17, 30817–30827.

Lynch M & Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290, 1151-1155.

Lyons SM, Fay MM, Akiyama Y, Anderson PJ & Ivanov P (2017) RNA biology of angiogenin: current state and perspectives. RNA

Biol 14, 171-178.

Magadum S, Banerjee U, Murugan P, Gangapur D & Ravikesavan R (2013) Gene duplication as a major force in evolution. J Genet 92, 155-161.

Maguid S, Fernandez-Alberti S & Echave J (2008) Evolutionary conservation of protein vibrational dynamics. Gene 422, 7–13.

Mandel AM, Akke M & Palmer AG 3rd (1995) Backbone dynamics of *Escherichia coli* Ribonuclease HI: Correlations with structure and function in an active enzyme. J Mol Biol. 246, 144–163.

Manley G & Loria JP (2012) NMR insights into protein allostery. Arch Biochem Biophys 519, 223– 231.

Maria-Solano MA, Serrano-Hervás E, Romero-Rivera A, Iglesias-Fernández J & Osuna S (2018) Role of conformational dynamics in the evolution of novel enzyme function. Chem Commun (Camb) 54, 6622-6634.

Marsh JA & Teichmann SA (2014) Parallel dynamics and evolution: protein conformational fluctuations and assembly reflect evolutionary changes in sequence and structure. Bioessays 36, 209–218.

Marsh L (2009) A model for protein sequence evolution based on selective pressure for protein stability: application to hemoglobins. Evol Bioinform Online 5, 107-118.

Miles AA, Misra SS & Irwin JO (1938) The estimation of the bactericidal power of the blood. J Hyg (Lond) 38 732-749.

Mittal S, Cai Y, Nalam MN, Bolon DN & Schiffer CA (2012) Hydrophobic core flexibility modulates enzyme activity in HIV-1 protease. J Am Chem Soc 134, 4163–4168.

Morcos F, Chatterjee S, McClendon CL, Brenner PR, Lopez-Rendon R, Zintsmaster J, Ercsey-Ravasz M, Sweet CR, Jacobson MP, Peng JW *et al*. (2010) Modeling conformational ensembles of slow functional motions in Pin1-WW. PLoS Comput Biol 6, e1001015.

Morin S (2011) A practical guide to protein dynamics from ^{15}N spin relaxation in solution. Prog Nucl Magn Reson Spectrosc 59, 245–262.

Moscato B, Swain M & Loria JP (2016) Induced fit in the selection of correct versus incorrect nucleotides by DNA polymerase β. Biochemistry 55, 382–395.

Nagel ZD & Klinman JP (2009) A 21st century revisionist's view at a turning point in enzymology. Nat Chem Biol 5, 543–550.

Narayanan C, Bernard DN, Bafna K, Gagné D, Chennubhotla CS, Doucet N & Agarwal PK (2018a) Conservation of dynamics associated with biological function in an enzyme superfamily. Structure 26, 1-11.

Narayanan C, Bernard DN, Bafna K, Gagné D, Agarwal PK & Doucet N (2018b) Ligand-induced variations in structural and dynamical properties within an enzyme superfamily. Front Mol Biosci 5, 54.

Narayanan C, Gagné D, Reynolds KA & Doucet N (2017) Conserved amino acid networks modulate discrete functional properties in an enzyme superfamily. Sci Rep 7, 3207.

Narayanan C, Bernard DN & Doucet N (2016) Role of conformational motions in enzyme function: selected methodologies and case studies. Catalysts 6:81.

Nashine VC, Hammes-Schiffer S & Benkovic SJ (2010) Coupled motions in enzyme catalysis. Curr Opin Chem Biol 14, 644–651.

Navarro S, Aleu J, Jiménez M, Boix E, Cuchillo CM & Nogués MV (2008) The cytotoxicity of eosinophil cationic protein/ribonuclease 3 on eukaryotic cell lines takes place through its aggregation on the cell membrane. Cell Mol Life Sci 65, 324-337.

Newton MS, Arcus VL, Gerth ML & Patrick WM (2017) Enzyme evolution: innovation is easy, optimization is complicated. Curr Opin Struct Biol 48, 110-116.

Nicholson LK, Yamazaki T, Torchia DA, Grzesiek S, Bax A, Stahl SJ, Kaufman JD, Wingfield PT, Lam PY, Jadhav PK *et al*. (1995) Flexibility and function in HIV-1 protease. Nat Struct Biol 2, 274–280.

Nikolaev DM, Shtyrov AA, Panov MS, Jamal A, Chakchir OB, Kochemirovski VA, Olivucci M & Ryazantsev MN (2018) A comparative study of modern homology modeling algorithms for rhodopsin structure prediction. ACS Omega 7, 7555-7566.

Nilsen TW & Graveley BR (2010) Expansion of the eukaryotic proteome by alternative splicing. Nature 463, 457-463.

Noé F & Fischer S (2008) Transition networks for modeling the kinetics of conformational change in macromolecules. Curr Opin Struct Biol 18, 154–162.

Noé F, Horenko I, Schütte C & Smith JC (2007) Hierarchical analysis of conformational dynamics in biomolecules: transition networks of metastable states. J Chem Phys 126, 155102.

Nogués MV, Vilanova M & Cuchillo CM (1995) Bovine pancreatic ribonuclease A as a model of an enzyme with multiple substrate binding sites. Biochim Biophys Acta 1253, 16-24.

Ojha S, Meng EC & Babbitt PC (2007) Evolution of function in the "two dinucleotide binding domains" flavoproteins. PLoS Comput Biol 3, e121.

Oyen D, Fenwick RB, Stanfield RL, Dyson HJ & Wright PE (2015) Cofactor-mediated conformational dynamics promote product release from *Escherichia coli* dihydrofolate reductase via an allosteric pathway. J Am Chem Soc 137, 9459–9468.

Pál C, Papp B & Lercher MJ (2006) An integrated view of protein evolution. Nat Rev Genet 7, 337-348.

Palmer AG 3rd (2014) Chemical exchange in biomacromolecules: past, present and future. J Magn Reson 241, 3-17.

Palmer AG 3rd (2004) NMR characterization of the dynamics of biomacromolecules. Chem Rev 104, 3623–3640.

Palmer AG 3rd (2001) NMR probes of molecular dynamics: Overview and comparison with other techniques. Annu Rev Biophys Biomol Struct 30, 129–155.

Pareek V, Samanta M, Joshi NV, Balaram H, Murthy MR & Balaram P (2016) Connecting active site loop conformations and catalysis in triosephosphate isomerase: Insights from a rare variation at residue 96 in the plasmodial enzyme. Chembiochem 17, 620-9

Pastor N & Amero C (2015) Information flow and protein dynamics: The interplay between nuclear magnetic resonance spectroscopy and molecular dynamics simulations. Front Plant Sci 6, 306.

Pelz B, Žoldák G, Zeller F, Zacharias M & Rief M (2016) Subnanometre enzyme mechanics probed by single-molecule force spectroscopy. Nat Commun 7:10848.

Pizzo E & D'Alessio G (2007) The success of the RNase scaffold in the advance of biosciences and in evolution. Gene 406, 8-12.

Prats-Ejarque G, Arranz-Trullén J, Blanco JA, Pulido D, Nogués MV, Moussaoui M & Boix E (2016) The first crystal structure of human RNase 6 reveals a novel substrate-binding and cleavage site arrangement. Biochem J 473, 1523-1536.

Premzl M (2013) Comparative genomic analysis of eutherian ribonuclease A genes. Mol Genet Genomics 10.1007/s00438-013-0801-5.

Prinz J-H, Wu H, Sarich M, Keller B, Senne M, Held M, Chodera JD, Schütte C & Noé F (2011) Markov models of molecular kinetics: generation and validation. J Chem Phys 134, 174105.

Ragland DA, Nalivaika EA, Nalam MN, Prachanronarong KL, Cao H, Bandaranayake RM, Cai Y, Kurt-Yilmaz N & Schiffer CA (2014) Drug resistance conferred by mutations outside the active site through alterations in the dynamic and structural ensemble of HIV-1 protease. J Am Chem Soc 136, 11956–11963.

Raines RT (1998) Ribonuclease A. Chem Rev 98, 1045-1066.

Raines RT, Toscano MP, Nierengarten DM, Ha JH & Auerbach R (1995) Replacing a surface loop endows ribonuclease A with angiogenic activity. J Biol Chem 270, 17180–17184.

Ramanathan A, Savol A, Burger V, Chennubhotla CS & Agarwal PK (2014) Protein conformational populations and functionally relevant substates. Acc Chem Res 47, 149-156.

Ramanathan A, Savol AJ, Langmead CJ, Agarwal PK & Chennubhotla CS (2011) Discovering conformational sub-states relevant to protein function. PLoS ONE 6, e15827.

Ramanathan A & Agarwal PK (2011) Evolutionarily conserved linkage between enzyme fold, flexibility, and catalysis. PLoS Biol 9, e1001193.

Ramanathan A & Agarwal PK (2009) Computational identification of slow conformational fluctuations in proteins. J Phys Chem B 113, 16669–16680.

Reynolds KA, McLaughlin RN & Ranganathan R (2011) Hot spots for allosteric regulation on protein surfaces. Cell 147, 1564-1575.

Richards FM (1972) The 1972 Nobel Prize for chemistry. Science 178, 492-493.

Risso VA, Martinez-Rodriguez S, Candel AM, Krüger DM, Pantoja-Uceda D, Ortega-Muñoz M, Santoyo-Gonzalez F, Gaucher EA, Kamerlin SCL, Bruix M *et al*. (2017) *De novo* active sites for resurrected Precambrian enzymes. Nat Commun 8, 16113.

Robert X & Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res 42, W320–W324.

Roche J, Louis JM & Bax A (2015) Conformation of inhibitor-free HIV-1 protease derived from NMR spectroscopy in a weakly oriented solution. Chembiochem 16, 214–218.

Romero PA & Arnold FH (2009) Exploring protein fitness landscapes by directed evolution. Nat Rev Mol Cell Biol 10, 866-876.

Romero-Rivera A, Garcia-Borràs M & Osuna S (2017) Role of conformational dynamics in the evolution of retro-aldolase activity. ACS Catal 7, 8524-8532.

Rosenberg HF (2015) Eosinophil-derived neurotoxin (EDN/RNase 2) and the mouse eosinophil-associated RNases (mEARs): expanding roles in promoting host defense. Int J Mol Sci 16, 15442-15455.

Rosenberg HF (2008) RNase A ribonucleases and host defense: an evolving story. J Leukoc Biol 83, 1079-1087.

Rosenberg HF & Domachowske JB (2001) Eosinophils, eosinophil ribonucleases, and their role in host defense against respiratory virus pathogens. J Leukoc Biol 70, 691-698.

Rosenberg HF & Dyer KD (1997) Diversity among the primate eosinophil-derived neurotoxin genes: a specific C-terminal sequence is necessary for enhanced ribonuclease activity. Nucleic Acids Res 25, 3532-3536.

Rosenberg HF (1995) Recombinant human eosinophil cationic protein. Ribonuclease activity is not essential for cytotoxicity. J Biol Chem 270, 7876-7881.

Rosenberg HF & Dyer KD (1995) Eosinophil cationic protein and eosinophil-derived neurotoxin. Evolution of novel function in a primate ribonuclease gene family. J Biol Chem 270, 21539-21544.

Rosenberg HF, Dyer KD, Tiffany HL & Gonzalez M (1995) Rapid evolution of a unique family of primate ribonuclease genes. Nat Genet 10, 219-223.

Sailer ZR & Harms MJ (2017) Molecular ensembles make evolution unpredictable. Proc Natl Acad Sci USA 114, 11938-11943.

Salazar VA, Rubin J, Moussaoui M, Pulido D, Nogués MV, Venge P & Boix E (2014) Protein post-translational modification in host defense. The antimicrobial mechanism of action of human eosinophil cationic protein in native forms. FEBS J 281, 5432-5446.

Salomon-Ferrer R, Gotz AW, Poole D, Le Grand S & Walker RC (2013) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. explicit solvent particle mesh Ewald. J Chem Theory Comput 9, 3878–3888.

Sample I (2018) Google's DeepMind predicts 3D shapes of proteins. The Guardian 2 Dec 2018.

Sawaya MR, Prasad R, Wilson SH, Kraut J & Pelletier H (1997) Crystal structures of human DNA polymerase β complexed with gapped and nicked DNA: Evidence for an induced fit mechanism. Biochemistry 36, 11205–11215.

Sawaya MR, Pelletier H, Kumar A, Wilson SH & Kraut J (1994) Crystal structure of rat DNA polymerase β: Evidence for a common polymerase mechanism. Science 264, 1930–1935.

Scherer MK, Trendelkamp-Schroer B, Paul F, Pérez-Hernández G, Hoffmann M, Plattner N, Wehmeyer C, Prinz J-H & Noé F (2015) PyEMMA 2: A software package for estimation, validation, and analysis of Markov models. J Chem Theory Comput 11, 5525-5542.

Schütte C, Fischer A, Huisinga W & Deuflhard P (1999) A direct approach to conformational dynamics based on hybrid Monte Carlo. J Comput Phys 151, 146-168.

Selvaratnam R, VanSchouwen B, Fogolari F, Mazhab-Jafari MT, Das R & Melacini G. (2012) The projection analysis of NMR chemical shifts reveals extended EPAC autoinhibition determinants. Biophys J 102, 630-639.

Serrano-Hervás E, Casadevall G, Garcia-Borràs M, Feixas F & Osuna S (2018) Epoxide hydrolase conformational heterogeneity for the resolution of bulky pharmacologically relevant epoxide substrates. Chemistry 24, 12254-12258.

Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, Bank JA, Jumper JM, Salmon JK, Shan Y *et al*. (2010) Atomic-level characterization of the structural dynamics of proteins. Science 330, 341–346.

Shen Y, Delaglio F, Cornilescu G & Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 44, 213–223.

Sheng J & Xu Z (2016) Three decades of research on angiogenin: a review and perspective. Acta Biochim Biophys Sin (Shanghai) 48, 399–410.

Shukla D, Hernandez CX, Weber JK & Pande VS (2015) Markov state models provide insights into dynamic modulation of protein function. Acc Chem Res 48, 414–422.

Shukla D, Meng Y, Roux B & Pande VS (2014) Activation pathway of SRC kinase reveals intermediate states as targets for drug design. Nat Commun 5, 3397.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J *et al*. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7, 539.

Singh A & Batra JK (2011) Role of unique basic residues in cytotoxic, antibacterial and antiparasitic activities of human eosinophil cationic protein. Biol Chem 392, 337-346.

Singhal N, Snow CD & Pande VS (2004) Using path sampling to build better markovian state models: Predicting the folding rate and mechanism of a tryptophan zipper β hairpin. J Chem Phys 121, 415–425.

Singhania NA, Dyer KD, Zhang J, Deming MS, Bonville CA, Domachowske JB & Rosenberg HF (1999) Rapid evolution of the ribonuclease A superfamily: adaptive expansion of independent gene clusters in rats and mice. J Mol Evol 49, 721–728.

Sorrentino S (2010) The eight "canonical" human ribonucleases: molecular diversity, catalytic properties, and special biological actions of the enzyme proteins. FEBS Lett 584, 2194-2200.

Spinelli S, Liu QZ, Alzari PM, Hirel PH & Poljak RJ (1991) The three-dimensional structure of the aspartyl protease from the HIV-1 isolate BRU. Biochimie 73, 1391–1396.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Starcevic D, Dalal S & Sweasy JB (2004) Is there a link between DNA polymerase β and cancer? Cell Cycle 3, 998–1001.

Stephenson JD, Kenyon JC, Symmons MF & Lever AM (2016) Characterizing 3D RNA structure by single molecule FRET. Methods 103, 57-67

Storz JF (2018) Compensatory mutations and epistasis for protein function. Curr Opin Struct Biol 50, 18-25.

Suel GM, Lockless SW, Wall MA & Ranganathan R (2003) Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nat Struct Biol 10, 59–69.

Sundquist WI & Krausslich HG (2012) HIV-1 assembly, budding, and maturation. Cold Spring Harb Perspect Med 2, a006924.

Sutto L & Gervasio FL (2013) Effects of oncogenic mutations on the conformational free-energy landscape of EGFR kinase. Proc Natl Acad Sci USA 110, 10616–10621.

Swaminathan GJ, Holloway DE, Veluraja K & Acharya KR (2002) Atomic resolution (0.98 A) structure of eosinophil-derived neurotoxin. Biochemistry 41, 3341-3352.

Swope WC, Pitera JW & Suits F (2004) Describing protein folding kinetics by molecular dynamics simulations. 1. Theory. J Phys Chem B 108, 6571-6581.

Tagle AM (1998) Percutaneous encoscopic gastrostomy. Rev Gastroenterol Peru 18, 56–61.

Thornton JW (2004) Resurrecting ancient genes: experimental analysis of extinct molecules. Nat Rev Genet 5, 366-375.

Tokuriki N Jackson CJ, Afriat-Jurnou L, Wyganowski KT, Tang R & Tawfik DS (2012) Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. Nat Commun 3, 1257.

Tokuriki N & Tawfik DS (2009) Protein dynamism and evolvability. Science 324, 203-207.

Torbeev VY, Raghuraman H, Hamelberg D, Tonelli M, Westler WM, Perozo E & Kent SB (2011) Protein conformational dynamics in the mechanism of HIV-1 protease catalysis. Proc Natl Acad Sci USA 108, 20982–20987.

