Université de Montréal

Étude numérique des premières étapes d'agrégation du peptide amyloïde GNNQQNY, impliqué dans une maladie à prion.

par
Jessica Nasica-Labouze

Département de physique
Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)
en Physique - option biophysique et physiologie moléculaires

Août, 2012

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée:

Étude numérique des premières étapes d'agrégation du peptide amyloïde GNNQQNY, impliqué dans une maladie à prion.

présentée par:

Jessica Nasica-Labouze

a été évaluée par un jury composé des personnes suivantes:

Laurent Lewis,	président-rapporteur
Normand Mousseau,	directeur de recherche
Rikard Blunck,	membre du jury
Nicolas Moitessier,	examinateur externe
Sven Joubert,	représentant du doyen de la FES

Thèse acceptée le: . . . . . .

RÉSUMÉ

Les protéines amyloïdes sont impliquées dans les maladies neurodégénératives comme Alzheimer, Parkinson et les maladies à prions et forment des structures complexes, les fibres amyloïdes. Le mécanisme de formation de ces fibres est un processus complexe qui implique plusieurs espèces d'agrégats intermédiaires. Parmi ces espèces, des petits agrégats, les oligomères, sont reconnus comme étant l'espèce amyloïde toxique, mais leur mécanisme de toxicité et d'agrégation sont mal compris. Cette thèse présente les résultats d'une étude numérique des premières étapes d'oligomérisation d'un peptide modèle GNNQQNY, issu d'une protéine prion, pour des systèmes allant du trimère au 50-mère, par le biais de simulations de dynamique moléculaire couplée au potentiel gros-grain OPEP. Nous trouvons que le mécanisme d'agrégation du peptide GNNQQNY suit un processus complexe de nucléation, tel qu'observé expérimentalement pour plusieurs protéines amyloïdes. Nous observons aussi que plusieurs chemins de formation sont accessibles à l'échelle du 20-mère et du 50-mère, ce qui confère aux structures un certain degré de polymorphisme et nous sommes capable de reproduire, dans nos simulations, des oligomères protofibrillaires qui présentent des caractéristiques structurelles observées expérimentalement chez les fibres amyloïdes.

Mots clés: fibres amyloïdes, prion, GNNQQNY, agrégation de protéines, nucléation, polymorphisme, oligomères, protofibres.

ABSTRACT

Amyloid proteins are involved in neurodegenerative diseases such as Alzheimer's, Parkinson's and prion diseases and form complex structures called amyloid fibrils. The fibril formation mechanism is a complex process, which involves several intermediary species. Among these species, small early aggregates, called oligomers, are thought to be the toxic amyloid species but their toxicity and aggregation mechanisms are poorly understood. This thesis aims at presenting the results of a numerical study of the first oligomerization steps of the model peptide GNNQQNY, from a prion protein, for system sizes ranging from the trimer to the 50-mer, via molecular dynamics simulations using the OPEP coarse-grained potential. We find that GNNQQNY's assembly follows a complex nucleation process, as observed experimentally for numerous amyloid proteins. We also observe that the 20-mer and 50-mer systems form polymorphic structures that are the byproducts of different formation pathways. We further report the spontaneous formation of protofibrillar oligomers with structural characteristics typical of experimentally determined amyloid fibril structures.

Keywords: amyloid fibrils, prion, GNNQQNY, protein aggregation, nucleation, polymorphism, oligomers, protofibrils.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

Je dédicace cette thèse à toutes les femmes qui n’ont pas accès à l’éducation.

REMERCIEMENTS

Je tiens tout d'abord à remercier mon superviseur, Normand Mousseau, pour son soutien et pour la confiance qu'il m'a accordée durant ma première année de maîtrise et mes quatre années de doctorat. J'ai appris, grâce à lui, à mieux questionner, synthétiser et présenter, qualités qui sont nécessaires pour la carrière en recherche que je souhaite poursuivre. Je voudrais remercier Philippe Derreumaux de l'IBPC à Paris, de même que Giorgio Colombo et Massimiliano Meli de l'ICRM à Milan pour une collaboration très enrichissante. Je tiens aussi à remercier tout particulièrement Sébastien Côté pour les nombreuses discussions scientifiques que nous avons eues et pour ses suggestions pertinentes tout au long de mon doctorat. Merci à mes autres collègues de bureau, Jean-François Joly, Lilianne Dupuis, Jean-François Saint-Pierre, Rozita Laghaei, Peter Brommer, Cindie Eugène, Laurent-Karim Béland et Saïd Bouzakraoui, pour une ambiance de travail des plus agréables et pour des parties de Scrabble et des discussions de super-héros absolument inoubliables.

J'aimerais remercier les organismes qui ont généreusement financé mes travaux : le GÉPROM, le Conseil de Recherches en Sciences Naturelles et en Génie du Canada (CRSNG), la Fondation des Chaires de Recherche du Canada, les Fonds Québécois de Recherche en Santé et la Faculté des Études Supérieures et Postdoctorales (FESP).

Sur le plan plus personnel, j'aimerais remercier mon petit-ami Alexi Morin-Duchesne qui a su me soutenir contre vents et marées pendant quatre années et dont l'excellence et l'intelligence m'ont inspirée à donner le meilleur de moi-même. Je tiens aussi à remercier les quatre femmes fortes les plus importantes à mes yeux : ma mère Patricia Nasica de qui je puise ma force de caractère et mon inspiration et dont les encouragements et l'amour sont infinis ; et mes trois soeurs Alexie, Déborah et Jade pour tous ces fous rires passés comme pour ceux à venir. Un merci tout particulier à mon père Gérard Labouze, qui m'a plongée dans les mathématiques très jeune et qui m'a inspirée à faire des études en physique, qui m'ont amenée où je suis aujourd'hui. Il a toujours su accepter mes choix quels qu'ils soient et je l'en remercie sincèrement. Merci à mes grands-parents Nadia Bussi et Guy et Monette Nasica, qui sont toujours agréablement

curieux de connaître le détail de l'avancement de mes travaux de recherche et qui sont si fiers de leur petite fille. Un gros merci à la famille Morin-Duchesne, ma deuxième famille à Montréal qui m'a accueillie les bras ouverts depuis quatre ans.

Et bien sûr merci aux amis : Sébastien G. et Mary B., Mohamed T.-M., Olivier P.-L., Marie Pier F., Julien M., Louis-Nicolas F., Myriam D., Pascale S., Michel L., Lison M., Yashar H., Alastair F., Josh et Ashley L., Karina V., Alexandre L., Frederic V., Claire L. et Karim B.

Merci.

INTRODUCTION

Une protéine, chaîne polypeptidique composée d'acides aminés, doit se replier correctement afin d'adopter une forme spécifique déterminée qui lui confère sa fonction cellulaire normale. Si la forme adoptée par une protéine dévie de sa forme fonctionnelle, la protéine mal repliée, ou mépliée, fonctionne anormalement et peut alors avoir des effets toxiques sur la cellule. Une classe de protéines mépliées, les protéines amyloïdes, sont impliquées dans des maladies neurodégénératives comme les maladies d'Alzheimer, de Parkinson et les maladies à prions [1] et sont ainsi la cible de nombreuses études expérimentales et numériques en quête d'une solution thérapeutique. Sous certaines conditions physiologiques, les protéines amyloïdes forment, par un processus d'auto-assemblage, des macrostructures filamenteuses complexes, appelées fibres amyloïdes, très résistantes face aux protéases[1], dont la présence est associée à une détérioration de la communication interneuronale qui entraîne la mort des cellules neuronales. Ces fibres présentent une grande diversité morphologique, ou polymorphisme, issue de possibilités mécanistiques de formation diverses pour une même protéine [2–5], mais possèdent toutes un élément structurel commun, la structure cross-β[2] qui leur confère leur grande stabilité en agissant au niveau du maintient de la structure fibrillaire. Des études expérimentales ont aussi montré que toutes les protéines auraient la propriété intrinsèque de former des fibres amyloïdes [6–8], ce qui aggrave le problème de contrôle cellulaire de ces agrégats.

La cinétique d'agrégation des protéines amyloïdes en fibres, ou polymérisation, a été observée comme étant un phénomène complexe de croissance par nucléation décrit, dans certains cas, par la théorie classique physique de la nucléation. En effet, lorsque la concentration de protéines mépliées est suffisamment élevée, on observe un assemblage autocatalytique spontané des protéines qui dépend de la formation de noyaux métastables ayant surmonté une barrière d'énergie libre permettant par la suite une croissance irréversible en fibres polymériques. De nombreux modèles expérimentaux révèlent cependant que la théorie classique de la nucléation ne suffit pas à elle seule à reproduire le

[1] Les protéases font partie d'un mécanisme de défense et de contrôle de la cellule face aux structures anormales comme les fibres amyloïdes.

[2] Vue plus en détails dans le chapitre 1.

phénomène d'agrégation des protéines amyloïdes en fibres et d'autres facteurs importants doivent être pris en compte comme la présence d'agrégats de tailles intermédiaires lors de la formation de fibres ou comme la nature biochimique même des protéines en jeu tel leur degré d'hydrophobicité, leur charge, la présence de cycles aromatiques ou la présence de motifs alternant acides aminés hydrophobes et hydrophiles. Il se peut aussi que dans certains cas, un processus de nucléation secondaire ait lieu qui accelère la vitesse de polymérisation des fibres, compléxifiant ainsi le processus d'agrégation et facilitant la formation de fibres. Aussi, la propriété la plus extraordinaire des protéines amyloïdes, leur diversité de possibilités mécanistiques de formation pour une même protéine, est une cause supplémentaire de la compléxité de leur cinétique d'agrégation et de la coexistence de nombreux intermédiaires amyloïdes, sur le chemin de formation des fibres, comme les protofibres et les oligomères. Les protofibres sont de jeunes agrégats ordonnés, précédant la fibre, qui jouent un rôle mécanistique important dans la cinétique d'agrégation. Les oligomères, quant à eux, sont des petits agrégats désordonnés reconnus comme étant l'espèce responsable de la toxicité amyloïde [9, 10], c'est-à-dire de la mort cellulaire. Leur mécanisme de toxicité est cependant mal compris mais des propriétés communes à tous les oligomères ont été identifiées et laissent penser qu'ils opèrent selon un mécanisme universel, indépendamment de la séquence en acides aminés [9, 11].

Les grandes questions qui émergent sont donc: 1) Est-ce que les oligomères sont sur le chemin de formation des fibres ou sont-ils une espèce indépendante? 2) Comment opèrent-ils pour détruire la cellule? 3) Comment peut-on mieux caractériser leur structure afin de concevoir une stratégie de recherche d'inhibiteurs? 4) Que peut-on apprendre de leur processus d'agrégation? 5) Peut-on caractériser et reproduire la formation des protofibres? 6) Peut-on reproduire la formation de la structure cross-β, élément structurel caractéristique des fibres amyloïdes?

Dans cette thèse, nous tentons d'apporter une réponse aux questions 1), 3), 4), 5) et 6) par l'étude numérique du court fragment amyloïde GNNQQNY[3], issu d'une pro-

[3]Les lettres G, N, Q et Y étant les symboles des acides aminés Glycine, Asparagine, Glutamine et Tyrosine.

téine prion de levure, système très étudié expérimentalement depuis plus d'une vingtaine d'années de par sa facilité à former des fibres amyloïdes. Par le biais de l'étude de ce peptide, nous espérons améliorer notre compréhension globale du mécanisme de formation des fibres amyloïdes et sa courte longueur de sept acides aminés nous permet d'accéder numériquement à des tailles d'oligomères conséquemment plus grandes. Nos objectifs de recherche sont de caractériser les premières étapes d'agrégation de ce peptide amyloïde par la caractérisation des structures accessibles à des tailles de sytème allant du trimère au 50-mère (Chapitres 4 et 6) et par la caractérisation du processus cinétique d'assemblage (Chapitre 5) afin d'apporter de nouveaux éléments pouvant servir à mieux comprendre les protéines et peptides amyloïdes en général.

La méthode numérique que nous utilisons (détaillée dans le Chapitre 2) est la dynamique moléculaire couplée à un potentiel gros-grain, OPEP, et à des méthodes d'échantillonnage accéléré. Cette méthode nous permet de simuler, à une résolution quasi atomique, des systèmes de tailles considérables et pour des temps pertinents d'un point de vue expérimental, comparé à la plupart des études numériques disponibles dans la littérature. Il existe plusieurs études numériques des toutes premières étapes d'agrégation de GNNQQNY [12–23] qui ont permis une avancée significative dans notre compréhension de la dynamique d'agrégation de cette séquence et incluent des études d'agrégation de petits agrégats (allant du dimère au 6-mère) [14, 17–21, 23], des études d'agrégation pré-nuclée du 8-mère [17], des études de désagrégation du 8-mère et 12-mère [16] et des études de stabilité de la structure cross-β [12, 13, 15, 22]. La plupart de ces études s'accordent pour dire que l'agrégation de GNNQQNY est un processus coopératif aux allures de nucléation et à forte tendance à former des feuillets-β. Cependant, une caractérisation complète de la cinétique complexe d'agrégation à une échelle atomique ou quasi-atomique manque à la littérature pour des agrégats de taille conséquente. Cette thèse vient compléter ses études en offrant la possibilité d'explorer les structures et la cinétique d'agrégats de GNNQQNY de plus grandes tailles, suffisamment grands pour permettre l'apparition de la formation de protofibres.

CHAPITRE 1

PROTÉINES ET PEPTIDES AMYLOÏDES

Ce chapitre fait état des connaissances générales acquises à ce jour sur les protéines amyloïdes: leur importance dans un contexte médical, leurs caractéristiques structurelles typiques, leur mécanisme de formation ainsi que leur mode de fonctionnement. Cette description sera suivie d'une justification du choix de système à l'étude, soit le peptide amyloïde GNNQQNY issu de la protéine prion Sup35.

1.1 Protéines amyloïdes

Les protéines amyloïdes sont impliquées dans de nombreuses maladies neurodégénératives. Parmi ces pathologies dévastatrices, on retrouve les maladies d'Alzheimer (causée par la protéine Amyloïde-β ou Aβ), de Parkinson (α-synucléine), de Creutzfeldt-Jakob (Prion), de Huntington (Huntingtine) et le diabète de type II (amyline), pour ne nommer que les plus connues. Dans ces cas, sous certaines conditions physiologiques, les protéines amyloïdes ont une forte propension à l'agrégation et se déposent intracellulairement pour former des inclusions ou extracellulairement pour former des plaques ou fibres [1]. De plus, des études expérimentales ont montré que toutes les protéines possèdent une prédisposition à former des agrégats amyloïdes qui partagent des caractéristiques structurelles communes et donc, par extension, des mécanismes communs d'agrégation [6–8].

1.1.1 Contexte biologique: Le mépliement des protéines amyloïdes

L'élément déclencheur pour l'agrégation est le mauvais repliement, ou mépliement, des protéines amyloïdes soumises à des conditions environnementales inhabituelles telles qu'une augmentation de la température, un changement de pH, un taux de glucose élevé ou la présence d'agents oxidatifs [24]. Malgré les nombreux mécanismes cellulaires de dégradation de protéines mépliées, un changement de conditions environnementales

peut générer de nombreux changements conformationnels, de protéines autrement fonctionnelles, qui entraînent alors une réaction en chaîne de production massive de ces protéines mépliées nocives qui saturent les systèmes naturels de contrôle de qualité des protéines dans la cellule. En effet, si la protéine sous sa forme native et soluble se replie en prenant soin de ne pas exposer ses régions hydrophobes au solvant, la forme mépliée, au contraire, les expose dans de nombreux cas et devient ainsi insoluble [25]. L'agrégation amyloïde – en d'autres termes, l'attraction mutuelle entre protéines mépliées – résulte très souvent de cette exposition des résidus hydrophobes chez ces protéines car leur forme mépliée est thermodynamiquement instable et non-favorable. Ainsi, l'agrégation est une façon de procurer une plus grande stabilité et une plus basse énergie aux protéines mépliées [24]. Ces changements environnementaux peuvent survenir post-translationnellement – dans le cas où une protéine perd sa structure native et adopte une forme mépliée – mais aussi pré-translationnellement – dans le cas où des modifications génétiques sont à l'origine d'une mutation entraînant le mépliement [26].

1.1.2 Propriétés structurelles

Durant l'agrégation, les protéines mépliées sont ainsi converties en agrégats fibrillaires supramoléculaires insolubles et particulièrement stables dont la structure est hautement organisée: les fibres amyloïdes, et dont la morphologie globale semble avoir des traits communs d'une protéine à l'autre. Les fibres amyloïdes sont des structures fibrillaires (filamenteuses) torsadées de 60 à 120 Å de diamètre [27] et de 0.1 à 10 μm de longueur composées de protofilaments enroulés, telles que déterminées par microscopie electronique [28] et par microscopie à force atomique [29] (Fig. 1.1) , dont la structure centrale consiste en une structure "*cross-*β", riche en feuillets-β. La structure cross-β, déterminée par diffractrométrie de rayons X [30, 31] et par crystallographie à rayons X [32], est composée d'une (ou plusieurs) paire de feuillets-β face à face qui intéragissent au moyen de l'interdigitation de leur chaînes latérales, dont l'axe de symétrie est perpendiculaire à l'axe de la fibre et dont chaque brin de chaque feuillet-β appartient à une copie différente d'une même protéine. Une description plus détaillée de la structure cross-β sera apportée dans la section 1.2.1.1.

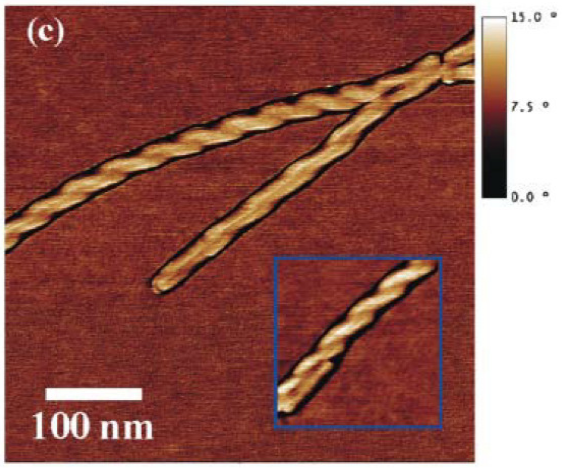

Figure 1.1: **Morphologie torsadée filamenteuse d'une fibre amyloïde d'insuline obtenue par microscopie à force atomique.** Cette figure provient de Jansen *et al.* (2005).

Les fibres amyloïdes ont des propriétés physico-chimiques spécifiques lorsqu'elles sont déposées dans les tissus, entre autre une biréfringence vert-pomme à la lumière polarisée lorsque liées au colorant rouge Congo ou à des dérivés de Thioflavine [25], ce qui permet une détection *ex-vivo* d'agrégats fibrillaires amyloïdes. Cependant seules les espèces amyloïdes ordonnées comme les fibres ou protofibres[1], c'est-à-dire riches en longs feuillets-β, se lient à ces colorants et sont donc détectables par ces moyens, tandis que les espèces plus jeunes et préliminaires, donc plus petites et plus désordonnées (et plus solubles), les oligomères, ne le sont pas [33] et ne sont observables qu'*in vitro* au moyen de techniques d'imagerie telles que la microscopie électronique (Fig. 1.2). La figure 1.2 compare l'apparence des fibres, protofibres et oligomères par microscopie électronique.

Bien qu'une structure commune ait été identifiée pour les fibres et protofibres amyloïdes (la structure cross-β), il existe peu de données appuyant une possibilité d'élément structurel commun pour les oligomères, certains étant riches en feuillets-β [34–38] et d'autres étant principalement désordonnés [39]. La communauté scientifique réfère ainsi

[1]Les protofibres sont une espèce ayant tous les traits caractéristiques d'une fibre mais n'ayant pas encore assez mûri pour en être une. En d'autres mots, une protofibre est une fibre en devenir.

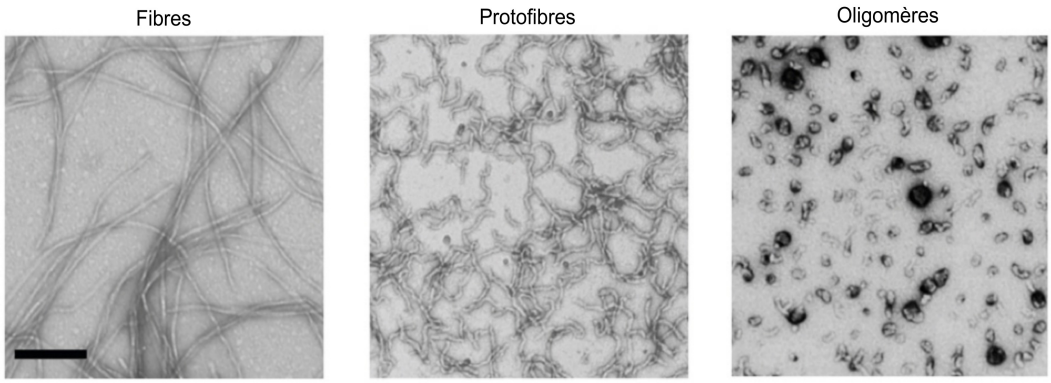

Figure 1.2: **Structure de Aβ par microscopie electronique.** Gauche: Fibres de Aβ. Milieu: Protofibres de Aβ. Droite: Oligomères de Aβ. Les oligomères apparaissent comme des amas plus petits et plus sphériques que les fibres et protofibres, dont l'apparence est similaire. Cette figure provient de Fändrich *et al.* (2012).

aux oligomères par le terme "*agrégats amorphes*" de façon plus générale. Malgré leur manque de conformité structurelle, les oligomères ont cependant des propriétés communes universelles très bien reconnues. En effet, il a été révélé expérimentalement que des oligomères provenant de protéines différentes partageaient une abilité commune à se lier à un anticorps spécifique [11]. De plus, les oligomères semblent avoir des effets similaires sur le métabolisme et la viabilité des cellules indépendamment de la séquence de leurs constituants [9]. Ces deux observations combinées laissent fortement penser qu'un mécanisme de formation commun régit l'agrégation en oligomères, et par extension le processus menant aux fibres.

1.1.3 Formation de fibres amyloïdes

Le mécanisme général de formation de fibres peut se voir comme un processus d'assemblage spontané et coopératif en trois étapes. La première étape, la plus rapide, est l'effondrement hydrophobe des monomères protéiques puis leur alignement pour former des feuillets-β, au moyen de liaisons hydrogène. Ici, les feuillets-β se forment rapidement et de façon réversible. La deuxième étape est la formation de la structure cross-β qui consiste en l'autocomplémentarité (i.e. l'interdigitation) de deux feuillets-β. Cette

phase nécessite des interactions de van der Waals et est plus lente que l'étape précédente car les chaînes latérales de deux feuillets-β doivent prendre le temps de trouver la meilleure orientation pour permettre leur interdigitation. Nelson *et al.* suggèrent qu'à ce stade la diminution entropique, due à une perte de liberté des monomères et à un gain dans le niveau d'ordre global de la structure, crée une barrière énergétique pour la formation de fibres [32]. L'organisation structurelle s'accompagne de la formation de liaisons hydrogène qui diminuent l'énergie globale du système après formation d'un noyau métastable issu d'un phénomène de nucléation (voir Annexe 1). Enfin, la troisième étape est l'interaction des paires de feuillets-β pour former des fibres par le biais de liaisons non covalentes. La formation de fibres contribue à une très grande barrière énergétique qui rend presqu'impossible la dissolution de fibres et donc qui fait de la fibrillisation un processus irréversible dans la plupart des cas [32].

Il a été montré expérimentalement cependant que, pour une protéine donnée, le mécanisme de formation d'agrégats amyloïdes n'emprunte pas un chemin unique et bien déterminé mais démontre plutôt une grande variété de possibilités mécanistiques amenant à la formation de fibres [2–5]. Ces chemins de formation possibles peuvent ainsi impliquer des espèces intermédiaires variées d'un chemin à l'autre, ce qui confère un certain degré de polymorphisme à la structure finale de la fibre, caractéristique bien connue des protéines amyloïdes (voir section I.5.2). Ainsi, une multitude d'espèces intermédiaires plus ou moins ordonnées structurellement coexistent avec les fibres, dont les oligomères. Les oligomères amyloïdes sont une espèce jeune qui apparait souvent tôt sur le chemin de formation des fibres et qui joue un rôle crucial, mais non moins mal compris, non seulement dans la morphologie finale des fibres mais aussi dans la toxicité amyloïde.

1.1.4 Mécanisme de toxicité

Pendant longtemps, les fibres amyloïdes furent ciblées comme étant l'espèce toxique responsable de la dégénérescence neuronale. Cependant, depuis plus d'une dizaine d'années, les oligomères sont de plus en plus mis en avant en tant qu'espèce pathogénique [9] alors que les fibres serviraient plutôt d'espèces détoxifiantes en séquestrant une par-

tie de ces oligomères toxiques et solubles en fibres insolubles et physiologiquement inactives [40, 41]. Bien que le mécanisme de toxicité des protéines amyloïdes soit peu caractérisé et généralement encore mal compris, certaines études semblent indiquer que le mécanisme de toxicité des oligomères opérerait par le biais d'une perturbation de la couche lipidique de la membrane cellulaire, bouleversant ainsi l'équilibre ionique entre l'espace intra- et extra-cellulaire, entraînant la mort de la cellule [10]. Ces observations soulèvent l'importance du rôle des oligomères d'un point de vue pathologique. Les oligomères sont donc de très bonnes cibles pour la recherche d'inhibiteurs capables d'annihiler leurs effets toxiques.

1.1.5 Pertinence de l'étude des oligomères

Si les oligomères sont de très bons candidats pour des études expérimentales visant à comprendre et tenter de reduire la toxicité causée par les protéines amyloïdes *in vivo*, ils sont aussi d'excellents candidats pour des études numériques pour plusieurs raisons:

1. La taille des oligomères de protéines amyloïdes est accessible à des simulations numériques de résolution presque atomique allant du dimère au tetramère.