Torrent M, de la Torre BG, Nogués VM, Andreu D & Boix E (2009) Bactericidal and membrane disruption activities of the eosinophil cationic protein are largely retained in an N-terminal fragment. Biochem J 421, 425-434.

Torrie GM & Valleau JP (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. J Comput Phys 23, 187–199.

Towle-Weicksel JB, Dalal S, Sohl CD, Doublie S, Anderson KS & Sweasy JB (2014) Fluorescence resonance energy transfer studies of DNA polymerase β: The critical role of fingers domain movements and a novel non-covalent step during nucleotide selection. J Biol Chem 289, 16541–16550.

Trendelkamp-Schroer B, Wu H, Paul F & Noé F (2015) Estimation and uncertainty of reversible Markov models. J Chem Phys 143, 174101.

Tribello GA, Bonomi M, Branduardi D, Camilloni C & Bussi G (2014) PLUMED 2: New feathers for an old bird. Comput Phys Commun 185, 604.

Tzoupis H, Leonis G, Avramopoulos A, Mavromoustakos T & Papadopoulos MG (2014) Systematic molecular dynamics, MM-PBSA, and *ab initio* approaches to the saquinavir resistance mechanism in HIV-1 PR due to 11 double and multiple mutations. J Phys Chem B 118, 9538–9552.

Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z *et al*. (2008) BioMagResBank. Nucleic Acids Res 36, D402–D408.

Ung PM, Dunbar JB Jr, Gestwicki JE & Carlson HA (2014) An allosteric modulator of HIV-1 protease shows equipotent inhibition of wild-type and drug-resistant proteases. J Med Chem 57, 6468–6478.

Vallurupalli P, Sekhar A, Yuwen T & Kay LE (2017) Probing conformational dynamics in biomolecules via chemical exchange saturation transfer: a primer. J Biomol NMR 67, 243-271.

Vallurupalli P, Bouvignies G & Kay LE (2012) Studying “invisible” excited protein states in slow exchange with a major conformation. J Am Chem Soc 134, 8148-8161.

Vallurupalli P, Hansen DF & Kay LE (2008) Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. Proc Natl Acad Sci USA 105, 11766–11771.

Vanderperre B, Lucier J-F, Bissonnette C, Motard J, Tremblay G, Vanderperre S, Wisztorski M *et al.* (2013) Direct detection of alternative open reading frames translation products in human significantly expands the proteome. PLoS ONE 8, e70698.

Vanwart AT, Eargle J, Luthey-Schulten Z & Amaro RE (2012) Exploring residue component contributions to dynamical network models of allostery. J Chem Theory Comput 8, 2949–2961.

Venkitakrishnan RP, Zaborowski E, McElheny D, Benkovic SJ, Dyson HJ & Wright PE (2004) Conformational changes in the active site loops of dihydrofolate reductase during the catalytic cycle. Biochemistry 43, 16046–16055.

Voelz VA, Jager M, Yao S, Chen Y, Zhu L, Waldauer SA, Bowman GR, Friedrichs M, Bakajin O, Lapidus LJ *et al*. (2012) Slow unfolded-state structuring in Acyl-coA binding protein folding revealed by simulation and experiment. J Am Chem Soc 134, 12565–12577.

Voelz VA, Bowman GR, Beauchamp K & Pande VS (2010) Molecular simulation of *ab initio* protein folding for a millisecond folder NTL9(1–39). J Am Chem Soc 132, 1526–1528.

Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J & Laue ED (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 59, 687–696.

Wales TE & Engen JR (2006) Partial unfolding of diverse SH3 domains on a wide timescale. J Mol Biol 357, 1592-1604.

Warshel A, Sharma PK, Kato M, Xiang Y, Liu H & Olsson MH (2006) Electrostatic basis for enzyme catalysis. Chem Rev 106, 3210–3235.

Watt ED, Rivalta I, Whittier SK, Batista VS & Loria JP (2011) Reengineering rate-limiting, millisecond enzyme motions by introduction of an unnatural amino acid. Biophys J 101, 411-420.

Watt ED, Shimada H, Kovrigin EL & Loria JP (2007) The mechanism of rate-limiting motions in enzyme function. Proc Natl Acad Sci USA 104, 11981-11986.

Webb B & Sali A (2014) Comparative protein structure modeling using MODELLER. Curr Protoc Bioinformatics 47, 5.6.1–5.6.32.

Wei G, Xi W, Nussinov R & Ma B (2016) Protein ensembles: How does nature harness thermodynamic fluctuations for life? The diverse functional roles of conformational ensembles in the cell. Chem Rev 116, 6516-6551.

Williamson MP (2013) Using chemical shift perturbation to characterize ligand binding. Prog Nucl Magn 73, 1–16.

Wishart DS, Sykes BD & Richards FM (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. J Mol Biol 222, 311-333.

Wojtas-Niziurski W, Meng Y, Roux B & Berneche S (2013) Self-learning adaptive umbrella sampling method for the determination of free energy landscapes in multiple dimensions. J Chem Theory Comput 9, 1885–1895.

Wolfenden R (2006) Degrees of difficulty of water-consuming reactions in the absence of enzymes. Chem Rev 106, 3379–3396.

Wu TD, Schiffer CA, Gonzales MJ, Taylor J, Kantor R, Chou S, Israelski D, Zolopa AR, Fessel WJ & Shafer RW (2003) Mutation patterns and structural correlates in human immunodeficiency virus type 1 protease following different protease inhibitor treatments. J Virol 77, 4836–4847.

Xia J, Deng NJ & Levy RM (2013) NMR relaxation in proteins with fast internal motions and slow conformational exchange: Model-free framework and Markov state simulations. J Phys Chem B 117, 6625–6634.

Yamada KJ, Barker T, Dyer KD, Rice TA, Percopo CM, Garcia-Crespo KE, Cho S, Lee JJ, Druey KM & Rosenberg HF (2015) Eosinophil-associated ribonuclease 11 is a macrophage chemoattractant. J Biol Chem 290, 8863–8875.

Yamazaki T, Hinck AP, Wang YX, Nicholson LK, Torchia DA, Wingfield P, Stahl SJ, Kaufman JD, Chang CH, Domaille PJ *et al*. (1996) Three-dimensional solution structure of the HIV-1 protease complexed with DMP323, a novel cyclic urea-type inhibitor, determined by nuclear magnetic resonance spectroscopy. Protein Sci 5, 495–506.

Yang D, Rosenberg HF, Chen Q, Dyer KD, Kurosaka K & Oppenheim JJ (2003) Eosinophil-derived neurotoxin (EDN), an antimicrobial protein with chemotactic activities for dendritic cells. Blood 102, 3396-3403.

Zhang C, Cantara W, Jeon Y, Musier-Forsyth K, Grigorieff N & Lyumkis D (2019) Analysis of discrete local variability and structural covariance in macromolecular assemblies using cryo-EM and focused classification. Ultramicroscopy 203, 170-180

Zhang J, Dyer KD & Rosenberg HF (2003) Human RNase 7: a new cationic ribonuclease of the RNase A superfamily. Nucleic Acids Res 31, 602-607.

Zhang J & Rosenberg HF (2002) Complementary advantageous substitutions in the evolution of an antiviral RNase of higher primates. Proc Natl Acad Sci USA 99, 5486-5491.

Zhang J, Dyer KD & Rosenberg HF (2002) RNase 8, a novel RNase A superfamily ribonuclease expressed uniquely in placenta. Nucleic Acids Res 30, 1169-1175.

Zhang N, Suen W-C, Windsor W, Xiao L, Madison V & Zaks A (2003) Improving tolerance of *Candida antarctica* lipase B towards irreversible thermal inactivation through directed evolution. Protein Eng 16, 599-605.

Zheng W, Brooks BR & Thirumalai D (2007) Allosteric transitions in the chaperonin GroEL are captured by a dominant normal mode that is most robust to sequence variations. Biophys J 93, 2289–2299.

Zhong X, Patel SS, Werneburg BG & Tsai MD (1997) DNA polymerase β: Multiple conformational changes in the mechanism of catalysis. Biochemistry 36, 11891–11900.

7 ANNEXES

7.1 Preuve mathématique de l'équation III

Considérons un échange conformationnel entre un état de base *i* et un état excité *j*. La constante d'échange k_{ex} est égale à la somme de la constante de vitesse de la transition directe et de la transition inverse, s'écrit également $k_{ex} = k_1 + k_{-1}$ (Dill et Bromberg 2011) :

Équation IX $$k_{ex} = k_{ij} + k_{ji}$$

L'équation d'Arrhenius (équation II) permet de réécrire sous la forme suivante :

Équation X $$k_{ex} = A_{ij} e^{-\frac{\Delta\Delta G_{ij}^{bar}}{k_B T}} + A_{ji} e^{-\frac{\Delta\Delta G_{ji}^{bar}}{k_B T}}$$

Comme le montre la figure 1.8, l'énergie de la barrière séparant *i* de *j* ($\Delta\Delta G_{ij}^{bar}$) est égale à la somme de l'énergie de la barrière séparant *j* de *i* ($\Delta\Delta G_{ji}^{bar}$) et de la différence énergétique entre les deux états ($\Delta\Delta G_{ji}$) :

Équation XI $$k_{ex} = A_{ij} e^{-\frac{\left(\Delta\Delta G_{ji}^{bar} + \Delta\Delta G_{ji}\right)}{k_B T}} + A_{ji} e^{-\frac{\Delta\Delta G_{ji}^{bar}}{k_B T}}$$

Ensuite, selon la loi du produit des puissances :

Équation XII $$k_{ex} = A_{ij} e^{-\frac{\Delta\Delta G_{ji}^{bar}}{k_B T}} e^{-\frac{\Delta\Delta G_{ji}}{k_B T}} + A_{ji} e^{-\frac{\Delta\Delta G_{ji}^{bar}}{k_B T}}$$

Ceci permet la factorisation suivante :

Équation XIII $$k_{ex} = A_{ij} e^{-\frac{\Delta\Delta G_{ji}^{bar}}{k_B T}} \left(A_{ji} e^{-\frac{\Delta\Delta G_{ji}}{k_B T}} + \frac{A_{ji}}{A_{ij}} \right)$$

Puisque A_{ij} et A_{ji} sont deux constantes, nous pouvons définir leur rapport comme une constante *C*, ce qui nous ramène à l'équation III.

7.2 Figures supplémentaires de l'article « Conservation of Dynamics Associated with Biological Function in an Enzyme Superfamily »

Tableau 7.1 PDB IDs of RNase homologs used for dynamical characterization on the µs-ms timescale

PDB code	Abbreviation	Species	RNase type
7RSA	BtRA	*Bos taurus*	RNase A
2K11	HsR1	*Homo sapiens*	RNase 1
3TSR	MmR1	*Mus musculus*	RNase 1
1RRA	RnRA	*Rattus norvegicus*	RNase A
1GQV	HsR2	*Homo sapiens*	RNase 2
1QMT	HsR3	*Homo sapiens*	RNase 3
PpECP*	PpR3	*Pongo pygmaeus*	RNase 3
MfECP*	MfR3	*Macaca fascicularis*	RNase 3
2HKY	HsR7	*Homo sapiens*	RNase 7
RNase6[#1]	HsR6	*Homo sapiens*	RNase 6
1ANG	HsR5	*Homo sapiens*	Angiogenin
1AGI	BtR5	*Bos taurus*	Angiogenin
2BWK	MmR5	*Mus musculus*	Angiogenin
4QFI	RnR5	*Rattus norvegicus*	Angiogenin
3ZBV	MmR52	*Mus musculus*	Angiogenin
3ZBW	MmR53	*Mus musculus*	Angiogenin
1RNF	HsR4	*Homo sapiens*	RNase 4
1KM9	RcRx	*Rana catesbeiana*	Bullfrog RNase
3SNF	RpRx	*Rana pipiens*	Onconase
4PER	GgR5	*Gallus gallus*	Angiogenin
2VQ9	DrR1	*Danio rerio*	RNase ZF-3E
3LJE	DrR5	*Danio rerio*	Angiogenin
2ZPO	CmRx	*Chelonia mydas*	Turtle RNase

* Homology model; [#] Structure not deposited; [1] The crystal structure of human RNase 6 was obtained in our research group. While this structure has not yet been deposited in the PDB, it shows a significant structural identity with the structure available in the PDB (4X09), with the root mean square deviation of 0.20 Å.

Tableau 7.2 Amino acid substitutions and their effect on the catalytic and biological activities of select RNase homologs

Protein	Mutation	Localization on structure	Effect on k_{cat}/K_M	Other effects	References
HsR3	R121A	Loop 7 (V_1)	No effect	Reduced bactericidal activity against Gram-positive *S. aureus*	(Carreras *et al*. 2003)
HsR3	W10K	Helix 1	2-fold decrease	Reduced bactericidal activity against Gram-positive *S. aureus* and reduced membrane-permeating capacity	(Carreras *et al*. 2003)
HsR3	R75A/ F76A	Loop 5 (V_1)	2-fold decrease	Reduced bactericidal activity against Gram-negative *E. coli*	(Carreras *et al*. 2003)
HsR3	R121A/ Y122A	Loop 7 (V_1)	1.25-fold increase	Reduced bactericidal activity against Gram-positive *S. aureus*	(Carreras *et al*. 2003)
HsR3	W35A/ R36A	Loop 2 (V_1)	1.6-fold increase	Reduced bactericidal activity against both *E. coli* and *S. aureus*, and reduced membrane-permeating capacity	(Carreras *et al*. 2003)
HsR3	R101A/ R104A	β-strand 4 (V_2)	2-fold increase	Reduced bactericidal activity against Gram-negative *E. coli* and reduced membrane-permeating capacity	(Carreras *et al*. 2003)
HsR3	115-122 EDN	Loop 7 (V_1)	No effect		(Carreras *et al*. 2003)
HsR3	Δ(115-122)	Loop 7 (V_1)	1.5-fold decrease	Reduced bactericidal activity against Gram-positive *S. aureus*	(Carreras *et al*. 2003)
Mouse EAR 11	K35R	Loop 2 (V_2)	Catalytically inactive (catalytic K)	No effect on leukocyte chemotaxis	(Yamada *et al*. 2015)
HsR2	W7A	Helix 1		Abolishment of EDN C-mannosylation (W7 was identified as the C-mannosylation site)	(Krieg *et al*. 1998)
HsR2	W10A	Helix 1		Abolishment of W7 C-mannosylation	(Krieg *et al*. 1998)
HsR2	W10F	Helix 1		Partial rescue of W7 C-mannosylation abolishment	(Krieg *et al*. 1998)
HsR2	R132T	β-strand 6 (V_1)	12-fold decrease		(Zhang and Rosenberg 2002)
HsR2	S64R	Loop 4 (V_1)	74-fold decrease		(Zhang and Rosenberg 2002)
HsR2	S64R/ R132T	V1	12-fold decrease	Reduced antiviral activity against RSV	(Zhang and Rosenberg 2002)
AtR2	T132R	β-strand 6 (V_1)	1.7-fold decrease		(Zhang and Rosenberg 2002)
AtR2	R64S	Loop 4 (V_1)	1.5-fold increase		(Zhang and Rosenberg 2002)
AtR2	R64S/ T132R	V_1	15-fold increase	No increase in antiviral activity against RSV (opposite mutation than in HsR2 S64R/R132T)	(Zhang and Rosenberg 2002)
HsR2	K38R	Loop 2 (V_2)	Catalytically inactive (catalytic K)	No antiviral activity against RSV	(Domachowske *et al*. 1998a)
HsR2	K38R	Loop 2 (V_2)	Catalytically inactive (catalytic K)	No antiviral activity against RSV	(Domachowske *et al*. 1998b)
HsR3	R97T (natural variant)	Loop 6 (V_2)		Reduced cytotoxicity and neurotoxicity	(Boix *et al*. 2012)
HsR3	R22A	Loop 1 (V_2)	1.6-fold decrease	Large reduction in cytotoxicity; small reduction in bactericidal activity against Gram-negative *E. coli*; large reduction in bactericidal activity against Gram-positive *B. subtilis*; large reduction of antiparasitic activity	(Singh and Batra 2011)