2. Étant donné que les oligomères possèdent des propriétés universelles, indépendantes de la séquence en acides aminés de leurs constituants, l'éventuel biais du potentiel choisi ne devrait pas transparaitre dans les simulations, augmentant ainsi la fiabilité des prédictions numériques.

3. La découverte de courtes séquences amyloïdes responsables de la formation d'agrégats amyloïdes permet, en isolant ces séquences, d'accéder numériquement à des systèmes de tailles considérables allant jusqu'au 50-mère à résolution quasi atomique.

1.2 L'approche des petits peptides amyloïdes

Il existe une panoplie de modèles qui tentent d'expliquer comment les protéines se convertissent en fibres [42, 43]. Cependant, tous ces modèles s'accordent pour dire

que lors de cette conversion, une partie de la protéine, normalement inaccessible par le solvant, est alors exposée et devient le catalyseur de la formation de fibres. Ce segment protéique anormalement exposé est ainsi responsable de la formation de fibres de protéines mépliées et est souvent appelé "*peptide amyloïdogénique*" lorsqu'isolé de la protéine.

Une méthode numérique de profilage 3D de courtes séquences pour la prédiction de formation de fibres a permis d'identifier de tels segments dans plusieurs protéines amyloïdes telles que la protéine τ, α-synucléine, β2-microglobuline, l'amyline, l'insuline ou encore Aβ [44] et leurs résultats confirment plusieurs expériences. Par la suite, Teng *et al.* [45] ont choisi d'insérer dix de ces fragments amyloïdes dans l'enzyme RNase A, reconnue pour être particulièrement stable. La protéine RNase A ne forme pas de fibres amyloïdes par elle-même mais lorsqu'elle contient un de ces fragments amyloïdes, elle est entraînée à former des fibres qui possèdent une biréfringence vert-pomme caractéristique lorsque liées au colorant rouge Congo. Il est ainsi conclu que ces peptides amyloïdes suffisent à entraîner l'agrégation et ainsi la formation de fibres.

1.2.1 Le peptide GNNQQNY issu de la protéine prion Sup35

La protéine prion Sup35 et, plus particulièrement, une courte portion de cette protéine ont suscité beaucoup d'intérêts dans le domaine des protéines amyloïdes depuis plus d'une vingtaine d'années. La protéine amyloïde Sup35 provient de la levure bourgeonnante *Saccharomyces cerevisiae* et est impliquée, dans sa forme native, dans le complexe facteur de terminaison de traduction [2], alors que dans sa forme amyloïde, elle forme des fibres *in vitro*. Sup35 est composée de 685 acides aminés dont les 123 premiers (région N-terminale) constituent le domaine prion, qui possède des similarités de séquence avec la protéine prion humaine PrP. Parmi la région N-terminale de Sup35, une courte séquence amyloïdogénique de 7 acides aminés de long, GNNQQNY, a été identifiée par Balbirnie *et al.* [27]. Ce peptide manifeste toutes les propriétés amyloïdes de la protéine complète Sup35 c'est-à-dire une cinétique d'agrégation coopérative, la

[2] Ce complexe est responsable de libérer une chaîne polypeptidique lors de la lecture d'un codon stop par le ribosome.

formation de fibres, une liaison au colorant rouge Congo ainsi qu'un motif de diffraction par rayons X caractéristique de la présence d'une structure cross-β. Il est donc considéré comme étant l'élément catalyseur d'agrégation pour la protéine Sup35.

1.2.1.1 Résolution de la structure cross-β pour le peptide GNNQQNY

Le peptide GNNQQNY est un peptide polaire, peu hydrophobe et riche en glutamines et asparagines, qui semblent être à l'origine de sa propension à l'agrégation [46]. Nelson *et al.* ont déterminé la structure atomique détaillée cross-β du peptide GNNQQNY par crystallographie à rayons X [32]. Leur étude rŐvŔle que le peptide est soluble[3], et qu'il forme, en quelques heures, des fibres dont la structure principale est la structure cross-β. Cette structure cross-β est composée de deux feuillets-β empilés, perpendiculairement à l'axe de la fibre, qui interagissent au moyen de l'interdigitation de leurs chaînes latérales (Fig. 1.3 (b), (c) et (d)). Les brins des feuillets-β sont entièrement parallèles et alignés (Fig. 1.3 (a)), les brins d'un feuillet étant antiparallèles aux brins de l'autre feuillet. Les paires de feuillets interagissent ensuite ensemble et Nelson *et al.* notent la présence de deux types d'interfaces: une interface "*sèche*" et une interface "*hydratée*" (Fig. 1.3 (c) et (d)). L'interface sèche est la zone d'interdigitation de chaînes latérales, ne contient presqu'aucune molécule d'eau et a une épaisseur de 8.5 Å , alors que l'interface hydratée, remplie de molécules d'eau est plus large et mesure 15 Å d'épaisseur (Fig. 1.3 (c)).

L'interdigitation et l'autocomplémentarité des chaînes latérales à l'interface sèche est ce qui donne sa grande stabilité à la structure cross-β pour le peptide GNNQQNY, mais aussi pour d'autres courtes séquences amyloïdogéniques qui présentent des structures cross-β similaires [47].

[3] ce qui permet l'obtention de microcrystaux

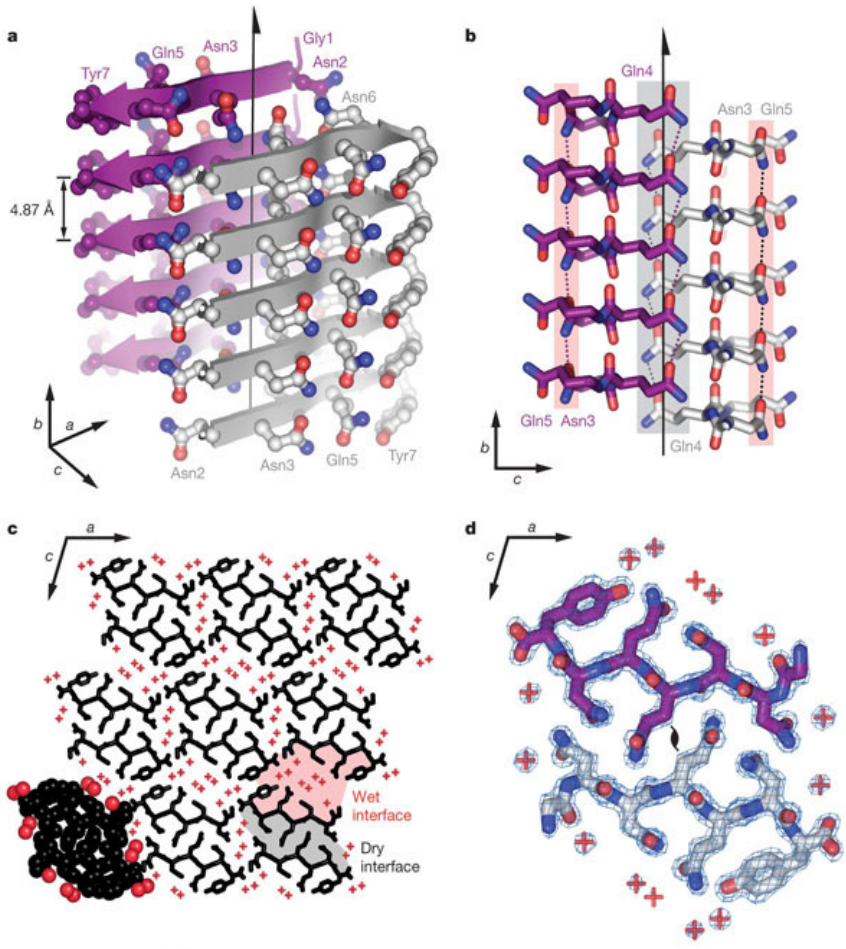

Figure 1.3: **Structure cross-β de GNNQQNY** (a) Structure des paires de feuillets montrant les brins parallèles dans chaque feuillet. Les feuillets sont antiparallèles les uns par rapport aux autres. Les chaînes latérales sont montrées en bâtons et sphères. (b) Interdigitation des chaînes latérales vue de côté (le long de l'axe a). (c) Vue de dessus du crystal de GNNQQNY le long de l'axe b montrant les interfaces sèche et hydratée. Il y a six rangées de feuillets-β dans la direction verticale. Les peptides sont montrés en noir et les molécules d'eau en rouge (signes "+"). Les atomes montrés en sphères en bas à gauche représentent les rayons de van der Waals. (d) Zoom sur l'interdigitation des chaînes latérales entre deux peptides GNNQQNY selon le même axe que la figure (c). Cette figure, ainsi qu'une partie de sa légende, provient de Nelson *et al.* (2005).

1.2.1.2 Travaux publiés sur les premières étapes d'agrégation du peptide GNNQQNY

En complément des nombreux travaux expérimentaux de détermination de structures pour le peptide GNNQQNY [32, 48–51] plusieurs études numériques ont été menées pour tenter de comprendre les premières étapes d'agrégation du peptide GNNQQNY à des échelles allant du dimère au 8-mère, à partir de structures désordonnées ou bien préformées [12–23]. Bien que ces travaux soient un pas en avant dans la compréhension de la dynamique des touts premiers agrégats, la taille de systèmes étudiés n'est pas suffisamment large pour apporter une information pertinente quant aux traits sous-jacents des morphologies oligomériques et de la cinétique d'agrégation du peptide GNNQQNY.

1.2.1.3 Objectifs de recherche

Par le biais d'une méthode numérique efficace d'échantillonnage en dynamique moléculaire couplée à un potentiel gros-grain (Chapitre 2), nous caractérisons les structures accessibles à des systèmes allant du trimère au 50-mère GNNQQNY en passant par le dodécamère et le 20-mère (Chapitres 4 et 6). En complément à ces études, nous apportons une revue exhaustive des connaissances en matière de phénomènes d'autoassemblage des protéines amyloïdes qui peuvent être vus comme des mécanismes complexes de nucléation (Chapitre 3 et Annexe 1), ce qui nous a permis d'apporter une description détaillée de la cinétique complexe d'agrégation du 20-mère GNNQQNY (Chapitre 5), et ce qui nous permettra de décrire la cinétique du 50-mère en dehors du cadre de cette thèse et en complément du chapitre 6.

CHAPITRE 2

MÉTHODOLOGIE

Face au constat des limites de la biologie expérimentale pour décrire des phénomènes tels que le repliement de protéines et les processus d'agrégation à haute résolution, il est devenu nécessaire d'intégrer d'autres disciplines, telles que la physique et la chimie computationnelles, à la biologie. C'est ainsi que naît la biophysique théorique, domaine dans lequel la modélisation biomoléculaire a fait ses preuves et a apporté de nombreuses contributions à l'amélioration de notre compréhension des systèmes biologiques complexes. Malgré des défis d'efficacité computationnelle, la modélisation biomoléculaire voit de nombreuses avancées, tant au niveau algorithmique que technologique, avec la montée des superordinateurs, devenus plus performants et plus accessibles pour la recherche académique et industrielle. Le développement et l'utilisation de méthodes numériques innovatrices pour complémenter les travaux expérimentaux sont ainsi devenus cruciaux à l'avancement de la recherche notamment dans le domaine du repliement des protéines, de leur dynamique et de l'étude de leur processus d'agrégation.

Ce chapitre décrit les outils numériques utilisés, dans nos travaux, pour étudier les premières étapes d'agrégation du peptide GNNQQNY, dont le potentiel gros-grain OPEP couplé à des dynamiques moléculaires et à des techniques d'échantillonnage accéléré comme l'échange de répliques. Les avantages et faiblesses des outils utilisés sont soulignés et, lorsque possible, une suggestion est apportée pour résoudre certains problèmes.

2.1 Le potentiel gros-grain OPEP - Optimized Potential for Efficient protein Prediction

Quel que soit le système à l'étude, la description des premières étapes d'agrégation à l'échelle des oligomères requiert une taille conséquente en terme du nombre de monomères à simuler. Ainsi, les potentiels gros-grain, c'est-à-dire dont la représentation atom-

ique est simplifiée, permettent d'augmenter la taille des systèmes accessibles aux simulations au dépend d'une perte de la résolution, et sont donc des méthodes idéales pour étudier les phénomènes d'agrégation. Pour l'étude du peptide GNNQQNY, nous avons choisi le potentiel gros-grain OPEP, développé par Philippe Derreumaux de l'IBPC-Paris [52] et utilisé depuis plus d'une dizaine d'années, couplé à une palette d'outils numériques tels que la méthode Monte-Carlo [53–57], la dynamique moléculaire [58–61], l'échange de répliques en dynamique moléculaire [62–66] et ART-nouveau [67–72].

2.1.1 Description du potentiel

En représentation OPEP, une chaîne polypeptidique est composée, pour la chaîne principale, de tous ses atomes détaillés (N, H, C_α, C et O) et d'une seule particule, ou bille, pour représenter la chaîne latérale (sauf pour les prolines) [52]. Les paramètres géométriques et rayons de van der Waals de cette bille, définis à partir d'un échantillon représentatif de structures diverses [73], sont différents pour chaque acide aminé (Fig. 2.1). Cette description des acides aminés permet une accélération du calcul de forces entre les particules, tout en gardant un certain degré de propriétés géométriques propre à chaque type de chaîne latérale. De plus, le potentiel OPEP considère les effets implicites du solvant – ici l'eau – dans sa fonction d'énergie.

2.1.1.1 Fonction d'énergie

La fonction d'énergie du potentiel OPEP inclut des termes d'interactions de courte et de longue portée qui peuvent être divisés en trois catégories: les interactions locales, les interactions non liantes et les liaisons hydrogène. L'énergie potentielle totale est ainsi:

$$E = E_{local} + E_{non\_liant} + E_{HB} \tag{2.1}$$

où HB signifie liaisons hydrogène.

Le terme d'interactions locales E_{local} traduit les forces associées aux déformations de longueurs de liaisons covalentes, d'angles de liaisons, mais aussi de torsions impropres (hors-plan) des chaînes latérales et liaisons peptidiques, et est exprimé par des potentiels

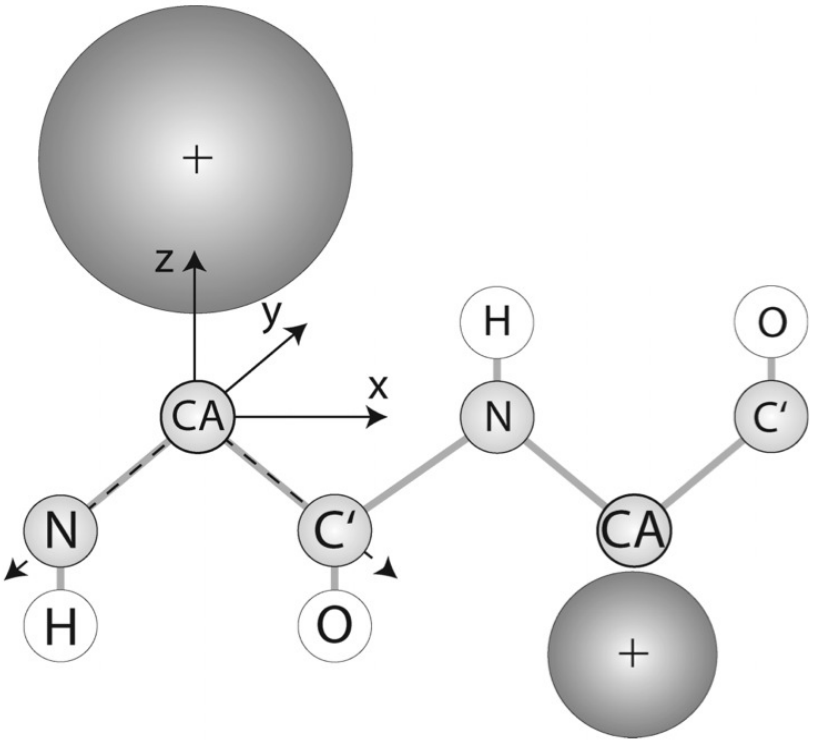

Figure 2.1: **Représentation gros-grain du potentiel OPEP.** Chaque chaîne latérale est representée par une bille (en gris foncé) dont la distance du centroïde par rapport aux atomes lourds (N, C_α et C) et le rayon de van der Waals sont différents pour chaque acide aminé. Tous les acides aminés sont représentés ainsi avec OPEP sauf les prolines qui sont représentées tout-atome. Ici, C_A signifie C_α. Cette figure est tirée de Mautpetit *et al.* (2007).

harmoniques qui assurent un maintien de la géométrie peptidique (approximation valide pour de petites déviations) et définis par:

$$E_{local} = w_b \sum_{liaisons} K_b(r - r_{eq})^2 + w_a \sum_{angles} K_\alpha(\alpha - \alpha_{eq})^2 + w_\Omega \sum_{tors\_imp} k_\Omega(\Omega - \Omega_{eq})^2 + w_{\phi,\psi}\left(\sum_\phi E_\phi + \sum_\psi E_\psi\right) \tag{2.2}$$

où l'indice "eq" indique les valeurs d'équilibre pour les liaisons, les angles et les torsions. K_b, K_α et k_Ω sont les constantes de forces (en kcal/mol/Å^2) qui donnent la rigidité des liaisons, angles et torsions et r, α et Ω sont les valeurs de longueur de liaisons, d'angles et de torsions, éloignées de l'équilibre. Les valeurs d'équilibre et les constantes de forces associées avec les atomes de la chaîne principale sont tirées du potentiel AMBER [74] et celles associées avec les chaînes latérales sont très similaires à celles d'AMBER. w_b, w_a et w_Ω sont des poids attribués aux différentes composantes énergétiques pour assurer un bon équilibre entre les forces de courte portée et de longue portée [54]. De plus, les potentiels E_ϕ et E_ψ de l'équation (2.2) assurent que les angles dihédraux des structures OPEP occupent de préférence (sans restrictions exclusives) les régions du graphe de Ramachandran normalement accessibles par les structures tout-atome [75] et suivent les potentiels harmoniques suivants:

$$E_\phi = k_{\phi\psi}(\phi - \phi_0)^2 \tag{2.3}$$

$$E_\psi = k_{\phi\psi}(\psi - \psi_0)^2 \tag{2.4}$$

où $k_{\phi\psi}$ est une constante de force. $\phi_0 = \phi$ si ϕ appartient à l'interval $[\phi_{inf} : \phi_{sup}]$ et $\phi_0 = \min(\phi - \phi_{inf}, \phi - \phi_{sup})$ autrement, avec $\phi_{inf} = -160°$ et $\phi_{sup} = -60°$. Les mêmes critères sont appliqués pour les angles ψ avec $\psi_{inf} = -60°$ et $\psi_{sup} = 160°$.

Le terme d'interactions non liantes, E_{non_liant}, est quant à lui donné par:

$$\begin{aligned} E_{non\_liant} =& w_{1,4} \sum_{1,4} E_{vdW} + w_{C_\alpha,C_\alpha} \sum_{C_\alpha,C_\alpha} E_{vdW} + \\ & w_{1>4} \sum_{P',P'} E_{vdW} + w_{1>4} \sum_{P',C_\alpha} E_{vdW} + \\ & w_{1>4} \sum_{P',cl} E_{vdW} + \sum_{cl,cl} w_{cl,cl} E_{vdW} \end{aligned} \tag{2.5}$$

où "1,4" désigne les interactions 1-4 (impliquant les atomes aux extrémités d'un angle dihédral), "P'" représente les atomes N, C, O et H de la chaîne principale, "*cl*" représente les chaînes latérales. Les interactions de courte portée (1-4) sont séparées des interactions de longue portée ($1 > 4$) et les poids associés aux différentes composantes sont calculés afin d'équilibrer ces deux types d'interactions.

E_{vdW}, le potentiel van der Waals "par paires" , dans l'équation (2.5) est donné par:

$$E_{vdW} = \varepsilon_{ij} \left(\left(\frac{r^0_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r^0_{ij}}{r_{ij}} \right)^{6} \right) H(\varepsilon_{ij}) - \varepsilon_{ij} \left(\frac{r^0_{ij}}{r_{ij}} \right)^{6} H(-\varepsilon_{ij}) \tag{2.6}$$

avec $H(x) = 1$ si $x \geq 0$ et $H(x) = 0$ si $x < 0$ (fonction échelon), r_{ij} la distance entre les particules i et j, $r^0_{ij} = (r^0_i + r^0_j)/2$ où r^0_i et r^0_j sont les rayons de van der Waals des particules i et j. Pour les particules qui ne sont pas des chaînes latérales, le potentiel de van der Waals est une fonction de Lennard-Jones de type 12-6 dont le premier terme est répulsif et le deuxième terme est attractif. Pour les chaînes latérales, deux cas sont considérés: 1) si les interactions *cl–cl* sont de nature hydrophobe ou si les deux chaînes latérales sont de charges contraires, le même potentiel 12-6 de Lennard-Jones est utilisé alors que 2) dans tous les autres cas (interactions *cl–cl* hydrophiles ou de charges électriques de même signe), le potentiel répulsif (troisième terme de l'équation (2.6)) est utilisé. Comme les forces de van der Waals sont des forces de courte portée, on applique une coupure progressive entre 15 et 16 Å (tel que $F = 0$ à $r = 16$ Å), distance au delà de laquelle elles ne sont pas calculées, pour alléger le coût des calculs. Les coefficients ε_{ij} sont calibrés et dépendent du type d'interaction entre deux particules données.

Finalement, le potentiel pour les liaisons hydrogène est composé de deux termes: un terme d'interaction à deux corps (E_{HB1}) et un terme d'interaction à quatre corps (E_{HB2}), qui tient compte de la coopérativité entre liaisons hydrogène. Le terme à deux corps est défini par:

$$E_{HB1} = w_{hb1-4} \sum_{ij,j=i+4} \varepsilon_{hb1-4} \mu(r_{ij}) \nu(\alpha_{ij}) + w_{hb1>4} \sum_{ij,j>i+4} \varepsilon_{hb1>4} \mu(r_{ij}) \nu(\alpha_{ij}) \quad (2.7)$$

où

$$\mu(r_{ij}) = 5 \left(\frac{\sigma}{r_{ij}} \right)^{12} - 6 \left(\frac{\sigma}{r_{ij}} \right)^{10} \quad (2.8)$$

et

$$\nu(\alpha_{ij}) = \begin{cases} cos^2 \alpha_{ij}, & \alpha_{ij} > 90^\circ, \\ 0, & \text{autrement}, \end{cases} \quad (2.9)$$

avec r_{ij} la distance O–H entre l'oxygène carboxyle d'une particule et l'hydrogène du groupe amine de l'autre particule, α_{ij} l'angle $\widehat{NHO}$ et σ la valeur d'équilibre de la distance O–H (1.8 Å). Les énergies de liaisons hydrogène de courte (hélices-α) et de longue (feuillets-β) portées sont distinguées et ε_{hb1-4} et $\varepsilon_{hb1>4}$ sont des paramètres qui dépendent de ces deux types d'interactions, respectivement.

À présent, le terme d'interactions à quatre corps, qui décrit la coopérativité entre les liaisons hydrogène ij et kl s'écrit:

$$\begin{aligned} E_{HB2} = & \sum \varepsilon_{\alpha}^{\text{coop}} \exp(-(r_{ij} - \sigma)^2/2) \exp(-(r_{kl} - \sigma)^2/2) \times \Delta(ijkl) \\ & + \sum \varepsilon_{\beta}^{\text{coop}} \exp(-(r_{ij} - \sigma)^2/2) \exp(-(r_{kl} - \sigma)^2/2) \times \Delta'(ijkl) \end{aligned} \quad (2.10)$$

où

$$\begin{aligned}
\Delta(ijkl) &= \begin{cases} 1, & \text{si } (k,l) = (i+1, j+1) \text{ et } j = i+4, l = k+4, \\ 0, & \text{autrement,} \end{cases} \\
\Delta'(ijkl) &= \begin{cases} 1, & \text{si } (k,l) = (i+2, j-2), \\ 1, & \text{si } (k,l) = (i+2, j+2), \\ 0, & \text{autrement.} \end{cases}
\end{aligned} \tag{2.11}$$

Les paramètres $\Delta(ijkl)$ et $\Delta'(ijkl)$ entrent en jeu dans la stabilisation des hélices-α et feuillets-β (aussi bien parallèles qu'anti-parallèles) et des boucles et segments désordonnés. Les paramètres $\varepsilon_\alpha^{\text{coop}}$ et $\varepsilon_\beta^{\text{coop}}$, quant à eux, prennent en compte la propension des acides aminés à former des feuillets-β ou des hélices-α.

2.1.1.2 Optimisation des paramètres du potentiel

Afin de maximiser la performance du potentiel OPEP en terme de prédiction de structures – c'est-à-dire afin qu'il puisse discriminer les structures natives des non natives – 261 paramètres libres (*i.e.* les poids dans la fonction d'énergie) doivent être calibrés à l'aide d'un ensemble de structures connues. Cette calibration des paramètres, à l'aide d'un algorithme génétique, satisfait plusieurs conditions sur la valeur d'énergie d'une structure native (N) comparée aux énergies de structures quasi natives (QN) ou de structures leurres (L) tel que:

$$\begin{cases} E(L) - E(N) > 0 \\ E(QN) - E(N) > 0 \\ E(L) - E(QN) > 0 \end{cases} \tag{2.12}$$

La fonction de coût pour chaque protéine est égale à -1 si ces conditions sont satisfaites, et est égale à 0 autrement. L'ensemble des structures choisies pour évaluer la fonction de coût et calibrer les paramètres du potentiel est composé d'une dizaine de protéines qui ont des structures secondaires variées (certaines étant seulement α, d'autres seulement β et d'autres étant un mélange α/β) et stables en solution sans contenir d'élements

stabilisants comme les liaisons disulfure ou les acides aminés non-naturels [76] pour un total de 7626 structures[1]. De plus, les protéines sélectionnées ne dépassent pas 16 à 25 acides aminés de long pour permettre un bon échantillonnage des conformations.
Un ensemble de 16 autres protéines (pour un total de 20926 structures leurres) a ensuite été utilisé pour valider la paramétrisation de OPEP. Pour $\sim$80 % des protéines, OPEP est capable d'identifier les structures natives ou quasi natives de façon discriminatoire et sa performance est comparable ou supérieure à des programmes comme DOPE [77, 78], ROSETTA [79–81], DFIRE [82] et ModPipe [83].