HsR3	R34A	Loop 2 (V_2)	1.25-fold decrease	Large reduction in cytotoxicity; small reduction in bactericidal activity against Gram-negative *E. coli*; large reduction in bactericidal activity against Gram-positive *B. subtilis*; large reduction of antiparasitic activity	(Singh and Batra 2011)
HsR3	R61A	Loop 4 (V_1)	1.4-fold decrease	Reduction in cytotoxicity; small reduction in bactericidal activity against Gram-negative *E. coli*; large reduction in bactericidal activity against Gram-positive *B. subtilis*; reduction of antiparasitic activity	(Singh and Batra 2011)
HsR3	H64A	Loop 4 (V_1)	1.2-fold increase	Reduction in cytotoxicity; small reduction in bactericidal activity against Gram-negative *E. coli*; large reduction in bactericidal activity against Gram-positive *B. subtilis*; reduction of antiparasitic activity	(Singh and Batra 2011)
HsR3	R77A	Loop 5 (V_1)	1.9-fold increase	Reduction in cytotoxicity; small reduction in bactericidal activity against Gram-negative *E. coli*; large reduction in bactericidal activity against Gram-positive *B. subtilis*; reduction of antiparasitic activity	(Singh and Batra 2011)
HsR3	R22A/ R34A	V_2	2-fold increase	Large reduction in cytotoxicity; large reduction in bactericidal activity against Gram-negative *E. coli*; large reduction in bactericidal activity against Gram-positive *B. subtilis*; large reduction of antiparasitic activity	(Singh and Batra 2011)
HsR3	R22A/ R61A		2-fold increase	Reduction in cytotoxicity; large reduction in bactericidal activity against Gram-negative *E. coli*; large reduction in bactericidal activity against Gram-positive *B. subtilis*; reduction of antiparasitic activity	(Singh and Batra 2011)
HsR3	R22A/ R77A		4-fold increase	Reduction in cytotoxicity; abolition of bactericidal activity against Gram- negative *E. coli*; large reduction of bactericidal activity against Gram-positive *B. subtilis*; reduction of antiparasitic activity	(Singh and Batra 2011)
HsR3	R22A/ H64A		5-fold increase	No reduction in cytotoxicity; small reduction in bactericidal activity against Gram-negative *E. coli*; no reduction in bactericidal activity against Gram-positive *B. subtilis*; reduction of antiparasitic activity	(Singh and Batra 2011)
HsR3	R34A/ H64A		1.25-fold decrease	Large reduction in cytotoxicity; small reduction in bactericidal activity against Gram-negative *E. coli*; large reduction in bactericidal activity against Gram-positive *B. subtilis*; large reduction of antiparasitic activity	(Singh and Batra 2011)
HsR3	R61A/ H64A	V_1	1.4-fold decrease	Large reduction in cytotoxicity; large reduction in bactericidal activity against Gram-negative *E. coli*; large reduction in bactericidal activity against Gram-positive *B. subtilis*; large reduction of antiparasitic activity	(Singh and Batra 2011)
HsR3	H64A/ R77A	V_1	1.25-fold decrease	Reduction in cytotoxicity; large reduction in bactericidal activity against Gram-negative *E. coli*; large reduction in bactericidal activity against Gram-positive *B. subtilis*; large reduction in antiparasitic activity	(Singh and Batra 2011)

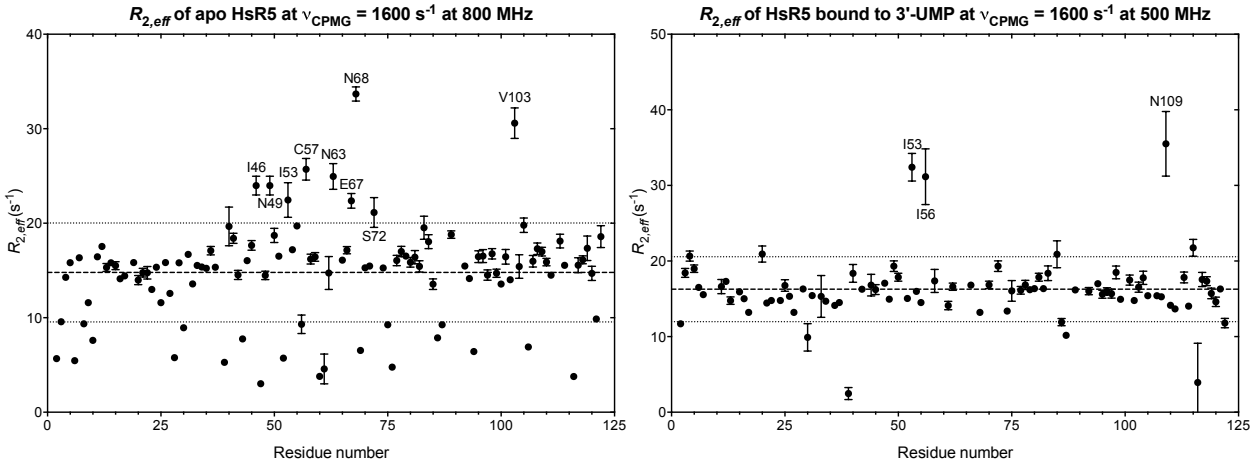
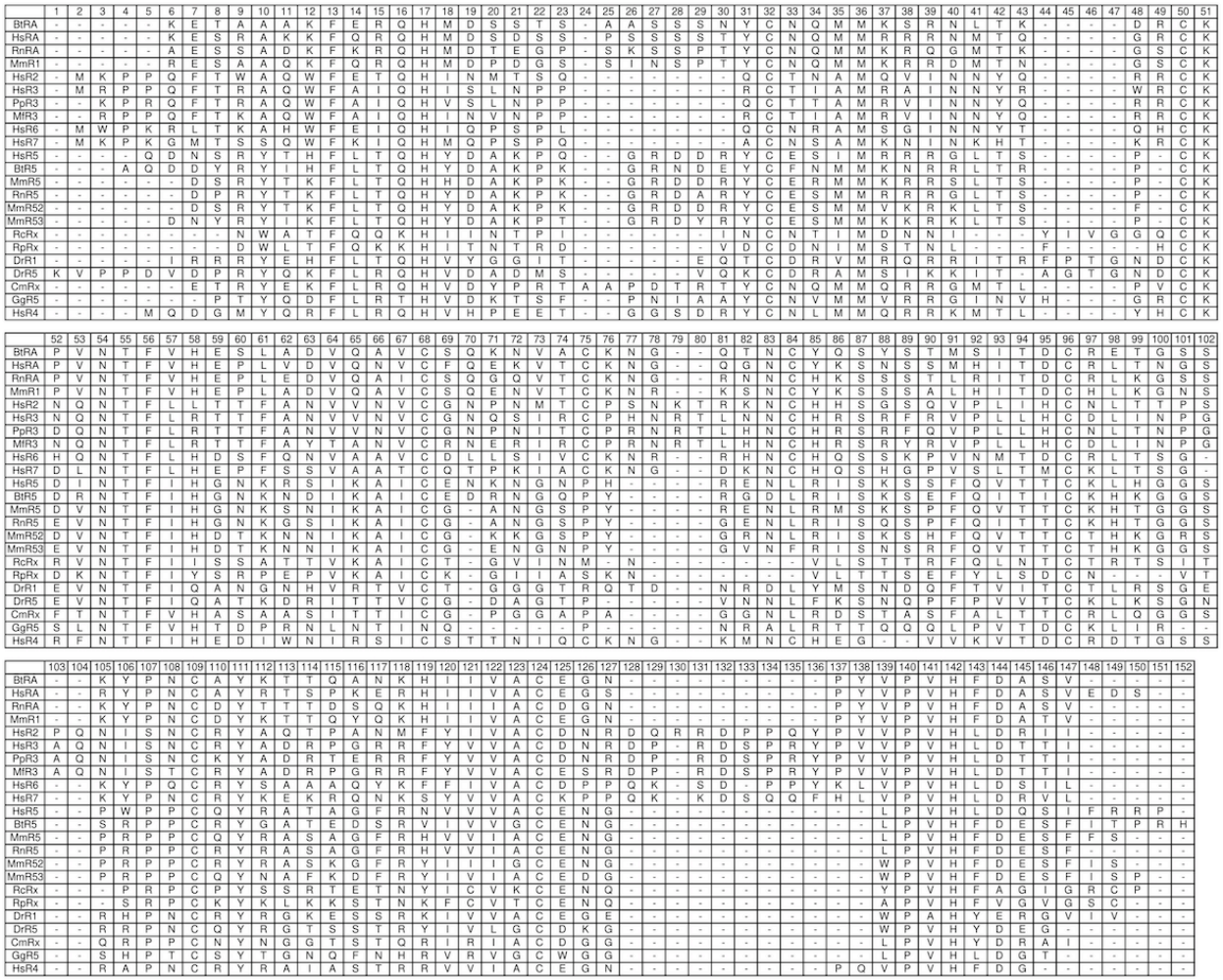


Figure 7.1 **Transverse relaxation rates (R_2) at ν_{CPMG} = 1600 s^{-1} (τ_{cp} = 0.625 ms) in free and ligand-bound HsR5.**

$R_{2,eff}$ values at ν_{CPMG} = 1600 s^{-1} are shown for all assigned residues of apo (left) and 3'-UMP- bound (right) HsR5. Dashed lines indicate the average $R_{2,eff}$ value and dotted lines indicate standard deviation. Residues exhibiting elevated $R_{2,eff}$ are identified on each graph. Upon closer examination of the full relaxation dispersion curves for these residues, most were disregarded since their $R_{2,eff}$ values were either inconsistently elevated and/or simply resulted from bad raw data. Only two residues displayed consistently elevated values: Asn68 in the apo protein, and Asn109 in the 3'-UMP bound protein. This could indicate the presence of dynamics at positions Asn68 (free form) and Asn109 (3'-UMP-bound form) on a microsecond timescale faster than that measurable by CPMG experiments. These observations further confirm the lack of conformational exchange on the CPMG timescale, and the overall likelihood that no significant functional dynamics is observed on the faster microsecond timescale in HsR5.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
BtRA	-	-	-	-	-	K	E	T	A	A	A	K	F	E	R	Q	H	M	D	S	S	T	S	-	A	A	S	S	S	N	Y	C	N	Q	M	M	K	S	R	N	L	T	K	-	-	-	-	D	R	C	K
HsRA	-	-	-	-	-	K	E	S	R	A	K	K	F	Q	R	Q	H	M	D	S	D	S	S	-	P	S	S	S	S	T	Y	C	N	Q	M	M	R	R	R	N	M	T	Q	-	-	-	-	G	R	C	K
RnRA	-	-	-	-	-	A	E	S	S	A	D	K	F	K	R	Q	H	M	D	T	E	G	P	-	S	K	S	S	P	T	Y	C	N	Q	M	M	K	R	Q	G	M	T	K	-	-	-	-	G	S	C	K
MmR1	-	-	-	-	-	R	E	S	A	A	Q	K	F	Q	R	Q	H	M	D	P	D	G	S	-	S	I	N	S	P	T	Y	C	N	Q	M	M	K	R	R	D	M	T	N	-	-	-	-	G	S	C	K
HsR2	-	M	K	P	P	Q	F	T	W	A	Q	W	F	E	T	Q	H	I	N	M	T	S	Q	-	-	-	-	-	-	-	Q	C	T	N	A	M	Q	V	I	N	N	Y	Q	-	-	-	-	R	R	C	K
HsR3	-	M	R	P	P	Q	F	T	R	A	Q	W	F	A	I	Q	H	I	S	L	N	P	P	-	-	-	-	-	-	-	R	C	T	I	A	M	R	A	I	N	N	Y	R	-	-	-	-	W	R	C	K
PpR3	-	-	K	P	R	Q	F	T	R	A	Q	W	F	A	I	Q	H	V	S	L	N	P	P	-	-	-	-	-	-	-	Q	C	T	T	A	M	R	V	I	N	N	Y	Q	-	-	-	-	R	R	C	K
MfR3	-	-	R	P	P	Q	F	T	K	A	Q	W	F	A	I	Q	H	I	N	V	N	P	P	-	-	-	-	-	-	-	R	C	T	I	A	M	R	V	I	N	N	Y	Q	-	-	-	-	R	R	C	K
HsR6	-	M	W	P	K	R	L	T	K	A	H	W	F	E	I	Q	H	I	Q	P	S	P	L	-	-	-	-	-	-	-	Q	C	N	R	A	M	S	G	I	N	N	Y	T	-	-	-	-	Q	H	C	K
HsR7	-	M	K	P	K	G	M	T	S	S	Q	W	F	K	I	Q	H	M	Q	P	S	P	Q	-	-	-	-	-	-	-	A	C	N	S	A	M	K	N	I	N	K	H	T	-	-	-	-	K	R	C	K
HsR5	-	-	-	-	Q	D	N	S	R	Y	T	H	F	L	T	Q	H	Y	D	A	K	P	Q	-	-	G	R	D	D	R	Y	C	E	S	I	M	R	R	R	G	L	T	S	-	-	-	-	P	-	C	K
BtR5	-	-	-	A	Q	D	D	Y	R	Y	I	H	F	L	T	Q	H	Y	D	A	K	P	K	-	-	G	R	N	D	E	Y	C	F	N	M	M	K	N	R	R	L	T	R	-	-	-	-	P	-	C	K
MmR5	-	-	-	-	-	-	D	S	R	Y	T	K	F	L	T	Q	H	H	D	A	K	P	K	-	-	G	R	D	D	R	Y	C	E	R	M	M	K	R	R	S	L	T	S	-	-	-	-	P	-	C	K
RnR5	-	-	-	-	-	-	D	P	R	Y	T	K	F	L	T	Q	H	Y	D	A	K	P	K	-	-	G	R	D	A	R	Y	C	E	S	M	M	R	R	R	G	L	T	S	-	-	-	-	P	-	C	K
MmR52	-	-	-	-	-	-	D	S	R	Y	T	K	F	L	T	Q	H	Y	D	A	K	P	K	-	-	G	R	D	D	R	Y	C	E	S	M	M	V	K	R	K	L	T	S	-	-	-	-	F	-	C	K
MmR53	-	-	-	-	-	D	N	Y	R	Y	I	K	F	L	T	Q	H	Y	D	A	K	P	T	-	-	G	R	D	Y	R	Y	C	E	S	M	M	K	K	R	K	L	T	S	-	-	-	-	P	-	C	K
RcRx	-	-	-	-	-	-	-	-	N	W	A	T	F	Q	Q	K	H	I	I	N	T	P	I	-	-	-	-	-	-	I	N	C	N	T	I	M	D	N	N	I	-	-	-	Y	I	V	G	G	Q	C	K
RpRx	-	-	-	-	-	-	-	-	D	W	L	T	F	Q	K	K	H	I	T	N	T	R	D	-	-	-	-	-	-	V	D	C	D	N	I	M	S	T	N	L	-	-	-	F	-	-	-	-	H	C	K
DrR1	-	-	-	-	-	I	R	R	R	Y	E	H	F	L	T	Q	H	V	Y	G	G	I	T	-	-	-	-	-	-	E	Q	T	C	D	R	V	M	R	Q	R	R	I	T	R	F	P	T	G	N	D	C	K
DrR5	K	V	P	P	D	V	D	P	R	Y	Q	K	F	L	R	Q	H	V	D	A	D	M	S	-	-	-	-	-	V	Q	K	C	D	R	A	M	S	I	K	K	I	T	-	A	G	T	G	N	D	C	K
CmRx	-	-	-	-	-	-	E	T	R	Y	E	K	F	L	R	Q	H	V	D	A	K	P	T	A	A	P	D	T	R	T	Y	C	N	Q	M	M	Q	R	R	G	M	T	L	-	-	-	-	P	V	C	K
GgR5	-	-	-	-	-	-	-	P	T	Y	Q	D	F	L	R	T	H	V	D	K	T	S	F	-	-	P	N	I	A	A	Y	C	N	V	M	M	V	R	R	G	I	N	V	H	-	-	-	G	R	C	K
HsR4	-	-	-	-	M	Q	D	G	M	Y	Q	R	F	L	R	Q	H	V	H	P	E	E	T	-	-	G	G	S	D	R	Y	C	N	L	M	M	Q	R	R	K	M	T	L	-	-	-	Y	H	C	K