2.1.1.3 Applications

OPEP est donc un potentiel fiable pour la prédiction de structures natives ou quasi natives, mais est aussi très adéquat pour décrire la cinétique et la thermodynamique des protéines. En effet, plusieurs études couplant le potentiel OPEP à la méthode d'échantillonnage ART-nouveau[2] ont reproduit des chemins de repliement en accord avec plusieurs autres études numériques [69, 84, 85] et de nombreuses études de dynamique moléculaire avec échange de répliques avec OPEP ont réussi à caractériser correctement la thermodynamique de structures comme la structure β en épingle à cheveux de la protéine G [62] et Aβ_{16-22} [86].
Depuis, le potentiel a aussi été couplé à des dynamiques moléculaires HREMD[3] pour permettre l'étude des premières étapes d'agrégation de protéines entières comme Aβ, amyline et polyglutamine [63, 65, 66, 87, 88] et à une méthode holographique de mouvements à grande échelle [72, 89].

2.1.2 Avantages et limites du potentiel

La simplification de la représentation des chaînes latérales et le choix d'un solvant implicite avec OPEP permettent de diminuer le coût numérique des calculs de forces,

[1] Structures leurres mépliées générées soit par dynamique moléculaire, par une méthode glouton (greedy algorithm), ou disponibles sur une base de données publique.
[2] Technique d'activation-relaxation.
[3] Hamiltonian Replica Exchange Molecular Dynamics.

opération la plus coûteuse d'une dynamique moléculaire. Dans une simulation de dynamique moléculaire avec OPEP, on peut ainsi accéder à des systèmes à plus grande échelle sur des temps plus longs, permettant d'atteindre des tailles et temps de plus en plus pertinents d'un point de vue expérimental, en particulier pour l'étude des oligomères amyloïdes. Un potentiel gros-grain à la résolution d'OPEP permet aussi de qualifier les tendances intrinsèques générales d'un système et est particulièrement adapté pour décrire le caractère universel de ces oligomères amyloïdes, qui ont des propriétés qui ne dépendent pas de la séquence, et l'effet potentiellement biaisé d'un potentiel se ressentira ainsi moins sur de tels systèmes.

D'un autre côté, le détail des chaînes latérales peut s'avérer crucial dans certains mécanismes protéiques; dans le cas d'intérêt ici, la formation de la structure cross-β. Il manque alors à notre description la possibilité de former un motif d'interdigitation des chaînes latérales qui est responsable de la stabilisation de la structure cross-β des agrégats amyloïdes. Une partie de la solution à ce problème serait l'utilisation d'un potentiel hybride, inspiré d'OPEP avec certaines chaînes latérales représentées tout-atome, développé récemment par Lilianne Dupuis, couplé à la méthode ART-nouveau et à une méthode holographique multi-échelle [72, 89] dont les résultats préliminaires sont très prometteurs.

La description implicite du solvant réduit de façon significative la taille du système et permet aussi d'accélérer considérablement l'échantillonnage des conformations des protéines de par l'absence d'effets de frottements avec le solvant. Cet absence de viscosité peut cependant affecter la dynamique de certaines protéines et, dans le meilleur des cas, un traitement des simulations avec une dynamique de Langevin, tenant compte de la viscosité du solvant, serait souhaitable [90]. Cependant, puisque nous étudions des courtes séquences monomériques qui ne se replient pas sur elles-mêmes, mais qui s'attirent entre elles, l'effet implicite de l'eau suffit à nous donner une bonne idée de l'influence générale de la présence d'un solvant sur l'agrégation et nous n'avons pas besoin de quantifier l'interaction entre les monomères et l'eau.

D'autre part, dans un solvant implicite, tous les effets du solvant ne peuvent pas être reproduits, en particulier si des molécules d'eau jouent un rôle spécifique dans une région

spécifique de la protéine. Dans notre cas, l'interaction entre les peptides amyloïdes et les molécules d'eau à l'échelle microscopique ne joue un rôle que dans le maintien de la structure cross-β [32]. De plus, aucun rôle important spécifique des molécules d'eau n'a été démontré dans le processus d'agrégation outre l'effet global macroscopique stabilisant ou déstabilisant de la présence de l'eau autour d'agrégats solubles ou insolubles. De nombreux modèles tout-atome existent pour modéliser l'eau (SPC, SPC/E, TIP3P, TIP4P et TIP5P) [91], mais il y a beaucoup de débat autour de leur validité et de leur robustesse et la modélisation de l'eau est tout un problème en soit qui requiert encore beaucoup d'avancées algorithmiques.

2.2 Dynamique moléculaire

Puisqu'on ne connaît pas a priori le chemin de formation ou même la structure exacte des oligomères amyloïdes de GNNQQNY, des méthodes comme *umbrella* et *transition-path sampling*, qui permettent d'étudier la transition entre deux états connus pour des événements rares, n'auraient aucune utilité. Au contraire, nous cherchons plutôt à faire de la prédiction de structures combinée à une recherche dans l'espace des conformations sans connaissance de l'état oligomérique, mais avec une connaissance détaillée de l'état fibrillaire (qui correspond à une des possibilités mécanistiques de formation des protéines amyloïdes) et nous cherchons aussi à caractériser un phénomène cinétique avec le moins de biais possible sans connaissance structurelle des intermédiaires impliqués. La dynamique moléculaire nous permet d'atteindre de tels objectifs.
Un protocole standard de dynamique moléculaire consiste en trois étapes:

1. Une procédure de minimisation
2. Une phase de thermalisation
3. La production d'une trajectoire

2.2.1 Description de la méthode

2.2.1.1 Conditions initiales

Au tout début de la simulation, les structures sont converties en format et représentation OPEP puis placées dans une boîte cubique périodique. Les conditions aux frontières créent un environnement périodique autour de la boîte de simulation, pour imiter la présence d'une infinité de copies du système en contact par chaque face. Ainsi, si un monomère sort d'un côté de la boîte, il réapparaîtra de l'autre côté avec la même vitesse et la même direction pour lui permettre de revenir interagir avec le reste du sytème. L'utilisation de conditions aux frontières périodiques nous permet donc d'avoir une concentration constante et de maintenir un certain degré d'interaction entre les monomères du système tout en les laissant diffuser.

2.2.1.2 Minimisation

L'étape de minimisation sert à relaxer les liaisons et angles déformés et les contacts incorrects qui pourraient avoir été générés lors de la construction de la structure initiale et s'assure de minimiser l'énergie potentielle (soit la force sur tous les atomes) et l'énergie cinétique (soit la vitesse de tous les atomes) du système. Initialement, les vitesses des atomes sont aléatoires. La minimisation se fait ensuite, dans nos simulations, au travers d'une dynamique moléculaire amortie. Une recherche est alors effectuée dans l'espace des conformations en amortissant les vitesses, à partir d'un point arbitraire, et la direction choisie, celle de la force $\overrightarrow{F}$, est celle qui minimise la puissance $P = \frac{dE}{dt} = \vec{v} \cdot \vec{F}$ Après 10 000 pas de dynamique moléculaire amortie, la minimisation est réussie si la force $F < 0.2\ kcal/mol/$Å et que la puissance P est proche de 0.

2.2.1.3 Thermalisation

La thermalisation dans le cas particulier de nos simulations se fait en cinq étapes de 10 000 pas où la température est progressivement augmentée jusqu'à atteindre la température cible, définie par la distribution de vitesses des particules du système.
D'après le théorème d'équipartition, chaque degré de liberté a une énergie $k_B T/2$ en

moyenne à l'équilibre. Pour un système de N particules, l'énergie cinétique moyenne du système est donc donnée par:

$$\langle E_{cin} \rangle = \frac{1}{2} \langle \sum_{i=1}^{3N} m_i v_i^2 \rangle = \frac{3Nk_BT}{2} \tag{2.13}$$

où i désigne la i-ième particule du système, m_i sa masse, v_i sa vitesse, k_B est la constante de Boltzmann et T est la température. Ainsi, la distribution de vitesses, définie suivant une distribution de Maxwell-Boltzmann, satisfait la condition:

$$\langle v_i^2 \rangle = \frac{k_BT}{m_i} \tag{2.14}$$

où T est la température progressivement augmentée en cinq étapes durant la thermalisation et est la température cible pour le reste de la simulation.

2.2.1.4 Simulation

La production de la trajectoire en dynamique moléculaire se fait ensuite à température constante. Pour simuler les mouvements des particules sous l'effet du potentiel OPEP, la méthode d'intégration des équations du mouvement utilisée est l'algorithme "*velocity-Verlet*" [92] tel que, à chaque nouveau pas de simulation au temps $t + \tau$:

$$\begin{cases} r_i(t+\tau) = r_i(t) + v_i(t)\tau + \frac{1}{2}\tau^2 \frac{F_i(t)}{m_i} \\ \\ v_i(t+\tau) = v_i(t) + \frac{F_i(t)+F_i(t+\tau)}{2m_i}\tau \end{cases} \tag{2.15}$$

où t est le temps et τ est le pas de simulation qui est égal à 1.5 fs. $F_i(t)$ et $F_i(t+\tau)$ sont les ancienne et nouvelle forces sur la particule i obtenues à partir des positions atomiques et suivant le potentiel OPEP. $r_i(t+\tau)$ est la nouvelle position de la particule i au temps $t+\tau$, $r_i(t)$ est l'ancienne position de la particule au temps t et similairement pour les vitesses $v_i(t+\tau)$ et $v_i(t)$.

L'algorithme de contrainte RATTLE [93] est appliqué à l'intégration des équations du

mouvement, pour tenir compte des contraintes rigides appliquées sur les liaisons, dont la période de vibrations est en moyenne de l'ordre de quelques fs [94], ce qui nous permet d'augmenter le pas de temps des simulations, et donc l'efficacité, sans dégrader la qualité de la trajectoire.

Afin de maintenir la température constante au travers des vitesses des atomes, un couplage à un bain thermique de Berendsen est effectué. Les équations du mouvements sont alors modifiées par l'ajout d'un terme qui contrôle le taux de relaxation du couplage au bain. Ainsi les vitesses, gouvernant la température, sont multipliées par un facteur c tel que:

$$c = \sqrt{1 - \frac{\tau}{\tau_c}\left(1 - \frac{T_0}{T}\right)} \tag{2.16}$$

où τ est le pas de simulation, τ_c est le temps associé au couplage (100 fs dans notre cas), T_0 est la valeur cible de la température et T est la température cinétique instantanée [95, 96].

2.2.2 Avantages et limites de la méthode

L'avantage d'utiliser la dynamique moléculaire est la possibilité d'étudier l'évolution temporelle d'un système biologique, mais aussi ses proporiétés à l'équilibre, ce qui en fait un outil parfaitement adapté pour l'étude des structures oligomériques amyloïdes et de leur cinétique d'agrégation. La méthode d'intégration numérique des équations du mouvement, "*velocity-Verlet*" est préférée ici car c'est un algorithme stable, reversible dans le temps[4] et efficace puisqu'il n'est pas nécessaire de garder plus d'une génération de vitesses, positions et forces en mémoire pour évaluer le pas suivant. Le couplage à un bain thermique de Berendsen est une façon simple et efficace de maintenir une température constante et est facile à implémenter. Cependant, la présence d'un tel thermostat supprime les fluctuations d'énergie cinétique du système puisque la température est corrigée pour que toute déviation diminue exponentiellement selon:

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau_c} \tag{2.17}$$

[4] *i.e.* la simulation suit le chemin exactement inverse sous la transformation $\vec{v} \Leftrightarrow -\vec{v}$.

et donc ne génère pas un réel ensemble canonique, en particulier pour les petits systèmes. En revanche, pour des systèmes approchant la centaine ou le millier de particules[5], l'approximation de Berendsen est satisfaisante. Une solution pour remédier au problème d'ensemble serait d'implémenter une dynamique de Langevin [90] ou d'utiliser un thermostat de Nosé-Hoover [97] pour générer un ensemble canonique correct.

2.3 Échange de répliques en dynamique moléculaire

Le but de la méthode d'échange de répliques est d'améliorer l'échantillonnage de l'espace des phases à travers une recherche stochastique des conformations accessibles à un système, en permettant aux conformations, bloquées dans un minimum local sur le paysage énergétique de repliement d'une protéine [98](Fig. 2.2), d'explorer plusieurs minima d'énergie et de pouvoir capturer des événements rares [99]. La méthode d'échanges de répliques est une méthode très populaire qui consiste à lancer plusieurs simulations de dynamique moléculaire en parallèle et à des températures différentes. Chaque paire de trajectoires adjacentes, en températures, peuvent alors se voir échanger leurs températures, tous les 5000 pas, suivant une probabilité déterminée par le critère de Métropolis:

$$p(i,j) = \min\left\{1.0, \exp\left[\left(\frac{1}{k_B T_i} - \frac{1}{k_B T_j}\right)(E_i - E_j)\right]\right\} \tag{2.18}$$

où la trajectoire i a une énergie E_i à température T_i et la trajectoire j a une énergie E_j à température T_j. Les trajectoires peuvent ainsi visiter des basses comme des hautes températures, permettant au système d'acquérir une plus grande flexibilité dans ses conformations. Cette méthode permet de caractériser efficacement les propriétés thermodynamiques d'un système au détriment de l'information dynamique que nous procurerait une dynamique moléculaire simple.