	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102
BtRA	P	V	N	T	F	V	H	E	S	L	A	D	V	Q	A	V	C	S	Q	K	N	V	A	C	K	N	G	-	-	Q	T	N	C	Y	Q	S	Y	S	T	M	S	I	T	D	C	R	E	T	G	S	S
HsRA	P	V	N	T	F	V	H	E	P	L	V	D	V	Q	N	V	C	F	Q	E	K	V	T	C	K	N	G	-	-	Q	G	N	C	Y	K	S	N	S	S	M	H	I	T	D	C	R	L	T	N	G	S
RnRA	P	V	N	T	F	V	H	E	P	L	E	D	V	Q	A	I	C	S	Q	G	Q	V	T	C	K	N	G	-	-	R	N	N	C	H	K	S	S	S	T	L	R	I	T	D	C	R	L	K	G	S	S
MmR1	P	V	N	T	F	V	H	E	P	L	A	D	V	Q	A	V	C	S	Q	E	N	V	T	C	K	N	R	-	-	K	S	N	C	Y	K	S	S	S	A	L	H	I	T	D	C	H	L	K	G	N	S
HsR2	N	Q	N	T	F	L	L	T	T	F	A	N	V	V	N	V	C	G	N	P	N	M	T	C	P	S	N	K	T	R	K	N	C	H	H	S	G	S	Q	V	P	L	I	H	C	N	L	T	T	P	S
HsR3	N	Q	N	T	F	L	R	T	T	F	A	N	V	V	N	V	C	G	N	Q	S	I	R	C	P	H	N	R	T	L	N	N	C	H	R	S	R	F	R	V	P	L	L	H	C	D	L	I	N	P	G
PpR3	D	Q	N	T	F	L	R	T	T	F	A	N	V	V	N	V	C	G	N	P	N	I	T	C	P	R	N	R	T	L	H	N	C	H	R	S	R	F	Q	V	P	L	L	H	C	N	L	T	N	P	G
MfR3	N	Q	N	T	F	L	R	T	T	F	A	Y	T	A	N	V	C	R	N	E	R	I	R	C	P	R	N	R	T	L	H	N	C	H	R	S	R	Y	R	V	P	L	L	H	C	D	L	I	N	P	G
HsR6	H	Q	N	T	F	L	H	D	S	F	Q	N	V	A	A	V	C	D	L	L	S	I	V	C	K	N	R	-	-	R	H	N	C	H	Q	S	S	K	P	V	N	M	T	D	C	R	L	T	S	G	-
HsR7	D	L	N	T	F	L	H	E	P	F	S	S	V	A	A	T	C	Q	T	P	K	I	A	C	K	N	G	-	-	D	K	N	C	H	Q	S	H	G	P	V	S	L	T	M	C	K	L	T	S	G	-
HsR5	D	I	N	T	F	I	H	G	N	K	R	S	I	K	A	I	C	E	N	K	N	G	N	P	H	-	-	-	-	R	E	N	L	R	I	S	K	S	S	F	Q	V	T	T	C	K	L	H	G	G	S
BtR5	D	R	N	T	F	I	H	G	N	K	N	D	I	K	A	I	C	E	D	R	N	G	Q	P	Y	-	-	-	-	R	G	D	L	R	I	S	K	S	S	E	F	Q	I	T	I	C	K	H	K	G	G	S
MmR5	D	V	N	T	F	I	H	G	N	K	S	N	I	K	A	I	C	G	-	A	N	G	S	P	Y	-	-	-	-	R	E	N	L	R	M	S	K	S	P	F	Q	V	T	T	C	K	H	T	G	G	S
RnR5	E	V	N	T	F	I	H	G	N	K	G	S	I	K	A	I	C	G	-	A	N	G	S	P	Y	-	-	-	-	G	E	N	L	R	I	S	Q	S	P	F	Q	I	T	T	C	K	H	T	G	G	S
MmR52	D	V	N	T	F	I	H	D	T	K	N	N	I	K	A	I	C	G	-	K	K	G	S	P	Y	-	-	-	-	G	R	N	L	R	I	S	K	S	H	F	Q	V	T	T	C	T	H	K	G	R	S
MmR53	E	V	N	T	F	I	H	D	T	K	N	N	I	K	A	I	C	G	-	E	N	G	N	P	Y	-	-	-	-	G	V	N	F	R	I	S	N	S	R	F	Q	V	T	T	C	T	H	K	G	G	S
RcRx	R	V	N	T	F	I	I	S	S	A	T	T	V	K	A	I	C	T	-	G	V	I	N	M	-	N	-	-	-	-	-	-	-	V	L	S	T	T	R	F	Q	L	N	T	C	T	R	T	S	I	T
RpRx	D	K	N	T	F	I	Y	S	R	P	E	P	V	K	A	I	C	K	-	G	I	I	A	S	K	N	-	-	-	-	-	-	-	V	L	T	T	S	E	F	Y	L	S	D	C	N	-	-	-	V	T
DrR1	E	V	N	T	F	I	Q	A	N	G	N	H	V	R	T	V	C	T	-	G	G	G	T	R	Q	T	D	-	-	N	R	D	L	Y	M	S	N	D	Q	F	T	V	I	T	C	T	L	R	S	G	E
DrR5	E	V	N	T	F	I	Q	A	T	K	D	R	I	T	T	V	C	G	-	D	A	G	T	P	-	-	-	-	-	V	N	N	L	F	K	S	N	Q	P	F	P	V	V	T	C	K	L	K	S	G	N
CmRx	F	T	N	T	F	V	H	A	S	A	A	S	I	T	T	I	C	G	-	P	G	G	A	P	A	-	-	-	-	G	G	N	L	R	D	S	T	A	S	F	A	L	T	T	C	R	L	Q	G	G	S
GgR5	S	L	N	T	F	V	H	T	D	P	R	N	L	N	T	I	N	Q	-	-	-	-	-	P	-	-	-	-	-	N	R	A	L	R	T	T	Q	Q	Q	L	P	V	T	D	C	K	L	I	R	-	-
HsR4	R	F	N	T	F	I	H	E	D	I	W	N	I	R	S	I	C	S	T	T	N	I	Q	C	K	N	G	-	-	K	M	N	C	H	E	G	-	-	V	V	K	V	T	D	C	R	D	T	G	S	S

	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152
BtRA	-	-	K	Y	P	N	C	A	Y	K	T	T	Q	A	N	K	H	I	I	I	V	A	C	E	G	N	-	-	-	-	-	-	-	-	P	Y	V	P	V	H	F	D	A	S	V	-	-	-	-	-
HsRA	-	-	R	Y	P	N	C	A	Y	R	T	S	P	K	E	R	H	I	I	V	A	C	E	G	S	-	-	-	-	-	-	-	-	P	Y	V	P	V	H	F	D	A	S	V	E	D	S	-	-
RnRA	-	-	K	Y	P	N	C	D	Y	T	T	T	D	S	Q	K	H	I	I	I	A	C	D	G	N	-	-	-	-	-	-	-	-	P	Y	V	P	V	H	F	D	A	S	V	-	-	-	-	-
MmR1	-	-	K	Y	P	N	C	D	Y	K	T	T	Q	Y	Q	K	H	I	I	V	A	C	E	G	N	-	-	-	-	-	-	-	-	P	Y	V	P	V	H	F	D	A	T	V	-	-	-	-	-
HsR2	P	Q	N	I	S	N	C	R	Y	A	Q	T	P	A	N	M	F	Y	I	V	A	C	D	N	R	D	Q	R	R	D	P	P	Q	Y	P	V	V	P	V	H	L	D	R	I	I	-	-	-	-
HsR3	A	Q	N	I	S	N	C	R	Y	A	D	R	P	G	R	R	F	Y	V	V	A	C	D	N	R	D	P	-	R	D	S	P	R	Y	P	V	V	P	V	H	L	D	T	T	I	-	-	-	-
PpR3	A	Q	N	I	S	N	C	K	Y	A	D	R	T	E	R	R	F	Y	V	V	A	C	D	N	R	D	P	-	R	D	S	P	R	Y	P	V	V	P	V	H	L	D	T	T	I	-	-	-	-
MfR3	A	Q	N	I	S	T	C	R	Y	A	D	R	P	G	R	R	F	Y	V	V	A	C	E	S	R	D	P	-	R	D	S	P	R	Y	P	V	V	P	V	H	L	D	T	T	I	-	-	-	-
HsR6	-	-	K	Y	P	Q	C	R	Y	S	A	A	A	Q	Y	K	F	F	I	V	A	C	D	P	P	Q	K	-	S	D	-	P	P	Y	K	L	V	P	V	H	L	D	S	I	L	-	-	-	-
HsR7	-	-	K	Y	P	N	C	R	Y	K	E	K	R	Q	N	K	S	Y	V	V	A	C	K	P	P	Q	K	-	K	D	S	Q	Q	F	H	L	V	P	V	H	L	D	R	V	L	-	-	-	-
HsR5	-	-	P	W	P	P	C	Q	Y	R	A	T	A	G	F	R	N	V	V	V	A	C	E	N	G	-	-	-	-	-	-	-	-	-	-	-	L	P	V	H	L	D	Q	S	I	F	R	R	P	-
BtR5	-	-	S	R	P	P	C	R	Y	G	A	T	E	D	S	R	V	I	V	V	G	C	E	N	G	-	-	-	-	-	-	-	-	-	-	-	L	P	V	H	F	D	E	S	F	I	T	P	R	H
MmR5	-	-	P	R	P	P	C	Q	Y	R	A	S	A	G	F	R	H	V	V	I	A	C	E	N	G	-	-	-	-	-	-	-	-	-	-	-	L	P	V	H	F	D	E	S	F	F	S	-	-	-
RnR5	-	-	P	R	P	P	C	R	Y	R	A	S	A	G	F	R	H	V	V	I	A	C	E	N	G	-	-	-	-	-	-	-	-	-	-	-	L	P	V	H	F	D	E	S	F	-	-	-	-	-
MmR52	-	-	P	R	P	P	C	R	Y	R	A	S	K	G	F	R	Y	I	I	I	G	C	E	N	G	-	-	-	-	-	-	-	-	-	-	-	W	P	V	H	F	D	E	S	F	I	S	P	-	-
MmR53	-	-	P	R	P	P	C	Q	Y	N	A	F	K	D	F	R	Y	I	V	I	A	C	E	D	G	-	-	-	-	-	-	-	-	-	-	-	W	P	V	H	F	D	E	S	F	I	S	P	-	-
RcRx	-	-	-	P	R	P	C	P	Y	S	S	R	T	E	T	R	Y	I	C	V	K	C	E	N	Q	-	-	-	-	-	-	-	-	-	-	-	Y	P	V	H	F	A	G	I	G	R	C	P	-	-
RpRx	-	-	-	S	R	P	C	K	Y	K	L	K	K	S	T	N	K	F	C	V	T	C	E	N	Q	-	-	-	-	-	-	-	-	-	-	-	A	P	V	H	F	V	G	V	G	S	C	-	-	-
DrR1	-	-	R	H	P	N	C	R	Y	R	G	K	E	S	S	R	K	I	V	V	A	C	E	G	E	-	-	-	-	-	-	-	-	-	-	-	W	P	A	H	Y	E	R	G	V	I	V	-	-	-
DrR5	-	-	R	R	P	N	C	Q	Y	R	G	T	S	S	T	R	Y	I	V	L	G	C	D	K	G	-	-	-	-	-	-	-	-	-	-	-	W	P	V	H	Y	D	E	G	-	-	-	-	-	-
CmRx	-	-	Q	R	P	P	C	N	Y	N	G	G	T	S	T	Q	R	I	R	I	A	C	D	G	G	-	-	-	-	-	-	-	-	-	-	-	L	P	V	H	Y	D	R	A	I	-	-	-	-	-
GgR5	-	-	S	H	P	T	C	S	Y	T	G	N	Q	F	N	H	R	V	R	V	G	C	W	G	G	-	-	-	-	-	-	-	-	-	-	-	L	P	V	H	L	D	G	T	-	-	-	-	-	-
HsR4	-	-	R	A	P	N	C	R	Y	R	A	I	A	S	T	R	R	V	V	I	A	C	E	G	N	-	-	-	-	-	-	-	-	-	P	Q	V	P	V	H	F	D	G	-	-	-	-	-	-	-

Figure 7.2 Consensus sequence corresponding to the multiple sequence alignment of the 23 RNase sequences shown in Figure 3.2.

The first column and row labels correspond to sequence identifiers and amino acid positions, respectively.

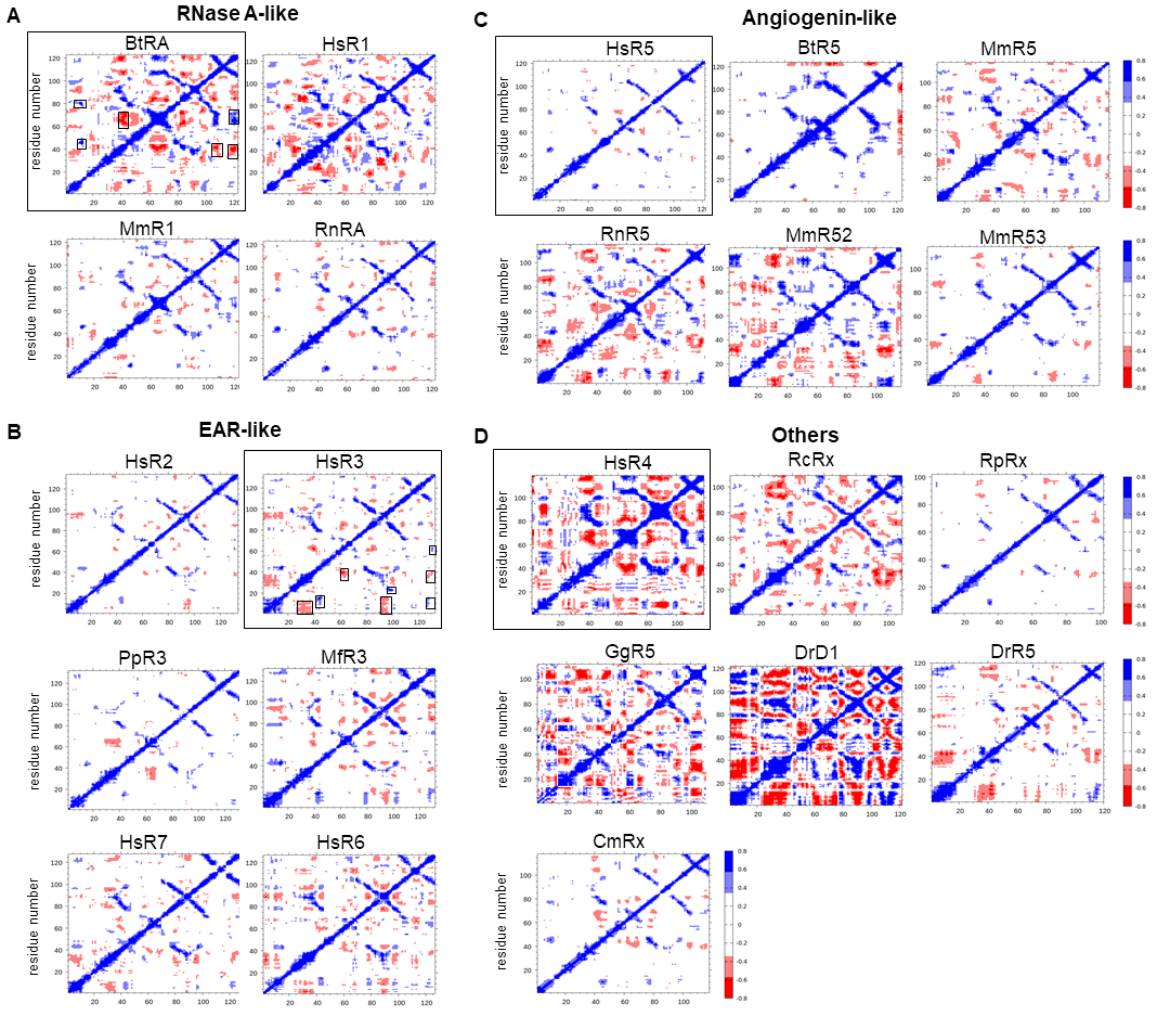


Figure 7.3 **Dynamic cross-correlation maps (DCCMs) for the 23 RNase homologs grouped into the four phylogenetic clusters.**

Values > 0.4 (blue end of the spectrum) represent correlated motions in the protein, while values < -0.4 (red end of the spectrum) represent anti-correlated motions. DCCMs identified using black outlines are discussed in the section “Microsecond dynamics” of the manuscript. Regions displaying (anti-)correlations are highlighted in the boxed DCCMs in A and B.

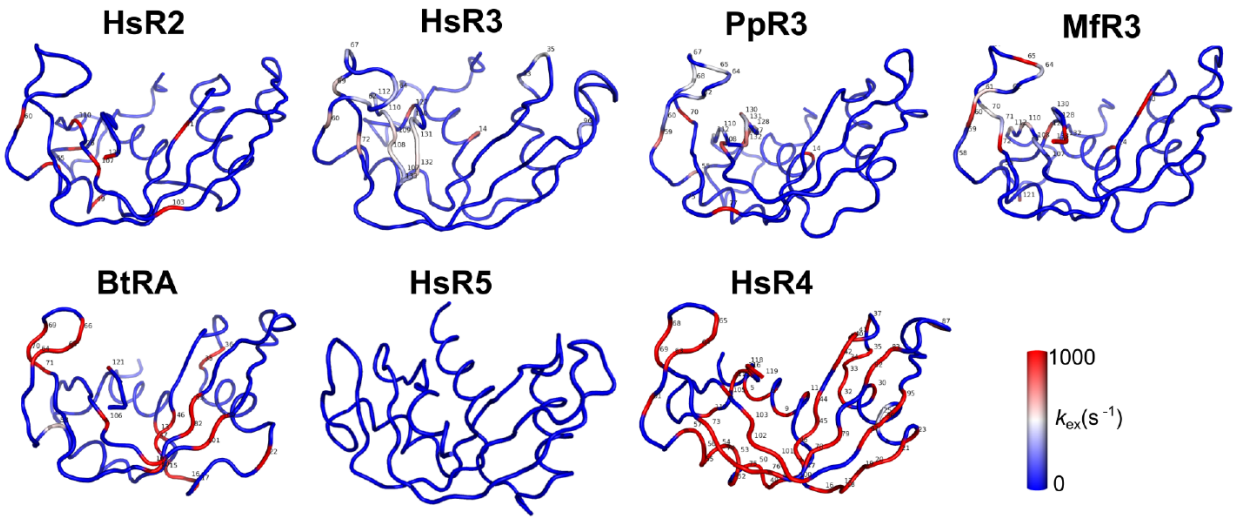


Figure 7.4 Conformational exchange rates of selected RNases.

Cartoon representation of *Homo sapiens* EDN (HsR2), *Homo sapiens* ECP (HsR3), *Pongo pygmaeus* ECP (PpR3), *Macaca fascicularis* ECP (MfR3), *Bos taurus* RNase A (BtRA), *Homo sapiens* angiogenin (HsR5) and *Homo sapiens* RNase 4 (HsR4). Structures are colored using a blue-white-red spectrum corresponding to residues exhibiting exchange rates (k_{ex}) ranging from 0 to 1000 s^{-1}. Residues experiencing enhanced conformational exchange appear in the red end of the spectrum while residues displaying no conformational exchange are in the blue end of the spectrum. Exchange rates (k_{ex}) were calculated for each individual residue by dual fitting 500 and 800 MHz CPMG data to the full single quantum relaxation-compensated CPMG equation (Manley and Loria 2012). Residues identified by NMR as undergoing conformational exchange on the millisecond time-scale are numbered according to their primary structure.