[5] comme nos trimère, dodécamère, 20-mère et 50-mère GNNQQNY.

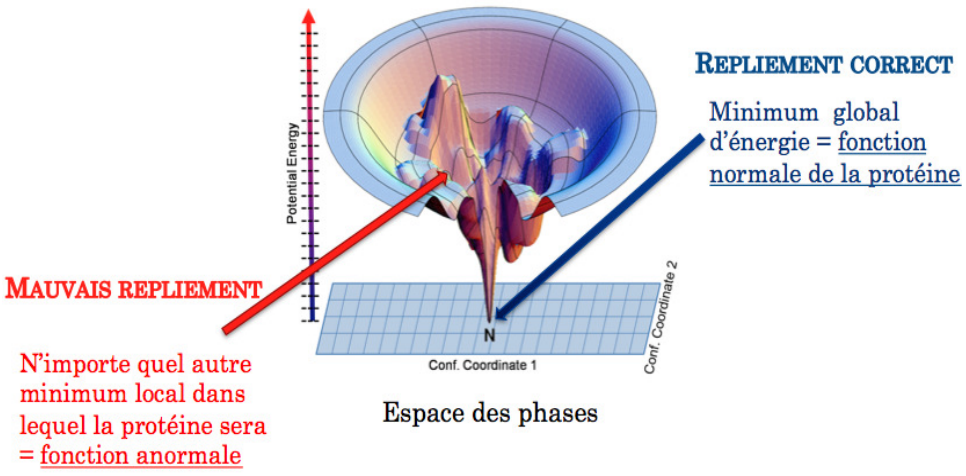

Figure 2.2: **Paysage énergétique de repliement d'une protéine.** (Dill, 1997)

2.3.1 L'importance du choix de températures

Lorsqu'on choisit l'échange de répliques en dynamique moléculaire, le choix et la taille de la distribution de températures sont cruciaux afin de permettre le meilleur échantillonnage possible. Cette distribution doit refléter le comportement de l'énergie potentielle du système en fonction de la température et, au point de fusion[6], *i.e.* dans les régions de températures où cette énergie varie soudainement, les températures de la distribution choisie pour les échanges de répliques doivent être rapprochées afin d'optimiser les échanges entre les trajectoires et d'assurer un échantillonnage uniforme. Si les températures ne sont pas suffisamment rapprochées autour du point de fusion, le système risque de voir ses configurations potentiellement bloquées dans un minimum local d'énergie puisque le critère de Métropolis, basé sur la facilité de recouvrement des énergies entre deux trajectoires, sera difficilement satisfait. La difficulté est de pouvoir prédire avec suffisamment de précision où se trouve le point de fusion avant de déterminer la distribution de températures et, dans certains cas, une distribution de températures autocorrective pourrait être un atout pour obtenir un meilleur échantillonnage au-

[6]Température de transition entre les états ordonnés et désordonnés du système.

tour de cette transition. L'implémentation de la méthode des échanges de répliques pour des systèmes de taille importante, pour lesquels le besoin d'avoir beaucoup de répliques est nécessaire pour un échantillonnage uniforme, peut aussi devenir un problème d'un point de vue du coût CPU, mais pour des systèmes de l'ordre du 50-mère GNNQQNY, le nombre de répliques requises ne dépasse pas 40.

2.4 Analyse des résultats

J'ai conçu et amélioré des outils d'analyse afin d'extraire les propriétés statistiques des structures obtenues en simulations REMD, mais aussi pour étudier les propriétés cinétiques d'agrégation en dynamique moléculaire simple pour l'étude du système GNNQQNY.

2.4.1 Détermination de clusters

La majeure partie de l'analyse des simulations présentées dans cette thèse a été effectuée à l'aide d'un outil de détermination de clusters, dont l'algorithme est adapté pour l'assemblage de courts peptides, et conçu pour classifier les groupes – ou clusters – de feuillets-β qui composent une structure. Nous supposons ici que tous les clusters qui se forment sont des feuillets-β, de par la forte propension de la séquence GNNQQNY à former des feuillets-β seulement. La classification et la détermination des clusters sont basées sur l'attachement séquentiel des brins, c'est-à-dire sur les liaisons hydrogène qui les relient à l'intérieur d'un feuillet. En d'autres termes, si trois brins A, B et C sont séquentiellement attachés dans un même feuillet-β tel que A forme des liaisons hydrogène avec B et B avec C, alors l'algorithme identifie le cluster comme étant le trimère ABC. Le critère de définition d'une liaison hydrogène est similaire à la définition DSSP [100], selon laquelle une liaison hydrogène a lieu entre deux acides aminés différents si:

$$E = 0.084 \left(\frac{1}{r_{ON}} + \frac{1}{r_{CH}} - \frac{1}{r_{OH}} - \frac{1}{r_{CN}} \right) \times 332 \text{ kcal/mol} < -0.5 \text{ kcal/mol} \qquad (2.19)$$

où E est l'énergie électrostatique de liaison entre les deux acides aminés, C et O sont les atomes du groupe carboxyle du premier acide aminé et N et H sont les atomes du groupe amine du deuxième acide aminé. Les r_{ij} sont les distances entre les atomes i et j. Un peptide, ou brin, appartient donc à un cluster s'il est attaché à un autre brin par au moins une liaison hydrogène. Un critère est aussi appliqué sur les angles dihédraux ψ et ϕ pour déterminer si un brin possède suffisamment d'acides aminés en conformation β (au moins trois par brin pour le 20-mère et un minimum de un par brin pour le 50-mère[7]) pour appartenir à un cluster. Si les angles ψ et ϕ correspondent à la région β du graphe de Ramachandran [75]: ψ(en degrés): [-180:-150;0:180], ϕ(en degrés): [-180:0;150:180], un acide aminé est en conformation β. Ainsi, si un peptide n'a pas assez d'acides aminés en conformation β, il est exclu du cluster. La détermination de clusters nous permet aussi de calculer des propriétés structurelles comme la longueur des feuillets-β, l'orientation des brins dans les feuillets (*i.e.* parallèles ou antiparallèles), la proportion de feuillets complètement parallèles ou complètement antiparallèles, mais aussi le degré de décalage des brins dans un feuillet.

Pour suivre l'évolution des propriétés cinétiques, l'algorithme permet aussi de calculer les taux d'association et de dissociation des clusters et de rassembler des statistiques sur les événements cinétiques microscopiques. Si on considère N_g la concentration de $g-$mères[8] pendant un processus d'agrégation, les phénomènes d'association et de dissociation peuvent se produire soit une molécule (un brin) à la fois (processus d'élongation de clusters par addition/perte de monomères) ou bien plusieurs molécules à la fois (soit un processus de fusion/fragmentation d'oligomères[9] ou bien un processus de formation/destruction d'oligomères de/en monomères) respectivement. Le taux net de création de $g-$mères est donné par:

$$\frac{\Delta N_g}{\Delta t} = C_g^{mono} + D_{g+1}^{mono} - (C_{g+1}^{mono} + D_g^{mono}) + C_g^{oligo} - D_g^{oligo} \tag{2.20}$$

[7]Le prérequis de trois acides aminés en conformation β est trop discriminatoire pour les clusters du 50-mère.

[8]cluster de g molécules.

[9]Un oligomère ici est défini comme étant tout agrégat constitué de deux molécules ou plus.

où C_g^{mono} et D_g^{mono} sont le taux de création de $(g-1)-$mères en $g-$mères et le taux de destruction de $(g+1)-$mères en $g-$mères alors que C_g^{oligo} and D_g^{oligo} sont les taux de création et de destruction de $g-$mères soit directement de/en monomères ou issus de la fusion/fragmentation d'oligomères. Tous les taux sont extraits par le code de détermination de clusters.

2.4.2 Autres outils d'analyse

Pour l'analyse thermodynamique des simulations REMD, la méthode d'analyse des histogrammes par assignation de poids (PTWHAM) est utilisée [101]. Cette analyse nous permet d'extraire la chaleur spécifique

$$C_v = \left(\frac{\partial E}{\partial T}\right)_V \tag{2.21}$$

où E est l'énergie potentielle du sytème et T est la température. Le pic de chaleur spécifique en fonction de la température permet d'identifier la température de fusion, transition entre les états ordonnés et désordonnés d'un système. Lorsque le pic de C_v est stable (*i.e.* à température fixe) dans le temps, la simulation a convergé, c'est-à-dire que le système a atteint un équilibre thermodynamique. D'autres quantités comme le rayon de gyration R_g peuvent être extraites de l'analyse WHAM.
Le programme STRIDE [102] est utilisé pour tout calcul de structure secondaire en dehors de la détermination de clusters.

2.5 Conclusion

Les méthodes de simulations et les outils d'analyse présentés dans ce chapitre constituent un ensemble d'algorithmes innovateurs visant à améliorer notre compréhension physique du comportement des systèmes biologiques complexes, dans notre cas l'agrégation du peptide GNNQQNY en agrégats amyloïdes. Les résultats obtenus de notre étude numérique des premières étapes d'agrégation de ce peptide sont détaillés

dans les chapitres 4, 5 et 6.

CHAPITRE 3

CINÉTIQUE D'AGRÉGATION AMYLOÏDE

Ce chapitre résume les points essentiels de la cinétique d'agrégation des protéines amyloïdes, développés plus en détails dans l'annexe 1, notamment les modèles soulignant l'importance de la présence d'espèces intermédiaires. Ces aspects de la cinétique d'agrégation seront ensuite explorés plus en détails dans le chapitre 5 dans le cas particulier de la séquence GNNQQNY.

3.1 L'assemblage des protéines amyloïdes: un processus complexe de nucléation

Les protéines amyloïdes présentent un comportement cinétique cohérent avec un phénomène de croissance par nucléation, tel qu'observé expérimentalement dans de nombreux cas. Cependant, ces phénomènes d'assemblage ne sont pas exactement bien décrits par la théorie classique de la nucléation et des éléments importants pour les décrire manquent à la théorie classique. Cette section se veut un court résumé des plus récents modèles, détaillés dans l'annexe 1, qui viennent expliquer ces éléments afin de compléter la théorie classique de la nucléation pour tenter de mieux comprendre et reproduire la cinétique d'agrégation amyloïde.

3.1.1 Nucléation et fibrillisation

Un phénomène de croissance par nucléation a lieu lorsque la polymérisation[1] ne peut avoir lieu qu'après la formation d'un noyau métastable critique qui nécessite le surmontage d'une importante barrière d'énergie libre (Figure 3.1) lorsque le système est supersaturé, tel que dicté par la théorie classique de la nucléation [103–106].

Cette même théorie ne prévoit, cependant, qu'un seul niveau de nucléation homogène et des modèles complémentaires [107–109] ajoutent un second niveau de nucléation (nucléation secondaire) pour décrire la fibrillisation de certaines protéines amyloïdes

[1] la croissance en fibres, aussi appelée fibrillisation

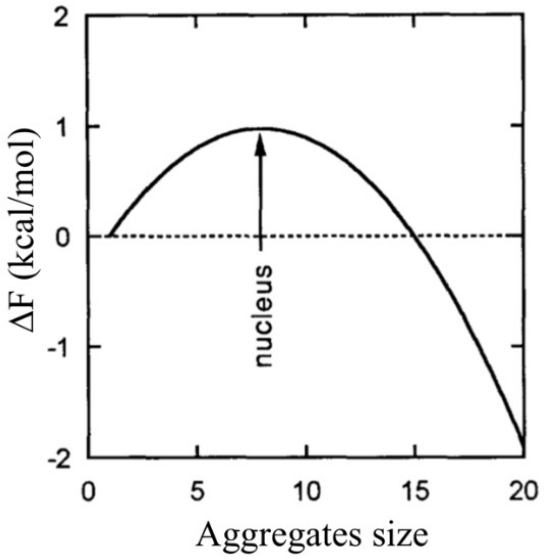

Figure 3.1: **Barrière d'énergie libre typique pour la formation d'un noyau critique.** Un noyau critique se forme lorsque le maximum d'énergie libre est surmontée par un agrégat. (Ferrone, 1999).

tels qu'observés expérimentalement. La nucléation secondaire peut se manifester sous plusieurs formes comme la fragmentation ou la nucléation hétérogène: la fragmentation consiste à augmenter le nombre d'extrémités de fibres disponibles pour recruter des monomères, agissant ainsi à titre de sites de nucléation [104, 109–114]. La nucléation hétérogène, quant à elle, consiste en la formation de nouveaux sites de nucléation le long d'une fibre existante et amenant à un phénomène d'épaississement des fibres [29, 115–120]. Les phénomènes de nucléation secondaires sont ainsi nécessaires à la description du phénomène d'agrégation de nombre de protéines amyloïdes et expliquent une croissance exponentielle des fibres lors de la polymérisation alors que la théorie classique de la nucléation prédit une vitesse de croissance quadratique. Une deuxième différence avec la théorie classique est la dépendance du comportement cinétique à la concentration initiale de monomères présents en solution. Alors que la théorie classique de la nucléation prévoit que le phénomène de croissance par nucléation dépende fortement de la quantité initiale de monomères, un modèle incluant la nucléation secondaire prévoit une dépendance notablement plus faible, confirmant ainsi plusieurs observations expéri-

mentales [107–109] (voir Annexe 1).
D'autres éléments manquant à la théorie classique de la nucléation incluent également la nature biochimique même de la séquence en acides aminés de la protéine considérée. En effet, plusieurs propriétés des protéines amyloïdes comme le degré d'hydrophobicité, la présence de résidus aromatiques ou bien la présence de motifs alternant résidus hydrophobes et hydrophiles, augmentent de manière significative le taux de nucléation amyloïde dans divers cas alors que d'autres propriétés comme la charge diminuent la vitesse d'agrégation [121]. Les effets extrinsèques de l'environnement cellulaire d'une protéine amyloïde peuvent également jouer un rôle non négligeable sur le taux de nucléation comme le pH, la température ou encore la présence d'une surface qui facilitent la nucléation [118]. Une vision d'ensemble est alors nécessaire afin de comprendre tous les paramètres en jeu qui dictent la dynamique d'agrégation chez les protéines amyloïdes et un des facteurs les plus importants à également considérer est le rôle que jouent les espéces intermédiaires dans le processus d'assemblage des protéines en fibres.

3.1.2 Rôle des espèces intermédiaires dans la fibrillisation

Les oligomères, espèce toxique préliminaire formée lors de la fibrillisation, peuvent mener à plusieurs morphologies suivant les protéines et les mécanismes en jeu : les oligomères sphériques désordonnés amorphes et les oligomères protofibrillaires plus ordonnés qui ressemblent structurellement aux fibres. Leur rôle est très important dans le mécanisme de formation des fibres et plusieurs modèles cinétiques les ont ainsi intégrer pour expliquer certaines observations expérimentales. Notamment, pour la protéine Aβ, un modèle proposé [122] suggère que les protofibres participent à la formation de fibres et sont créées par un processus de nucléation (Figure 3.2 (a)). En parallèle, d'autres hypothèses [33, 43] existent sur la nature de l'association des protofibres en fibres (Figure 3.2 (b)) et sur leur rôle potentiel de réservoirs de monomères [123] (Figure 3.2 (c)).
D'autre part, des modèles soulignent le rôle des oligomères amorphes, soit comme espèce embryonnaire préliminaire au noyau critique lors d'un processus de nucléation par addition de monomères [103, 107, 112, 114, 124–129] (Figure 3.3 (a)), soit comme acteur principal d'un mécanisme de fusion d'oligomères [43, 130–135] (Figure 3.3 (b)) qui

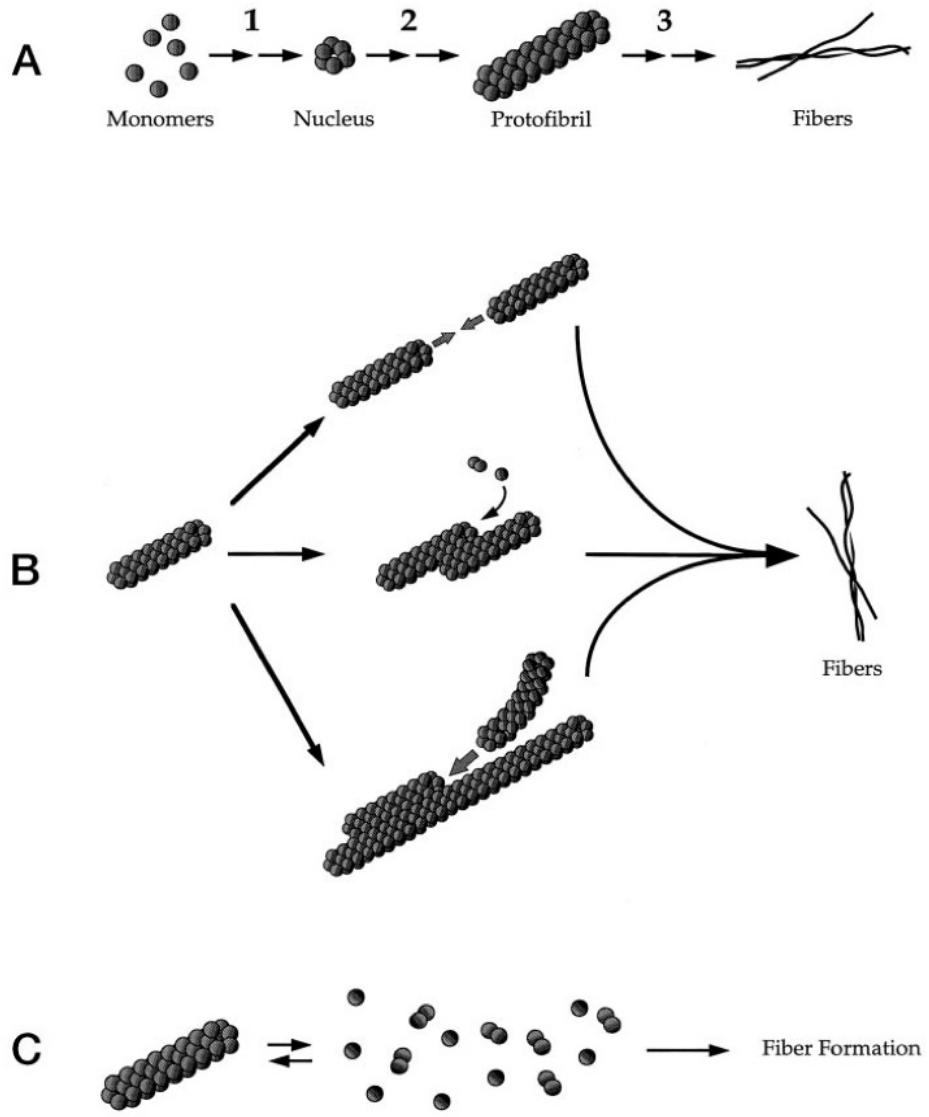

Figure 3.2: **Modèles decrivant la fibrillisation à partir de protofibres pour la protéine Aβ (Walsh *et al.*, 1997).** (a) Fibrillisation de Aβ en tant que phénomène de nucléation impliquant les protofibres. (b) Hypothèses sur la nature de l'association entre protofibres pour produire des fibres (Jansen *et al.*, 2005) (c) Protofibres en tant que réservoirs de monomères. Walsh *et al.* (1997).

pourrait agir de concert avec le mécanisme de nucléation. Bhak et al. [131] proposent alors que les fibres issues d'un processus de nucléation ou de fusion d'oligomères ont des morphologies différentes. Ainsi, un mécanisme de formation différent produirait des fibres de morphologies différentes pour une même protéine. Ce phénomène est appelé polymorphisme et est une propriété très répandue chez les protéines amyloïdes, notamment pour la protéine Aβ [3].

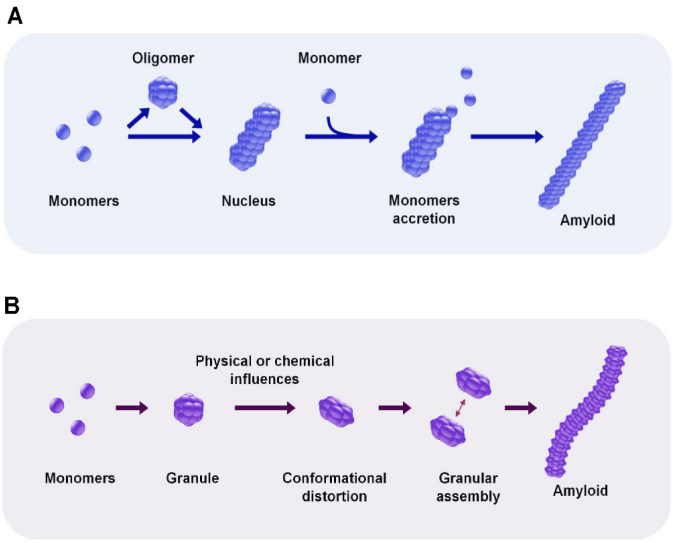

Figure 3.3: **Comparaison entre deux différents modèles décrivant la formation de fibres en impliquant des oligomères amorphes.** par (a) un processus de croissance par nucléation et (b) un processus de fusion d'oligomères. Rochet and Lansbury (2000).

3.2 Conclusions

Les protéines amyloïdes subissent ainsi un mécanisme de fibrillisation à la cinétique complexe dans lequel nombre de facteurs peuvent intervenir afin de faciliter la croissance en fibres. Dans le chapitre 5, nous explorons les mécanismes microscopiques de la cinétique d'agrégation du 20-mère de GNNQQNY en termes des éléments résumés dans ce chapitre et développés plus en détails dans l'annexe 1. Nous explorons également le degré de polymorphisme présent pour les agrégats allant du trimère au 50-mère de GNNQQNY dans les chapitres 4 et 6.

CHAPITRE 4

A MULTISCALE APPROACH TO CHARACTERIZE THE EARLY AGGREGATION STEPS OF THE AMYLOID-FORMING PEPTIDE GNNQQNY FROM THE YEAST PRION PROTEIN SUP35

4.1 Objectifs

Cette étude numérique multi-échelle, publiée dans le journal *PLoS Computational Biology* (PLoS Comput Biol, 2011, 7(5): e1002051. doi:10.1371/journal.pcbi.1002051), a pour objectifs d'étudier les premières étapes d'agrégation du peptide GNNQQNY pour des systémes de tailles différentes allant du trimère au 20-mère. En combinant l'efficacité d'un potentiel gros-grain à l'exhaustivité d'un potentiel tout-atome, les auteurs s'assurent de la stabilité et relevance de leurs résultats numériques. Cet article explore la thermodynamique associée à l'agrégation d'oligomères de GNNQQNY ainsi que les morphologies accessibles à ces oligomères. L'étude des oligomères amyloïdes, en tant qu'espèce pathogène reconnue, est cruciale dans le contexte du développement de molécules inhibitrices d'agrégats toxiques amyloïdes.

4.2 Contributions des auteurs

- (Dans l'ordre d'apparition dans l'article) Jessica Nasica-Labouze (JNL), Massimiliano Meli (MM), Philippe Derreumaux (PD), Giorgio Colombo (GC) et Normand Mousseau (NM) ont conçu ensemble le plan d'action pour cette étude numérique multi-échelle.
- JNL et MM ont réalisé les simulations, l'extraction et l'analyse des données.
- JNL, MM, PD, GC et NM ont chacun contribué à l'élaboration des outils de simulations, d'extraction de données et d'analyse.
- JNL, MM, PD, GC et NM ont participé à la rédaction de l'article.

- JNL était responsable de rassembler les travaux des trois groupes de recherches.

4.3 Author summary

The formation of amyloid fibrils is associated with many neurodegenerative diseases such as Alzheimer's, Creutzfeld-Jakobs, Parkinson's, the Prion disease and diabetes mellitus. In all cases, proteins misfold to form highly ordered insoluble aggregates called amyloid fibrils that deposit intra- and extracellularly and are resistant to proteases. All proteins are believed to have the instrinsic capability of forming amyloid fibrils that share common specific structural properties that have been observed by X-ray crystallography and by NMR. However, little is known about the aggregation dynamics of amyloid assemblies and their toxicity mechanism is therefore poorly understood. It is believed that small amyloid oligomers, formed on the aggregation pathway of full amyloid fibrils, are the toxic species. A detailed atomic characterization of the oligomerization process is thus necessary to further our understanding of the amyloid oligomer's toxicity. Our approach here is to study the aggregation dynamics of a 7-residue amyloid peptide GNNQQNY through a combination of numerical techniques. Our results suggest that this amyloid sequence can form fibril-like structures and is polymorphic, which agrees with recent experimental observations. The ability to fully characterize and describe the aggregation pathway of amyloid sequences numerically is key to the development of future drugs to target amyloid oligomers.

4.4 Abstract

The self-organization of peptides into amyloidogenic oligomers is one of the key events for a wide range of molecular and degenerative diseases. Atomic-resolution characterization of the mechanisms responsible for the aggregation process and the resulting structures is thus a necessary step to improve our understanding of the determinants of these pathologies. To address this issue, we combine the accelerated sampling properties of replica exchange molecular dynamics simulations based on the OPEP coarse-grained potential with the atomic resolution description of interactions provided by all-atom MD

simulations, and investigate the oligomerization process of the GNNQQNY for three system sizes: 3-mers, 12-mers and 20-mers. Results for our integrated simulations show a rich variety of structural arrangements for aggregates of all sizes. Elongated fibril-like structures can form transiently in the 20-mer case, but they are not stable and easily interconvert in more globular and disordered forms. Our extensive characterization of the intermediate structures and their physico-chemical determinants points to a high degree of polymorphism for the GNNQQNY sequence that can be reflected at the macroscopic scale. Detailed mechanisms and structures that underlie amyloid aggregation are also provided.