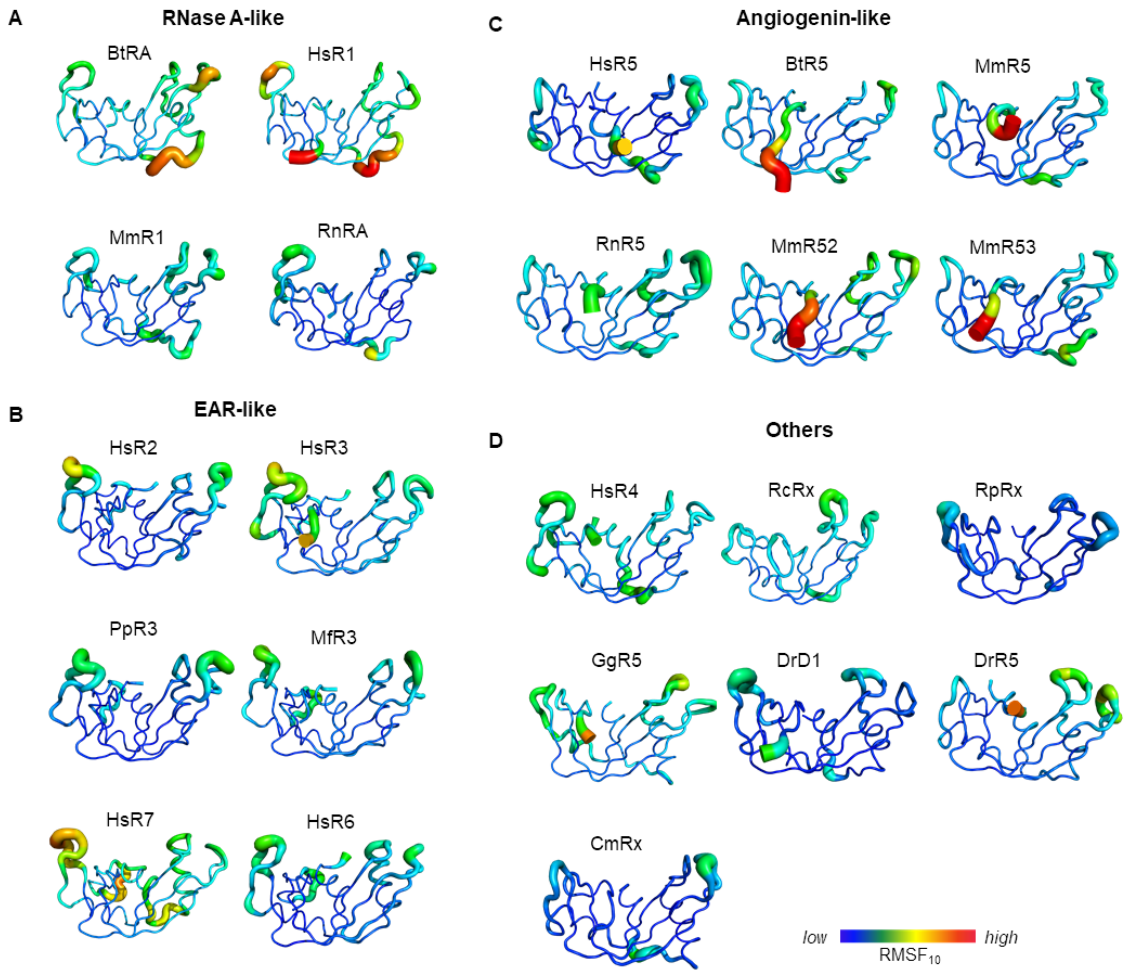


Figure 7.5 RMSFs corresponding to the top ten modes of all 23 RNase homologs from each of the four phylogenetic groups.

Structures are shown using a tube representation where the thickness of the tube corresponds to the flexibility of residues in each protein, with thicker tubes corresponding to flexible regions and thinner tubes representing less flexible regions. The dynamical range represented using the color spectrum is consistent across all sequences with the blue and red ends of the spectrum corresponding to low and high dynamic regions, respectively. N-terminal residues, which showed larger RMSFs (see left panel), are hidden for clarity.

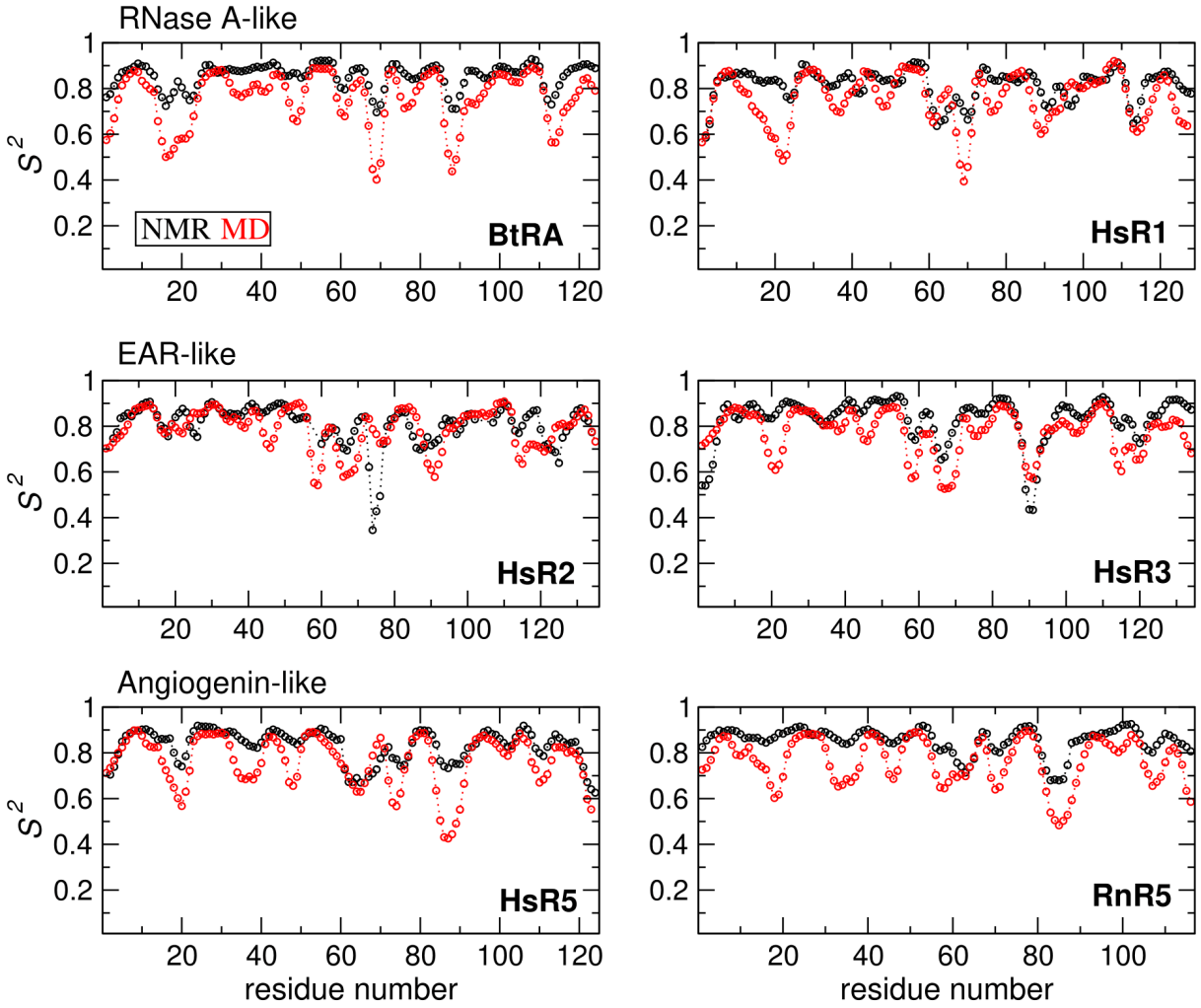


Figure 7.6 Faster timescale motions in representative RNases.

Comparison of the calculated order parameters for the experimental (black) and MD (red) data for two representative RNases each from the three functionally distinct RNase sub-families (RNase A-like, EAR-like and Angiogenin-like) defined in Figure 3.2C.

To determine if MD simulations capture the motions observed experimentally on the faster ps-ns timescale, we compared the square of the generalized order parameter (S^2). Using ShiftX2 (Han *et al*. 2011), we calculated the backbone chemical shifts for two representative members each from the three functionally distinct sub-families defined in Figure 3.2C. NMR chemical shifts for the selected RNases were obtained from the BMRB (Ulrich *et al*. 2008). Order parameters were calculated from the chemical shifts using TALOS+ (Shen *et al*. 2009). Comparison of the order parameters showed qualitative agreement between the experimental (black) and MD (red) data (Figure 6.6), suggesting that MD simulations capture the motions observed experimentally.

7.3 **Figures supplémentaires de l'article « Ligand-induced variations in structural and dynamical properties within an enzyme superfamily »**

Tableau 7.3 List of residues displaying $\Delta\delta_{obs} \geq 0.1$ ppm upon 3'-UMP and 5'-AMP binding to the five RNases

Residues displaying $\Delta\delta_{obs} \geq 0.1$ ppm upon UMP binding									
RNA_Bovine		RN2_Human		RN3_Human		RN4_Human		RN5_Human	
Residue	Δδ, ppm	Residue	Δδ, ppm	Residue	Δδ, ppm	Residue	Δδ, ppm	Residue	Δδ, ppm
11	0.145	7	0.147	4	0.281	5	0.105	11	0.125
12	0.137	11	0.241	11	0.134	20	0.125	42	0.172
17	0.149	15	0.195	14	0.194	42	0.349	50	0.113
41	0.223	16	0.102	16	0.162	44	0.310	51	0.157
44	0.490	37	0.229	31	0.129	103	0.114	53	0.120
45	1.197	42	0.213	40	0.335	116	0.127	56	0.124
46	0.258	43	0.247	48	0.118	117	0.261	61	0.107
47	0.110	64	0.145	55	0.109	118	0.121	65	0.125
51	0.180	66	0.101	57	0.105			72	0.106
66	0.152	82	0.162	62	0.123			84	0.159
82	0.163	83	0.209	64	0.302			103	0.137
96	0.149	103	0.108	66	0.238				
97	0.108	129	0.249	69	0.113				
99	0.101	130	0.477	75	0.279				
100	0.130	131	0.361	77	0.189				
101	0.187			84	0.346				
105	0.107			129	0.568				
119	0.225			130	0.136				
120	0.403								
122	0.206								

Residues displaying $\Delta\delta_{obs} \geq 0.1$ ppm upon AMP binding									
RNA_Bovine		RN2_Human		RN3_Human		RN4_Human		RN5_Human	
Residue	Δδ, ppm	Residue	Δδ, ppm	Residue	Δδ, ppm	Residue	Δδ, ppm	Residue	Δδ, ppm
11	0.166	7	0.101	10	0.279	5	0.183	12	0.122
43	0.107	8	0.208	41	0.112	11	0.100	14	0.179
45	0.227	11	0.244	64	0.118	20	0.118	25	0.158
46	0.133	14	0.250	110	0.132	44	0.125	106	0.106
51	0.239	15	0.164	128	0.235	50	0.158	117	0.112
58	0.249	39	0.109	129	0.169	66	0.178		
64	0.134	64	0.211	130	0.194	68	0.228		
65	0.450	66	0.101			70	0.201		
66	0.193	82	0.100			71	0.157		
68	0.139	83	0.105			115	0.110		
69	0.145	95	0.190			116	0.137		
71	0.190	128	0.236			117	0.625		
72	0.194	129	0.460			118	0.152		
101	0.116	130	0.332			119	0.208		
112	0.104	131	0.464						
118	0.149								
119	0.259								
120	0.325								
121	0.107								

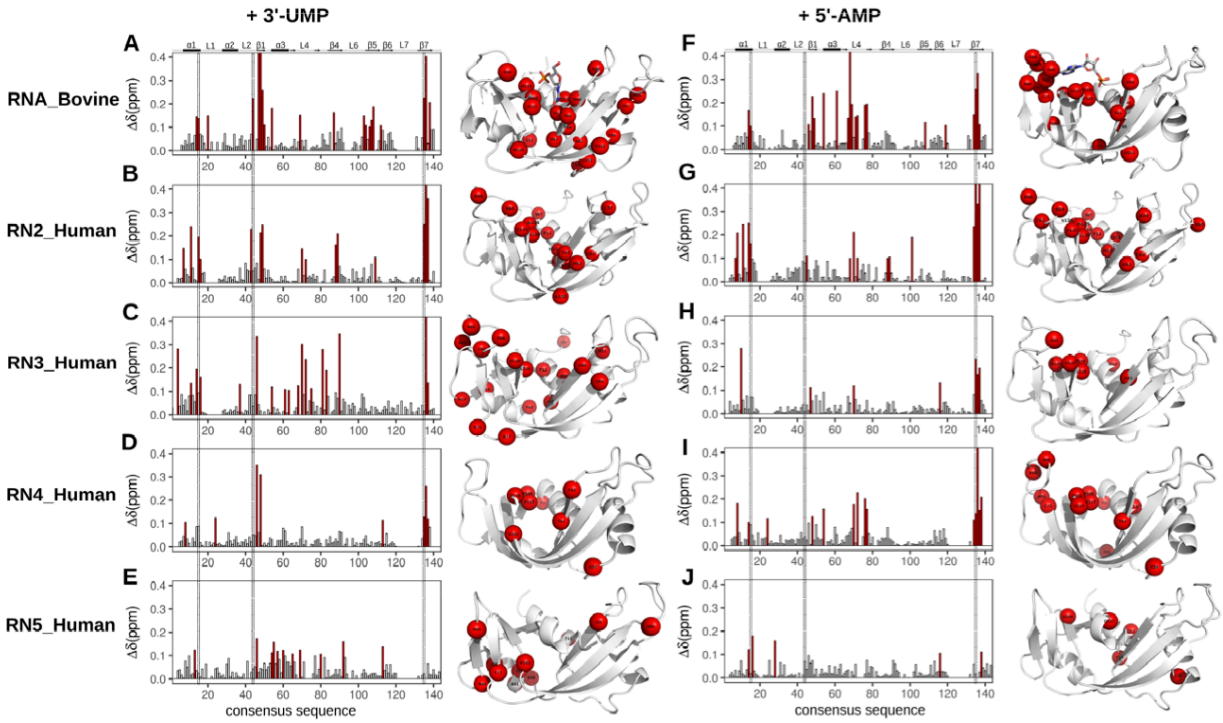


Figure 7.7 Chemical shift perturbations upon ligand binding to RNases

Compounded chemical shift variations ($\Delta\delta_{obs}$ at the highest ligand concentration) relative to the apo form upon binding of two mononucleotides 3'-UMP (A-E) and 5'-AMP (F-J) for bovine RNase A (A,F) and human RNases 2 (B,G), 3 (C-H), 4 (D,I) and 5 (E,J) are plotted as function of consensus sequence. Residues displaying $\Delta\delta_{obs}$ greater (lower) than 0.1 ppm are shown as red (white) bars. Active-site residues (His12, Lys41, His119, RNase A numbering) are highlighted using dashed lines. Residues with $\Delta\delta_{obs}$ > 0.1 ppm at the highest ligand concentration are depicted as spheres on the three-dimensional structures to the right of the plots.

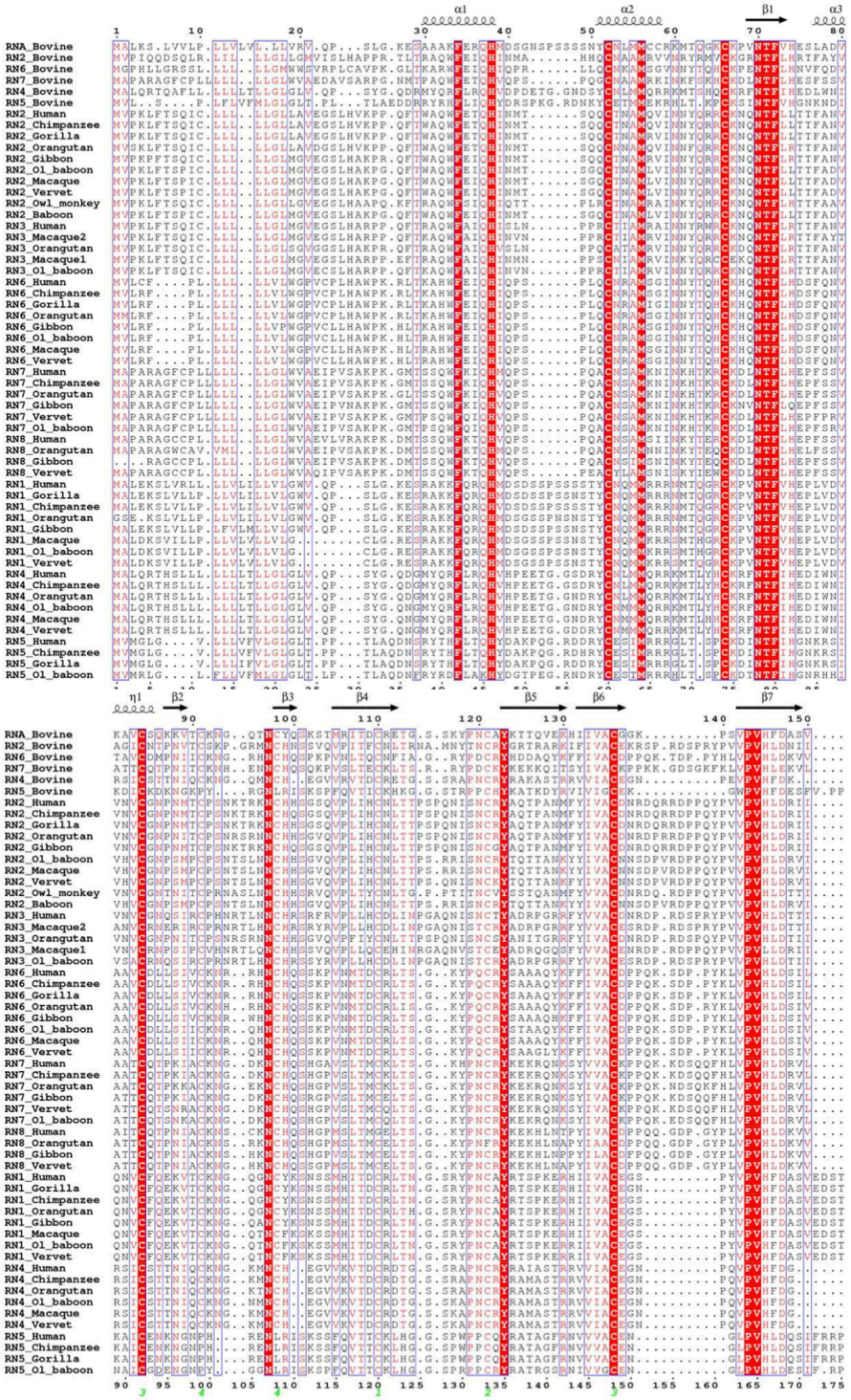


Figure 7.8 Multiple sequence alignment of bovine and Hominidae RNases 1-8

The primary structure numbering for bovine RNase A (RNA_Bovine) and the consensus sequence numbering are shown above and below the alignment, respectively. The α-helical and β-strands are identified above the alignment. Cysteine disulphide linkages are identified in green below the consensus sequence numbering.

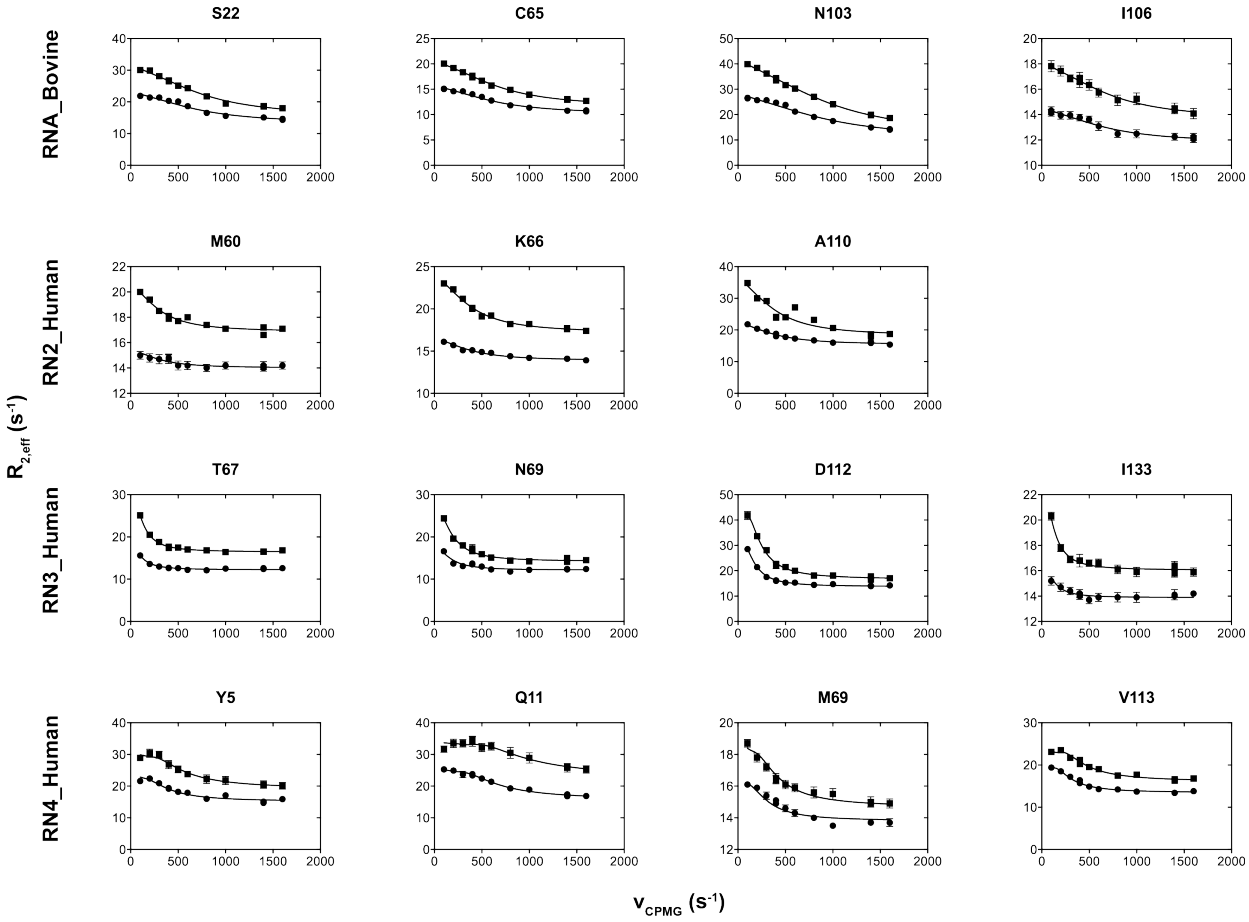


Figure 7.9 Representative ^{15}N-CPMG relaxation dispersion curves for the different RNases

Relaxation dispersion curves of select residues from the four RNases that display conformational exchange in the apo state. As RN5_Human does not display any such behavior, no curves are shown for this enzyme. Squares and circles respectively correspond to data recorded at 800 MHz and at 600 MHz (RNA_Bovine) or 500 MHz (RN2, RN3 and RN4_Human). Data were dual-fitted to the full Carver-Richards relaxation dispersion equation. Residues displaying (un)correlated dynamical changes were determined as described in the Methods section of the manuscript.

7.4 **Figures supplémentaires de l'article « Divergence of conformational motions along the evolutionary pathway in the eosinophil ribonuclease subfamily »**

Tableau 7.4 RNases depicted in Figure 5.2

Name	Type	Organism
GgR1	Ribonuclease 1	*Gorilla gorilla*
PaR1	Ribonuclease 1	*Pongo abelii*
MmR1	Ribonuclease 1	*Macaca mulatta*
PtR1	Ribonuclease 1	*Pan troglodytes*
HsR1	Ribonuclease 1	*Homo sapiens*
MoR1	Ribonuclease 1	*Mus musculus*
BtR1	Ribonuclease 1	*Bos taurus*
BtR2	Ribonuclease 2	*Bos taurus*
AtR2	Ribonuclease 2	*Aotus trivirgatus*
PpR2	Ribonuclease 2	*Pongo pygmaeus*
PaR2	Ribonuclease 2	*Pongo abelii*
GgR2	Ribonuclease 2	*Gorilla gorilla*
HsR2	Ribonuclease 2	*Homo sapiens*
PtR2	Ribonuclease 2	*Pan troglodytes*
Mo5R23	Eosinophil-Associated Ribonuclease 5	*Mus musculus*
Mo1R23	Eosinophil-Associated Ribonuclease 1	*Mus musculus*
MmR3	Ribonuclease 3	*Macaca mulatta*
MfR3	Ribonuclease 3	*Macaca fascicularis*
HsR3	Ribonuclease 3	*Homo sapiens*
PpR3	Ribonuclease 3	*Pongo pygmaeus*
PaR3	Ribonuclease 3	*Pongo abelii*
BtR4	Ribonuclease 4	*Bos taurus*
MoR4	Ribonuclease 4	*Mus musculus*
MmR4	Ribonuclease 4	*Macaca mulatta*
PaR4	Ribonuclease 4	*Pongo abelii*
PtR4	Ribonuclease 4	*Pan troglodytes*
HsR4	Ribonuclease 4	*Homo sapiens*
MoR5	Ribonuclease 5	*Mus musculus*
BtR5	Ribonuclease 5	*Bos taurus*
GgR5	Ribonuclease 5	*Gorilla gorilla*
PtR5	Ribonuclease 5	*Pan troglodytes*
HsR5	Ribonuclease 5	*Homo sapiens*
DrR5	Ribonuclease 5	*Danio rerio*
MoR6	Ribonuclease 6	*Mus musculus*
BtR6	Ribonuclease 6	*Bos taurus*
MmR6	Ribonuclease 6	*Macaca mulatta*
PaR6	Ribonuclease 6	*Pongo abelii*
HsR6	Ribonuclease 6	*Homo sapiens*
PtR6	Ribonuclease 6	*Pan troglodytes*
GgR6	Ribonuclease 6	*Gorilla gorilla*
PaR7	Ribonuclease 7	*Pongo abelii*
HsR7	Ribonuclease 7	*Homo sapiens*
PtR7	Ribonuclease 7	*Pan troglodytes*
BtR7	Ribonuclease 7	*Bos taurus*
PpR8	Ribonuclease 8	*Pongo pygmaeus*
HsR8	Ribonuclease 8	*Homo sapiens*

Table 7.5 Enzyme and substrate concentrations used in steady-state kinetics experiments

Enzyme	Enzyme concentration	Range of UpA concentration
	μM	*mM*
BtRA	0.002	0.1 – 1.5
HsR2	0.05	0.0 – 3.0
AtR2	2.0	0.7 – 4.2
HsR3	1.0	0.5 – 4.2
MfR3	2.0	0.2 – 3.5
PpR3	1.0	0.5 – 3.2
PaR3	1.9	0.2 – 3.0
HsR4	0.05	0.3 – 2.1
HsR5	9.0	0.8 – 4.0
HsR6	0.1	0.1 – 1.6

Tableau 7.6 Exchange rates determined by ^{15}N-CPMG

Position in alignment	BtRA	HsR2	AtR2	HsR3	MfR3	PpR3	PaR3
1	0	0	0	0	0	0	0
2	-	0	0	0	0	0	0
3	-	0	0	0	0	0	0
4	-	0	835.9	0	0	0	0
5	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0
10	898.7	0	0	0	0	0	0
11	0	0	513.8	0	0	0	0
12	0	0	283.8	0	0	0	0
13	0	0	0	0	0	0	0
14	0	0	355.4	120.9	1392	532.8	0
15	0	0	0	0	0	0	0
16	3069	0	0	0	0	0	0
17	808.7	0	0	0	0	0	0
18	4704	0	0	0	0	0	0
19	2747	-	-	-	-	-	-
20	2676	-	-	-	-	-	-
21	0.3898	-	-	-	-	-	-
22	1.857	0	0	0	0	0	0
23	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0
25	2119	0	0	0	0	0	0
26	0	0	0	0	0	0	0
27	0	-	-	-	-	-	-
28	1733	-	-	-	-	-	-
29	0	-	-	-	-	-	-
30	0	0	0	0	0	0	0
31	0	0	0	0	0	0	4185
32	6311	0	0	0	0	0	0
33	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0
35	0	0	0	0	0	0	0
36	1633	0	0	0	0	0	0
37	1279	0	393.3	0	0	0	0
38	957.6	0	0	0	0	0	0
39	552.1	0	0	308	0	0	0
40	783.6	0	0	0	0	0	0
41	0	0	0	473.6	0	0	0
42	0	0	0	0	0	0	0
43	0	0	0	0	0	0	0
44	1107	0	0	0	0	0	0
45	0	0	0	0	0	0	0
46	86.59	0	0	0	1377	0	0
47	0	1477	85.39	0	0	0	0
48	0	0	0	0	0	0	0
49	2663	0	0	0	0	0	0
50	116.1	0	0	0	0	0	557.9
51	1942	0	0	0	0	0	0
52	0	0	0	0	0	0	0
53	0	0	0	0	0	0	0
54	0	0	0	0	0	0	0
55	0	44.28	0	0	0	0	0
56	0	0	0	0	0	0	0
57	0	0	0	0	0	0	1614
58	0	0	0	0	0	0	0
59	0	939.7	0	0	0	0	1531
60	0	1137	0	0	0	0	1907
61	51.02	1342	323.3	0	16.85	336.2	888.2
62	0	0	0	0	0	0	0
63	0	0	0	0	0	0	0
64	0	0	263.7	0	17.9	0	0
65	1942	0	365.7	0	64	391.8	0
66	745.5	1061	309	623.1	452.9	153.9	0
67	1696	0	143.3	0	296.8	0	521.6
68	1915	0	339.7	408	14.24	223.8	0
69	577.9	0	0	0	0	0	0
70	0	0	220.2	182.1	235.7	144.5	1948
71	-	0	265.4	0	544.2	247.8	0
72	-	0	0	0	0	0	0
73	0	0	235.7	406.3	8.833	161.2	0
74	1621	0	113.9	857.5	0	244.7	935.2
75	2706	0	0	545	0	0	0
76	1825	0	169.6	624.9	262.9	2301	912.6
77	1288	0	350	0	198.4	0	0
78	0	0	341.2	617.1	556.1	0	733.9
79	1056	0	211.1	0	0	0	0
80	0	0	342.6	4662	0	0	0
81	0	0	322.3	0	0	35.18	0
82	0	0	0	0	0	0	0
83	0	0	223.6	0	0	2925	0
84	0	0	3.914	0	0	0	0
85	2893	0	0	0	0	0	0
86	0	0	251.7	0	0	0	1901
87	322.7	0	0	0	0	0	0
88	2663	0	0	0	0	0	0
89	0	0	0	0	0	0	0
90	0	0	0	0	0	0	0
91	0	0	0	0	0	0	0
92	0	0	0	0	0	0	0
93	0	0	0	0	0	0	0
94	0	0	0	0	0	0	0
95	0	0	-	0	0	0	0
96	-	0	0	0	0	0	0
97	0	0	-	0	0	0	0
98	0	0	0	0	0	0	0
99	-	0	0	0	0	0	0
100	0	0	0	0	0	0	0
101	0	0	0	0	0	0	0
102	0	0	0	330.9	0	0	0
103	0	0	0	0	0	0	0
104	0	0	0	0	0	0	0
105	0	0	0	0	0	0	0
106	0	0	0	0	0	0	0
107	3000	0	0	0	0	0	0
108	3193	0	0	0	0	0	0
109	0	1248	0	0	0	0	0
110	2915	0	0	0	0	0	1.881
111	0	0	0	0	0	0	862.8
112	0	0	0	0	0	0	1473
113	2066	0	220.9	337.9	34.54	0	882.8
114	0	0	326.8	519.4	163.4	429.6	1784
115	0	0	265.4	493.8	0	0	0
116	898.7	1222	240.6	377.9	304.1	205	1537
117	0	0	242.4	0	0	0	0
118	0	0	335	388.3	287.1	274.7	901.7
119	0	0	218.7	0	0	0	0
120	0	0	315.8	0	0	0	0
121	-	0	0	0	0	0	0
122	-	0	0	0	0	0	0
123	-	0	0	0	0	0	0
124	-	0	-	-	-	-	-
125	-	0	0	0	0	0	0
126	-	0	0	0	0	0	0
127	-	0	0	0	0	0	0
128	-	0	0	0	1825	0	0
129	-	0	0	0	0	0	0
130	0	0	0	0	0	0	0
131	0	0	0	0	0	0	0
132	0	0	0	0	0	0	0
133	0	0	0	0	0	0	0
134	0	0	0	547.3	459.7	1016	0
135	1950	1067	255.4	0	45.16	301.6	929.6
136	0	0	0	0	0	0	0
137	2132	0	334.4	342.1	157.7	369	0
138	0	80.53	0	514.6	0	233.1	1500
139	4279	1681	94.25	539.8	776.4	361.7	1704
140	0	834.3	255.4	375.2	581.6	0	0

k_{ex} values are given in s^{-1}. Dashes correspond to gaps in the sequence alignment.

Tableau 7.7 Exchange rates determined by ^{15}N-CEST

Residue in alignment	HsR2	AtR2	HsR3	MfR3	PpR3	PaR3
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0
6	0	0	0	0	0	0
7	0	0	0	0	0	0
8	0	0	0	0	0	0
9	0	0	0	0	0	0
10	0	0	0	0	0	0
11	0	0	0	0	0	0
12	0	0	0	0	0	0
13	0	0	0	0	0	0
14	0	0	0	0	0	0
15	0	0	0	0	0	0
16	0	0	0	0	0	0
17	0	0	0	0	0	0
18	0	0	0	0	0	0
19	-	-	-	-	-	-
20	-	-	-	-	-	-
21	-	-	-	-	-	-
22	0	0	0	0	0	0
23	0	0	0	0	0	0
24	0	0	0	0	0	0
25	0	0	0	0	0	0
26	0	0	0	0	0	0
27	-	-	-	-	-	-
28	-	-	-	-	-	-
29	-	-	-	-	-	-
30	0	0	0	0	0	0
31	0	0	0	0	0	0
32	0	0	0	0	0	0
33	0	0	0	0	0	0
34	0	0	0	0	0	0
35	0	0	0	0	0	0
36	0	0	0	0	0	0
37	0	0	0	0	0	0
38	0	0	0	0	0	0
39	0	0	0	0	0	0
40	0	0	0	0	0	0
41	0	0	0	0	0	0
42	0	0	0	0	0	0
43	0	0	0	0	0	0
44	0	0	0	0	0	0
45	0	0	0	0	0	0
46	0	0	0	0	0	0
47	0	0	0	0	0	0
48	0	0	0	0	0	0
49	0	0	0	0	0	0
50	0	0	0	0	0	0
51	0	0	0	0	0	0
52	0	0	0	0	0	0
53	0	0	0	0	0	0
54	0	0	0	0	0	0
55	0	0	0	0	0	0
56	0	0	0	0	0	0
57	0	0	0	0	0	0
58	0	0	0	0	0	0
59	0	0	0	0	0	0
60	0	426.8	0	0	0	0
61	0	426.8	0	0	0	0
62	0	0	0	0	0	0
63	0	0	0	0	0	0
64	0	426.8	458.72	0	0	0
65	0	0	0	182.53	0	0
66	0	426.8	458.72	182.53	362.76	0
67	0	287	0	0	0	0
68	0	0	302	270.6	362.76	0
69	0	0	0	0	0	0
70	0	426.8	458.72	36.47	199.01	0
71	0	29.07	0	270.6	362.76	0
72	0	0	0	0	0	0
73	0	0	0	0	362.76	0
74	0	0	25	0	362.76	0
75	0	0	0	0	0	0
76	0	426.8	0	36.47	600.75	0
77	0	426.8	0	0	0	0
78	0	426.8	0	0	0	0
79	0	15.74	0	0	0	0
80	0	15.74	0	0	0	0
81	0	426.8	0	0	0	0
82	0	15.74	0	0	0	0
83	0	15.74	0	0	0	0
84	0	306.11	0	0	0	0
85	0	0	0	0	0	0
86	0	0	0	0	0	0
87	0	0	0	0	0	0
88	0	0	0	0	0	0
89	0	0	0	0	0	0
90	0	0	0	0	0	0
91	0	0	0	0	0	0
92	0	0	0	0	0	0
93	0	0	0	0	0	0
94	0	0	0	0	0	0
95	0	-	0	0	0	0
96	0	0	0	0	0	0
97	0	-	0	0	0	0
98	0	0	0	0	0	0
99	0	0	0	0	0	0
100	0	0	0	0	0	0
101	0	0	0	0	0	0
102	0	0	0	0	0	0
103	0	0	0	0	0	0
104	0	0	0	0	0	0
105	0	0	0	0	0	0
106	0	0	0	0	0	0
107	0	0	0	0	0	0
108	0	0	0	0	0	0
109	0	0	0	0	0	0
110	0	0	0	0	0	0
111	0	0	0	0	0	0
112	0	0	0	0	0	0
113	0	15.74	0	0	0	0
114	0	57.04	458.72	182.53	0	0
115	0	0	0	0	0	0
116	0	426.8	458.72	182.53	362.76	0
117	0	0	0	0	0	0
118	0	426.8	458.72	182.53	0	0
119	0	0	0	0	0	0
120	0	0	0	0	0	0
121	0	0	0	0	0	0
122	0	0	0	0	0	0
123	0	0	0	0	0	0
124	0	-	-	-	-	-
125	0	0	0	0	0	0
126	0	0	0	0	0	0
127	0	0	0	0	0	0
128	0	0	0	0	0	0
129	0	0	0	0	0	0
130	0	0	0	0	0	0
131	0	0	0	0	0	0
132	0	0	0	0	0	0
133	0	0	0	0	0	0
134	0	0	0	182.53	0	0
135	0.0059	426.8	458.72	182.53	362.76	0
136	0	0	0	182.53	0	0
137	0	0	0	182.53	362.76	0
138	0	0	0	0	362.76	0
139	0	0	0	270.6	362.76	0
140	0	0	0	0	0	0

k_{ex} values are given in s^{-1}. Dashes correspond to gaps in the sequence alignment.

Tableau 7.8 Implied timescales determined by MSM

Position in alignment	HsR2	AtR2	HsR3	MfR3	PpR3	PaR3
1	51.83	nan	63.57	nan	nan	nan
2	nan	nan	nan	nan	nan	237.11
3	nan	603.99	nan	nan	nan	383.34
4	nan	132.97	nan	nan	249.44	469.84
5	nan	76.99	148.15	nan	nan	nan
6	nan	58.66	nan	nan	nan	nan
7	69.55	nan	100.53	nan	nan	nan
8	nan	nan	nan	nan	nan	nan
9	nan	nan	nan	nan	nan	nan
10	nan	nan	nan	nan	nan	70.55
11	nan	nan	nan	nan	nan	nan
12	nan	nan	nan	nan	154.45	nan
13	nan	nan	nan	nan	104.20	nan
14	nan	177.31	153.05	nan	146.50	nan
15	nan	205.63	77.76	111.85	149.89	nan
16	nan	76.89	nan	52.85	nan	167.00
17	nan	125.81	92.30	nan	nan	98.72
18	nan	nan	nan	nan	81.46	111.02
19	---	---	---	---	---	---
20	---	---	---	---	---	---
21	---	---	---	---	---	---
22	nan	89.09	82.37	nan	nan	86.68
23	nan	nan	50.91	nan	nan	64.59
24	nan	nan	74.69	nan	nan	242.22
25	nan	243.37	75.25	nan	nan	242.02
26	nan	nan	nan	nan	89.52	80.30
27	---	---	---	---	---	---
28	---	---	---	---	---	---
29	---	---	---	---	---	---
30	nan	nan	nan	nan	70.42	nan
31	nan	nan	nan	nan	nan	nan
32	nan	nan	nan	nan	nan	nan
33	nan	nan	nan	nan	nan	103.10
34	nan	nan	nan	nan	nan	88.36
35	nan	nan	nan	nan	nan	nan
36	nan	nan	nan	nan	nan	107.38
37	nan	nan	nan	nan	nan	nan
38	nan	205.71	nan	nan	nan	nan
39	234.64	229.71	nan	776.13	75.88	163.83
40	nan	69.93	131.99	763.13	nan	nan
41	70.88	nan	68.00	nan	nan	nan
42	nan	nan	51.24	nan	nan	nan
43	nan	nan	nan	nan	nan	nan
44	nan	nan	nan	nan	nan	nan
45	nan	nan	67.91	nan	nan	nan
46	nan	nan	251.46	nan	nan	nan
47	nan	nan	nan	nan	nan	nan
48	nan	nan	nan	nan	122.44	nan
49	nan	190.13	225.15	nan	nan	nan
50	nan	80.41	617.20	113.09	82.22	nan
51	nan	198.27	nan	nan	nan	174.95
52	nan	73.06	81.06	89.55	nan	233.73
53	nan	nan	nan	173.27	nan	161.88
54	nan	nan	nan	505.86	nan	nan
55	nan	nan	nan	118.25	nan	nan
56	nan	nan	nan	56.46	nan	176.24
57	nan	85.87	nan	72.57	nan	nan
58	nan	nan	nan	nan	nan	nan
59	nan	nan	nan	nan	nan	nan
60	nan	nan	nan	68.94	nan	nan
61	nan	108.61	110.38	74.56	nan	nan
62	nan	91.13	424.45	76.56	nan	nan
63	218.91	97.32	128.06	nan	nan	nan
64	nan	174.80	223.06	63.08	nan	nan
65	167.96	57.28	521.22	85.38	nan	nan
66	nan	155.07	545.52	nan	nan	nan
67	nan	270.93	nan	nan	753.82	95.87
68	nan	288.82	159.62	nan	183.28	103.23
69	nan	nan	nan	nan	nan	nan
70	281.58	122.82	nan	nan	546.29	101.10
71	279.14	336.20	227.78	nan	nan	75.10
72	274.16	229.65	166.09	nan	nan	78.79
73	nan	288.31	87.97	nan	166.68	98.07
74	288.14	230.63	130.62	nan	752.87	60.02
75	284.63	103.50	520.04	nan	753.46	nan
76	280.98	150.91	455.42	nan	100.51	nan
77	272.83	280.61	517.99	nan	104.45	104.34
78	171.75	82.32	nan	nan	94.15	nan
79	nan	58.87	nan	nan	nan	nan
80	nan	nan	431.84	nan	nan	nan
81	nan	nan	148.60	67.37	nan	nan
82	nan	nan	82.66	nan	nan	nan
83	nan	100.13	116.30	nan	nan	nan
84	56.40	78.21	431.19	118.83	131.62	60.99
85	nan	195.46	180.47	52.03	nan	nan
86	nan	243.35	384.95	nan	121.01	89.89
87	nan	81.57	87.91	nan	nan	115.66
88	37.60	121.20	423.15	51.38	236.47	146.27
89	nan	nan	65.85	nan	nan	nan
90	nan	nan	241.45	nan	134.55	165.18
91	nan	nan	nan	228.59	nan	139.74
92	nan	nan	nan	77.69	nan	335.85
93	422.29	nan	128.57	847.40	nan	204.22
94	424.04	nan	nan	nan	nan	397.25
95	423.66	---	nan	156.74	nan	132.87
96	423.04	nan	469.70	858.10	98.99	91.86
97	421.58	---	894.62	855.50	101.05	150.83
98	417.61	227.92	nan	856.15	nan	182.31
99	nan	nan	nan	nan	nan	nan
100	nan	nan	nan	nan	nan	nan
101	269.78	nan	131.41	nan	nan	118.56
102	nan	nan	nan	nan	76.24	122.90
103	nan	nan	nan	nan	nan	60.55
104	nan	nan	nan	nan	nan	nan
105	nan	nan	nan	nan	81.38	nan
106	nan	103.83	nan	nan	nan	106.43
107	nan	nan	161.19	nan	nan	nan
108	nan	83.44	nan	nan	nan	71.94
109	nan	157.22	nan	nan	nan	nan
110	nan	182.13	431.62	95.15	nan	nan
111	nan	245.91	122.96	nan	nan	97.72
112	nan	77.09	136.24	nan	140.73	53.95
113	nan	nan	428.63	nan	nan	nan
114	nan	nan	nan	nan	nan	nan
115	80.82	nan	nan	nan	nan	nan
116	nan	nan	nan	nan	nan	nan
117	nan	nan	nan	nan	nan	nan
118	nan	nan	nan	nan	nan	nan
119	nan	nan	nan	nan	nan	nan
120	nan	144.13	127.11	nan	108.62	86.61
121	nan	64.23	180.76	nan	66.24	102.96
122	nan	nan	543.42	nan	nan	nan
123	nan	210.25	442.07	nan	60.35	93.70
124	72.52	---	---	---	---	---
125	nan	99.28	nan	80.84	84.61	68.97
126	nan	nan	nan	nan	nan	nan
127	nan	nan	180.03	nan	nan	nan
128	nan	nan	556.99	nan	nan	nan
129	nan	179.12	555.99	nan	nan	nan
130	nan	nan	250.04	nan	nan	nan
131	nan	nan	229.48	nan	nan	nan
132	nan	nan	nan	nan	nan	75.34
133	nan	nan	nan	nan	nan	nan
134	nan	nan	nan	74.42	nan	114.45
135	nan	nan	nan	57.59	nan	nan
136	nan	62.43	194.19	82.55	nan	nan
137	140.47	265.11	223.23	nan	82.72	87.74
138	nan	236.92	67.19	nan	nan	nan
139	nan	156.59	338.03	nan	nan	nan
140	nan	nan	nan	nan	nan	nan

Values are given in ns. Dashes correspond to gaps in the sequence alignment. Values of “nan” are “not a number” and correspond either to timescales < 50 ns or to data that has not converged.

Figure 7.10 Sequence alignment of the RNases used in the phylogenetic tree.

Sequences were taken from UniProt Knowledge Base, and aligned using ClustalΩ. Catalytic residues are indicated with an asterisk above the alignment.

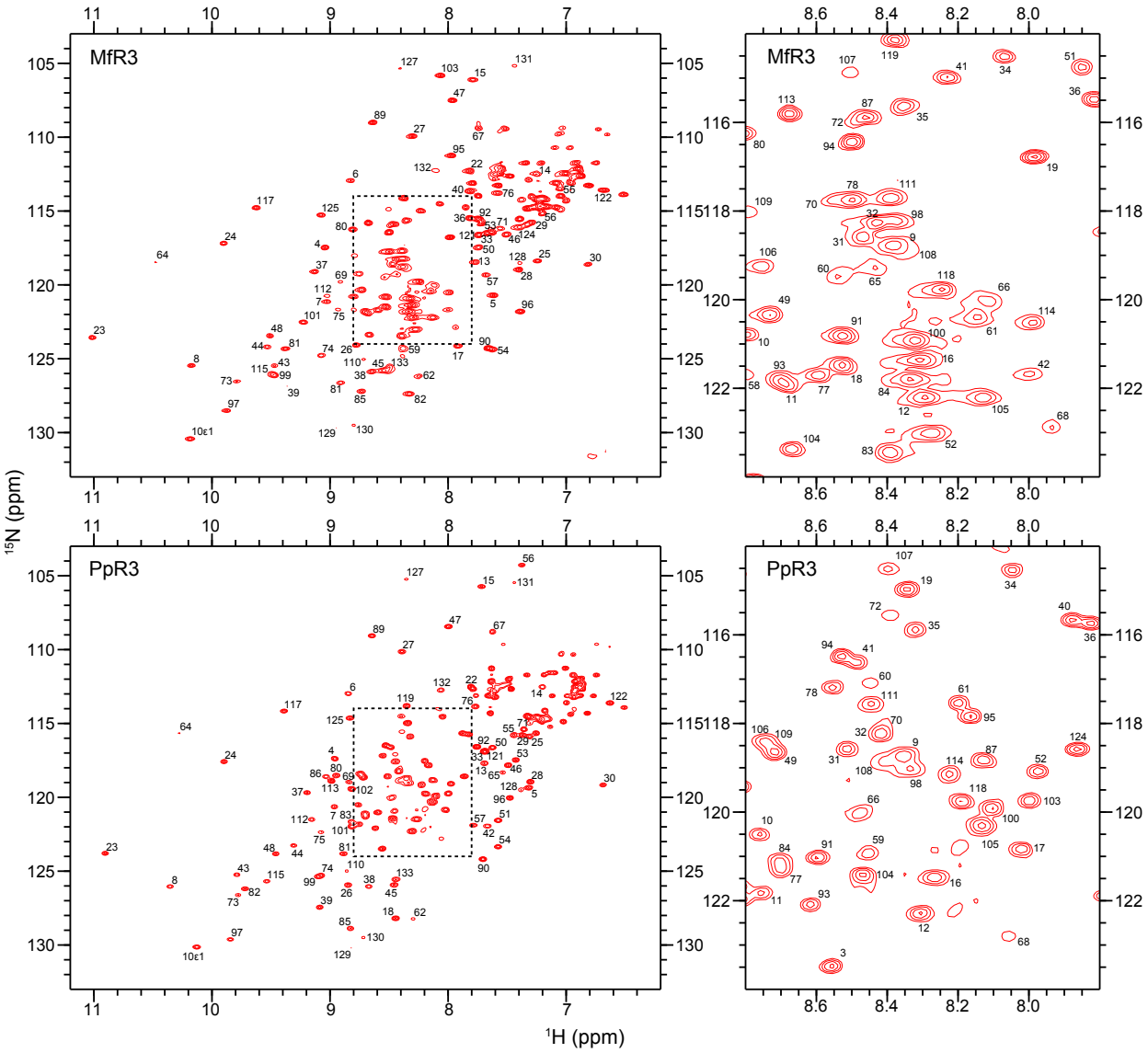


Figure 7.11 {^{1}H,^{15}N}-HSQC spectra of MfR3 and PpR3

Spectra were acquired at 298 K in 15 mM sodium acetate at pH 5.0 with 10% D_2O. Resonances are labeled using their position in the sequence of the mature protein, thus excluding the signal peptide sequence. The central region of each spectrum has been magnified to the right of its associated spectrum.

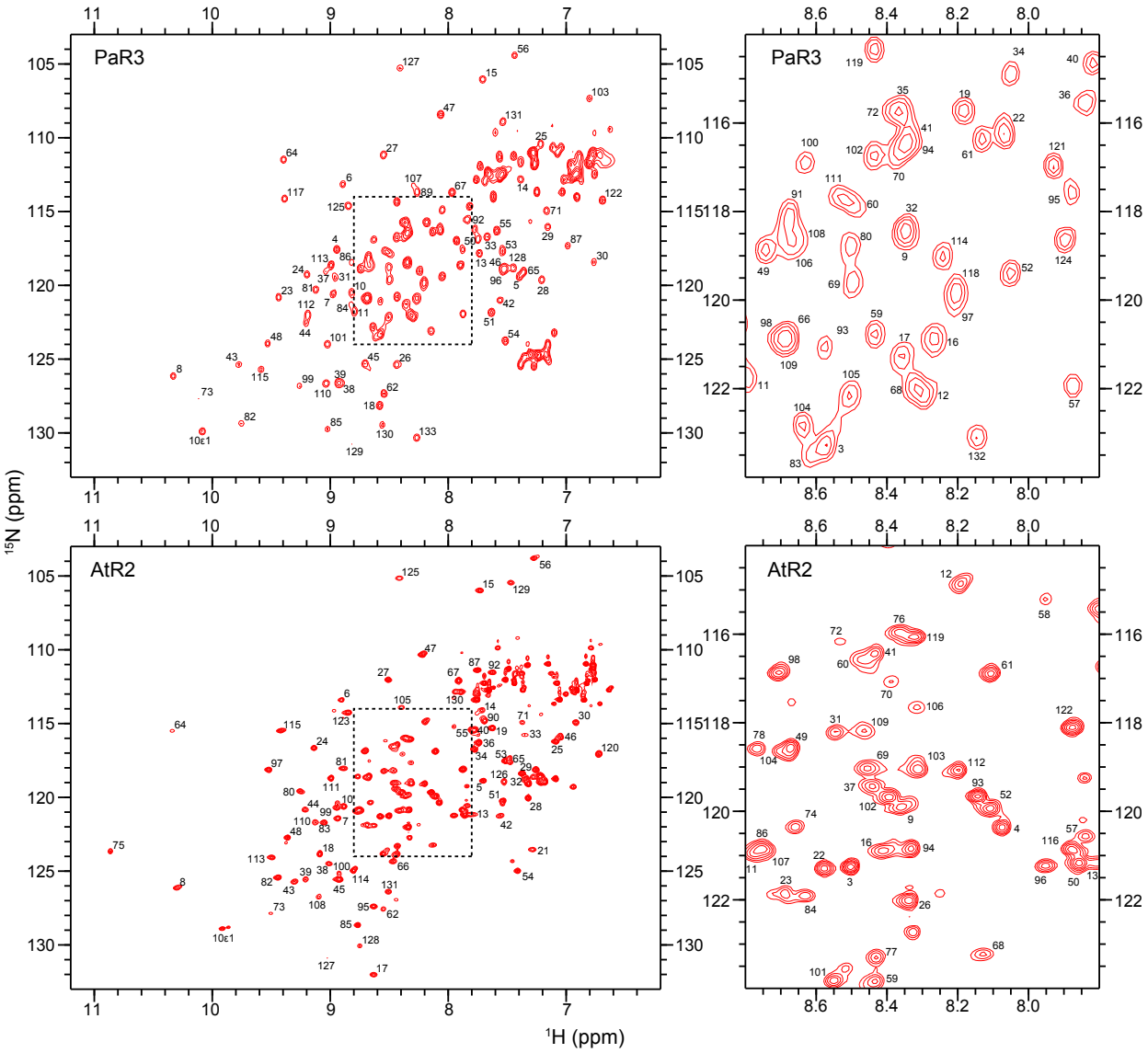


Figure 7.12 {^{1}H,^{15}N}-HSQC spectra of PaR3 and AtR2

Spectra were acquired at 298 K in 15 mM sodium acetate at pH 5.0 with 10% D_2O. Resonances are labeled using their position in the sequence of the mature protein, thus excluding the signal peptide sequence. The central region of each spectrum has been magnified to the right of its associated spectrum.

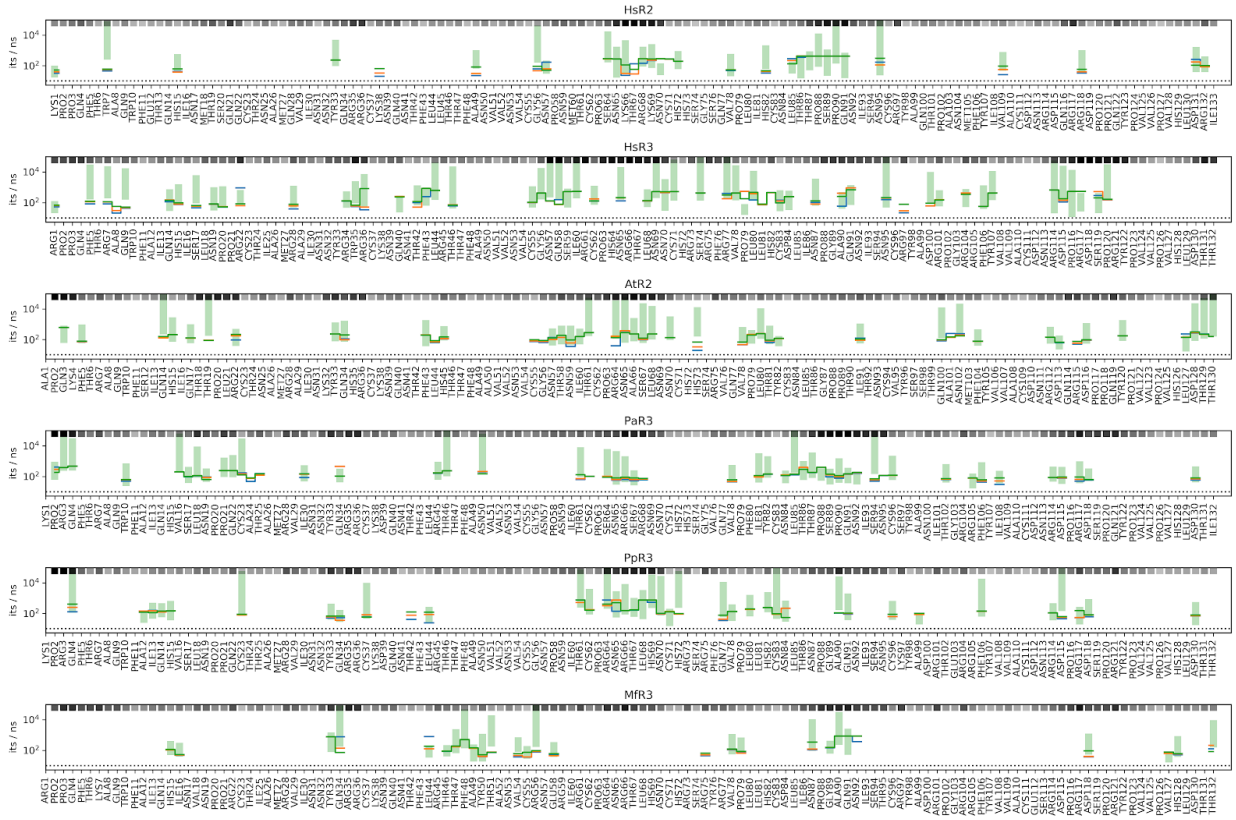


Figure 7.13 Error estimation for all 6 RNase MSM implied timescales

Only converged models are shown. Discretization was performed with 50, 100 and 150 cluster centers (depicted in orange, blue and green, respectively) to estimate discretization artifacts. Error bars (green) are obtained by Bayesian sampling of the posterior (Trendelkamp-Schroer *et al*. 2015) using the finest state discretization.

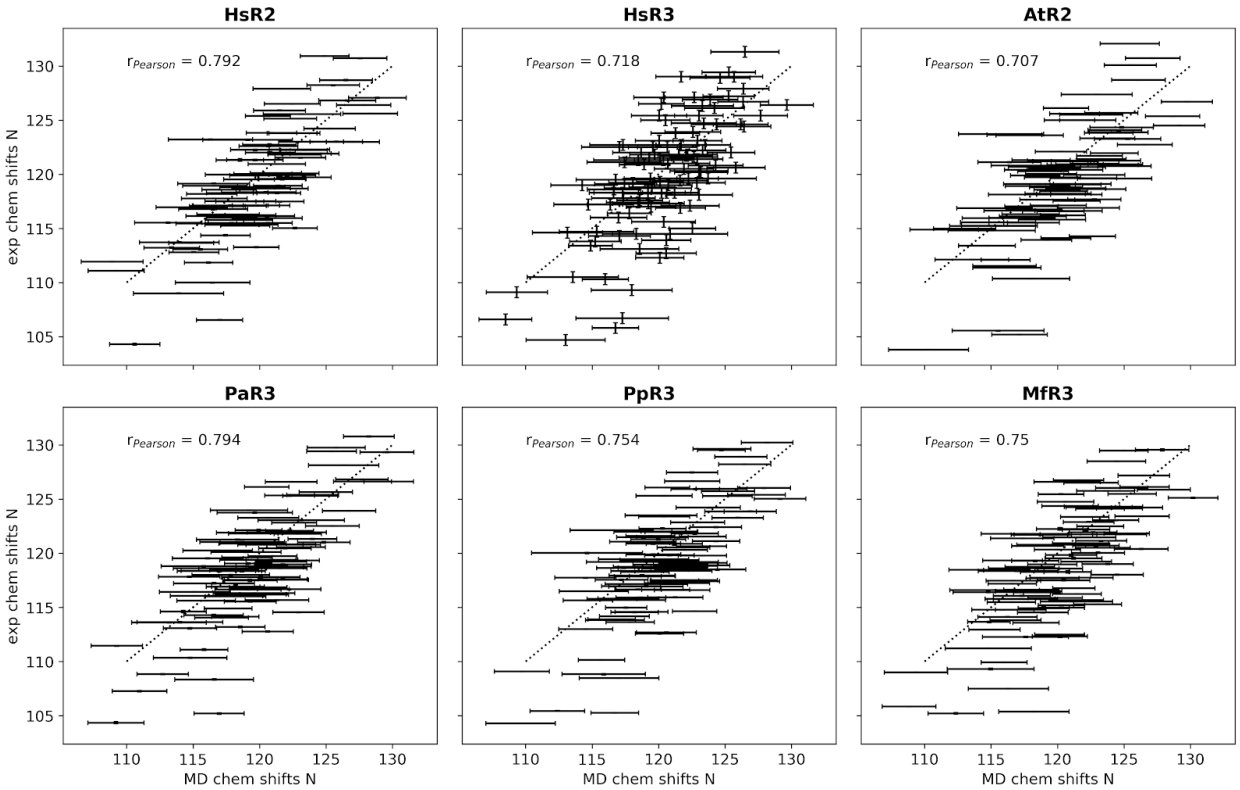


Figure 7.14 Validation of homology models

Comparison of ^{15}N chemical shifts derived from MD simulations (*x*-axis, mean over trajectories with standard deviation error bars) and NMR (*y*-axis). Pearson's correlation coefficient is annotated.

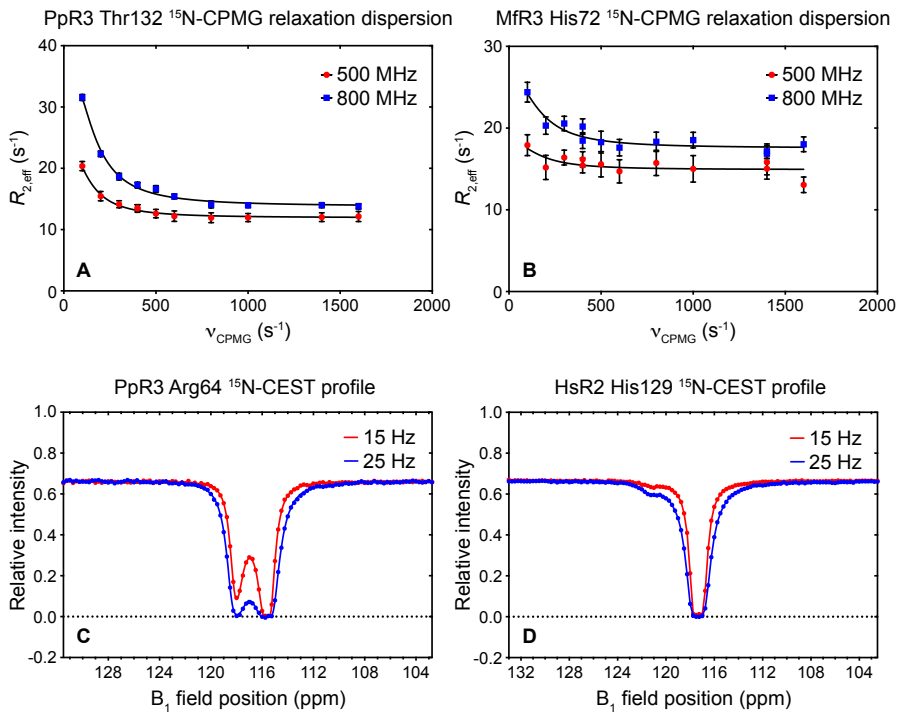


Figure 7.15 Typical NMR conformational exchange curves in eosinophil RNases

The depicted ^{15}N-CPMG relaxation dispersion curves exemplify both A) well-defined and B) ambiguous fits, as calculated using the full relaxation dispersion equation on GraphPad Prism. The ^{15}N-CEST curves depicted also exemplify both C) well-defined and D) ambiguous exchange data, as calculated by Chemex (Vallurupalli *et al*. 2012).

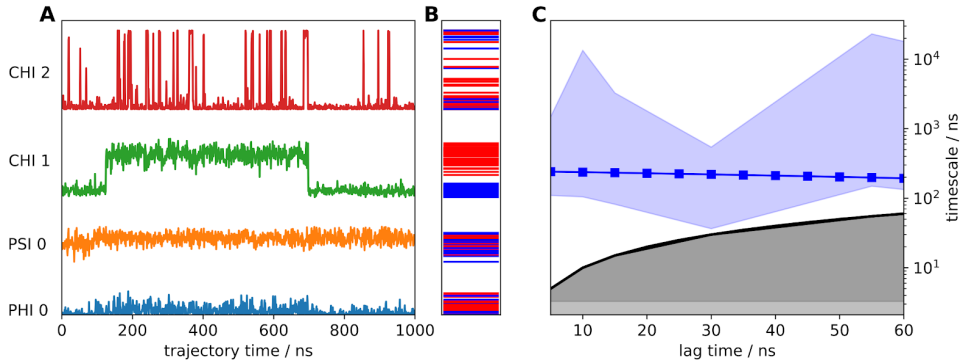


Figure 7.16 Example of timescale estimation in the ns-µs regime

A) Trajectory of single residue backbone torsion angles of His82 from PpR3 extracted from raw MD. Please note that the Chi1 angle describes the metastable dynamics. It is captured by the HMM eigenvector, which changes sign along this angle but not along the other angles (panel B, color coded eigenvector along torsion angles, shared *y*-axis with panel A). HMM implied timescales are estimated after discretization and transition matrix estimation (not shown) and checked for convergence (panel C). Maximum likelihood estimate (squares) is converged while the confidence interval (blue shaded area) mirrors that the metastable process has only been sampled once in both directions.

7.5 Articles supplémentaires rédigés au cours de mon doctorat

Tel que mentionné dans la section remerciements, j'ai eu la chance de travailler sur de nombreuses collaborations au cours de mon doctorat, et quelques-unes ont mené à des publications. J'ai pu travailler dans le cadre de l'imposant projet de recherche de Ugo Dionne (laboratoire de Pr. Nicolas Bisson, Université Laval) sur la protéine adaptatrice Nck1. Afin d'étudier l'effet de la phosphorylation d'un résidu tyrosine particulier d'un domaine SH3, j'ai dû déterminer l'effet de différentes mutations sur la structure du domaine, et j'ai fait le titrage RMN du domaine SH3 WT ou muté avec le peptide que reconnaît le WT. Ces travaux ont été publiés dans l'article « Direct phosphorylation of SH3 domains by tyrosine kinase receptors disassembles ligand-induced signaling networks » (Dionne *et al*. 2018), et un total de 19 chercheurs y ont contribué.

J'ai également pu apporter mon soutien au projet de recherche de Sarah Martinez (laboratoire de Pr. Éric Déziel, INRS) sur l'évolution dirigée de *Burkholderia thailendensis* pour une production accrue de rhamnolipides. J'ai prédit l'effet structural de mutations identifiées au sein d'un gène particulier, la protéine régulatrice de « quorum sensing » QsmR, ainsi que l'effet que ces mutations pourraient avoir sur la régulation génétique. Les résultats de ces travaux ont mené à la rédaction de l'article « Implementation of a directed evolution strategy for improved production of the microbial surfactant rhamnolipid » (Martinez *et al*. en préparation). L'article est entièrement rédigé et est sur le point d'être soumis.

Outre les collaborations avec d'autres laboratoires, j'ai également travaillé sur plusieurs projets de recherche au sein du laboratoire Doucet. J'ai apporté ma contribution à l'article « Ligand Binding Enhances Millisecond Conformational Exchange in Xylanase B2 from *Streptomyces lividans* » (Gagné *et al*. 2016) en déterminant les valeurs de R_1, R_2 et hNOE de la XlnB2 et en analysant la répartition structurale de ces différents taux de relaxation.

J'ai participé à la rédaction de l'article de revue « Enzyme dynamics : looking beyond a single structure » (Agarwal *et al*. soumis) qui a été soumis au Journal of Biological Chemistry, qui s'est intéressé aux techniques permettant d'étudier les mouvements internes chez les enzymes, à l'effet de l'environnement cellulaire sur ceux-ci, à leur rôle dans l'ingénierie des protéines et aux problématiques qui demeurent encore non résolues à leur sujet. J'ai rédigé les sections concernant les méthodes RMN de CPMG et de CEST, la technique bioinformatique de time-structure-based independent component analysis (tICA), ainsi que la section concernant le rôle de la flexibilité conformationnelle dans l'évolution dirigée d'enzymes.

Finalement, j'ai participé à un projet de recherche ayant mené à la rédaction de l'article « Insights into structural and dynamical changes experienced by human RNase 6 upon ligand binding » (Narayanan *et al*. accepté avec corrections). Pour cet article, j'ai effectué l'attribution RMN de la protéine (déposée dans la BMRB sous le code 27659), qui s'est avérée posséder deux conformations différentes n'échangeant pas entre elles, tel que confirmé par ^{15}N-CPMG, ^{15}N-CEST et ZZ exchange). J'ai également effectué le titrage de cette protéine avec le ligand dACACA, et les expériences de CPMG en présence et en absence de ce ligand.

7.5.1 Références

Agarwal PK, Bernard DN, Bafna K et Doucet N (soumis) Enzyme dynamics : looking beyond a single structure.

Dionne U, Chartier FJM, Lopez de los Santos Y, Lavoie N, Bernard DN, Banerjee SL, Otis F, *el al*. (2018) Direct phosphorylation of SH3 domains by tyrosine kinase receptors disassembles ligand-induced signaling networks, Molecular Cell 70, 1-13

Gagné D, Narayanan C, Nguyen-Thi N, Roux LD, Bernard DN, Brunzelle JS, Couture J-F, Agarwal PK & Doucet N (2016) Ligand Binding Enhances Millisecond Conformational Exchange in Xylanase B2 from Streptomyces lividans, Biochemistry 55, 4184-4196

Martinez S, Bernard DN & Déziel E (en préparation) Implementation of a directed evolution strategy for improved production of the microbial surfactant rhamnolipid

Narayanan C, Pham H, Bernard DN, Gagnon J, Couture J-F, Doucet N (accepté avec corrections) Biochemistry