4.5 Introduction

The aggregation of soluble peptides and proteins first into soluble oligomeric assemblies and then into insoluble amyloid fibrils is associated with the onset of misfolding diseases such as Alzheimer's disease, Parkinson's disease, type II diabetes and transmissible spongiform encephalopathies [25, 26, 136–138]. Though there is no sequence similarity, the final products of all amyloidogenic proteins display a similar cross-β structure [32, 47] and the soluble oligomers of several proteins appear to share similar structural properties [11], suggesting common pathways for amyloid formation [11, 31, 139]. Structural similarity does not, however, exclude diversity or polymorphism in the intermediates and products of amyloid assembly [2, 3, 5, 12, 13, 120, 140–147].

Many studies have shown that soluble oligomeric intermediates are more toxic than the full fibrils themselves [9, 148]. These transient oligomers include low molecular weight aggregates (e.g. dimers [149] and tetramers [150]) and high molecular weight species (e.g., β-sheet rich annular protofibrils similar to pore-forming toxins [14, 151–153]). While oligomers are considered as primary toxic species for most neurodegenerative diseases, there is recent experimental evidence that fragmentation or breakage of fibrils can contribute to the kinetics of aggregation and the amyloid cytotoxicity itself [109, 154].

One important way for investigating amyloid fibril formation, polymorphism and

cytotoxicity is offered by short protein fragments. Among them, GNNQQNY, from the N-terminal prion-determining domain of the yeast protein Sup35, is a paradigmatic example of a short sequence with the same properties as its corresponding full-length protein [27, 155]. These properties include an amyloid fibril with a core cross-β spine, Congo-red binding and a nucleated-growth aggregation process [27]. In particular, X-ray diffraction of several micro-crystals provides a detailed atomic structure for different GNNQQNY fibrillar morphologies where the side chains form self-complementing steric zippers [27, 32, 47, 48, 155]. As for all amyloid sequences, structural characterization of the intermediate GNNQQNY oligomers has been however precluded experimentally due to the high complexity of the aggregation process, and the short-lived and meta-stable character of the early aggregates.

Computer simulations have proved useful complements to experiments for looking at the initial aggregation steps providing information, for example, about the presence of amorphous states in dynamic equilibrium with fibrillar and annular states [60, 156–158] and the final steps of the polymerization-nucleation process [12, 15]. They can provide atomic-resolution insights into several factors, ranging from the effect of sequence variations on aggregation tendencies to information on the stability of aggregates and the kinetics of aggregation. Due to lighter computational costs, short peptides are more amenable to simulations of the aggregation process than full-length proteins. For example, a number of numerical works have characterized the structures and free energy of small GNNQQNY aggregates ranging from 2-mers to 8-mers starting from disordered states or studied the stability of pre-formed GNNQQNY assemblies with cross-β or annular morphologies [12–14, 14–23].

In this paper, we push the boundaries of the GNNQQNY oligomer size and investigate, through a multi-scale simulation approach, the aggregation and polymorphism of three GNNQQNY oligomer sizes: 3-mers, 12-mers and 20-mers. Our approach takes advantage of the accelerated sampling properties of replica exchange molecular dynamics (REMD) simulations [99] based on coarse-grained models and of the accurate description of the physico-chemical interactions between the peptides and the solvent by using an all-atom model. More precisely, we first use REMD simulations [99] with the coarse-

grained potential OPEP [52, 53], and then analyze the stability and conformational properties of selected aggregates by room temperature MD as well as REMD simulations using the GROMOS force-field [91]. In total, we accumulated more than 23.60 μs and 2.66 μs of coarse-grained and all-atom simulations, respectively, allowing relevant statistical analysis. To our knowledge, the present study reports the largest simulations of spontaneous self-organization carried out at the atomic resolution on an amyloid peptide without any pre-formed seed. Overall, the results of our integrated simulations and analysis show the existence of a high degree of polymorphism for the GNNQQNY sequence, even for oligomeric assemblies containing as many as 20 monomers.

4.6 Materials and methods

Simulations and analyses presented here couple a number of approaches, which are described briefly in this section. The first set of simulations uses the coarse-grained OPEP potential with replica-exchange molecular dynamics (REMD). These are followed by all-atom simulations using GROMACS with MD and REMD. All simulations are labeled as follows: a number, which indicates the number of monomers, two letters indicating the force field (OP for OPEP and GR for GROMACS), a letter or number indicating the simulation and a label for the specific conformation studied (when appropriate) giving, for example: 01OP2-A1.

4.6.1 Replica-exchange molecular dynamics (REMD)

REMD is a thermodynamical sampling method that requires the running of N MD trajectories (or replica) in parallel at N different temperatures selected in order to optimize thermodynamical sampling [99]. At regular time intervals, conformational exchanges are attempted between adjacent simulation pairs according to the Metropolis criterion with accept-reject probability:

$$p(i,j) = \min\left\{1.0, \exp\left[\left(\frac{1}{k_B T_i} - \frac{1}{k_B T_j}\right)(E_i - E_j)\right]\right\} \tag{4.1}$$

where, before the exchange, trajectory i at temperature T_i has an energy E_i and trajectory j has an energy E_j at temperature T_j .

This broadly used method allows for conformations in a deep local minimum to explore other regions of the energy landscape by migrating to higher temperatures. While thermodynamical properties converge faster than with single temperature standard MD, dynamical information is lost due to temperature exchanges. It is still possible, however, to derive thermodynamically putative aggregation pathways by following the continuous trajectories through temperature space.

4.6.2 The Optimal Potential for Efficient peptide-structure Prediction (OPEP) force-field

OPEP is a coarse-grained protein model that uses a detailed representation of all backbone atoms (N, H, C_α, C and O) and reduces each side chain to one single bead with appropriate geometrical parameters and van der Waals radius. The OPEP energy function, which includes implicit effects of aqueous solution, is expressed as a sum of local potentials (taking into account the changes in bond lengths, bond angles, improper torsions of the side chains and backbone torsions), non-bonded potentials (taking into account the hydrophobic and hydrophilic properties of each amino acid) and hydrogen-bonding potentials (taking into account two- and four- body interactions) [52]. OPEP has been extensively tested on peptides using multiple approaches such as the activation-relaxation technique [159], Monte Carlo [53], MD [60] and REMD simulations [160], and greedy-based algorithms [161, 162]. OPEP is also appropriate for simulations of GNNQQNY. Preliminary test simulations on this peptide's dimer indicate that, at 300 K, the GNNQQNY relative orientation is a 60 to 40 probability in favor of the antiparallel dimer with a least two hydrogen bonds. This result is in general agreement with what was found by Strodel *et al.* with CHARMM19 and the implicit solvation potential EEFI [19] where both orientations of the strands are visited with similar probabilities.

4.6.3 OPEP simulation details

REMD were carried out using a 1.5 fs time-step, periodic boundary conditions with box sizes depending on the systems and a weak coupling to an external temperature bath [59, 62, 95]. Replica exchanges were attempted every 5000 steps and configurations saved every 5000 steps. Initial structures for 3-mer and the 20-mer simulations were constructed by placing random coil monomers between 12 Å to 50 Å apart (Fig. 4.1(a) and Fig. 4.2). For the 12-mer, the initial chains occupied four rows, with each peptide separated from the others by 15 Å (Fig. 4.1(b)). Because of the extensive sampling of REMD, all results are independent from this initial setup. For the 20-mer system, three OPEP-REMD simulations were launched. A preliminary REMD simulation (20OPp) was used to obtain a first estimate of the melting temperature (T_m = 283 K) and generate some representative conformations for all-atom MD refinement. The 20 initial temperatures were logarithmically distributed between 230 K and 450 K. Despite 200-ns simulation per replica, we found that the configuration space was not optimally sampled because of the existence of a large discontinuity in the potential energy when the system orders. Thus, a Gaussian distribution of temperatures around 283 K was deemed preferable to allow a better sampling of the phase space. The other two REMD simulations, running for 400 ns at each temperature, were started from the same random configuration (Fig. 4.2), but with an optimized Gaussian temperature distribution centered around 283 K: 20OP1 uses 20 temperatures (in Kelvins: 223.8, 249.2, 260.1, 266.0, 270.3, 273.8, 277.1, 280.1, 283.0, 285.9, 288.9, 292.2, 295.7, 300.1, 305.9, 316.8, 342.2, 370.1, 398.0, 426.0) and 20OP2 uses 22 temperatures, with two more temperatures below the transition, at 236.5 and 254.7 K, to increase exchanges between low-energy structures. All REMD simulations are summarized in Table 4.I.

Determining whether equilibrium has been reached, even for the trimer, is difficult. It is always possible that a system is stuck in a minimum and thermodynamical properties will then appear as though they are converged. Here, we use the specific heat to track convergence. This quantity, the second derivative of the free energy, is very sensitive to convergence at all temperatures, and provides a very stringent test even near transitions.

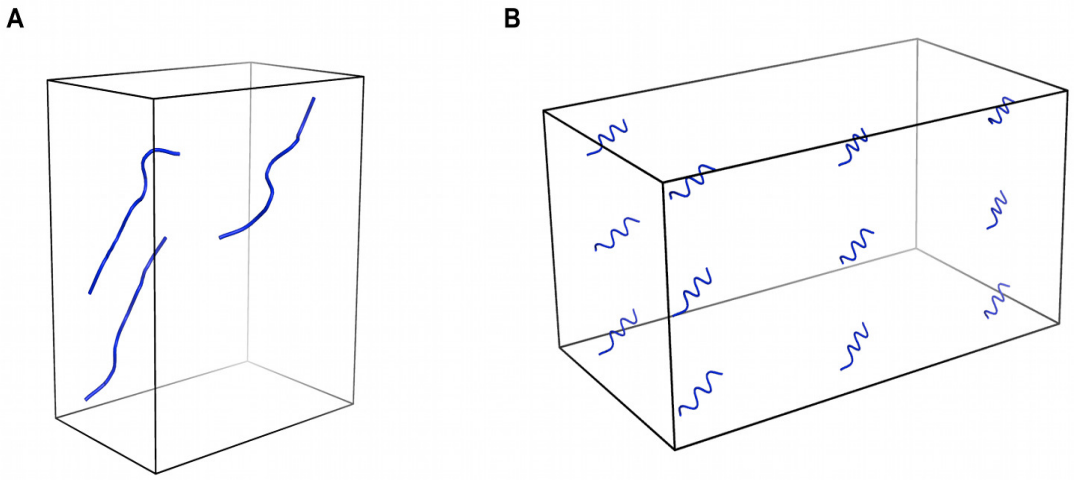

Figure 4.1: **Starting structures for a) the trimer and b) dodecamer.** The concentration for both systems is set at 4.15 mM. The random coil monomers are placed 15 Å apart.

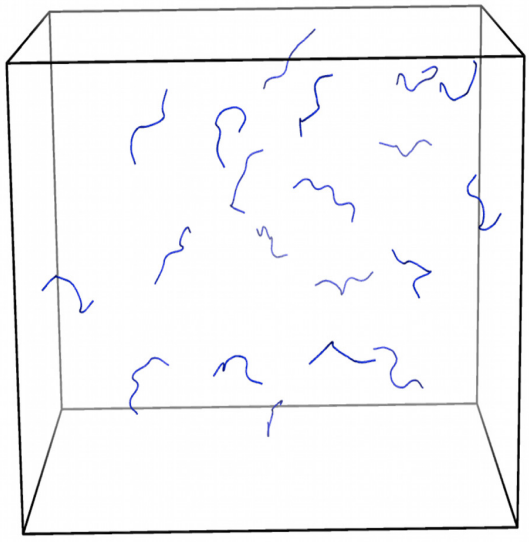

Figure 4.2: **Starting structure with random coils and no seed for the 20-mer simulations.** The concentration is also 4.15 mM. The monomers are randomly placed 12 to 50 Å apart.

Because we are mostly interested in the qualitative properties of the systems under study here, we consider that a system is converged when the overall shape of the specific heat near the transition is converged. This ensures that the dominant structures are found with the proper weight, within the limits of our simulations.

4.6.4 OPEP analysis and structure selection

Analysis for these simulations was performed, in part, using a new clustering code that enables us to identify the dominant configuration types in terms of clusters formed in β-sheet structures based on strand attachment. The criterion set to define a hydrogen bond between two given strands is similar to the one used in the DSSP algorithm. [163]. A cutoff of one hydrogen bond is used for distinguishing random from β-strands since we are dealing with a very short sequence and not considering the hydrogen bonds with the N-terminal glycines. The configuration types are defined here in terms of the number of sheets and the number of strands per sheet in the structure. For instance, a configuration type 8 7 5 for the 20-mer describes a structure with 3 β-sheets containing 8, 7 and 5 strands, respectively. The clustering code also provides information about the orientation of the strands in a β-sheet (*i.e.*, parallel or anti-parallel), alignment of the β-strands within a β-sheet (*i.e.*, in register or out-of register) and nature of the β-sheets (*i.e.* fully parallel, full anti-parallel or mixed orientations within a sheet). In addition to the clustering analysis, a PTWHAM analysis [101] was also performed on all of our data to compute thermodynamical properties.

In all cases, structures for all-atom simulations were taken among those of lower-energy OPEP that resisted most efficiently to a temperature increase during replica exchanges. For one preliminary simulation (20OPp) however, the structures were selected based on their frequency of occurrence.

	Length of OPEP simulations (ns)[a]	OPEP - LABEL of structures extracted[b]	Temperatures min-max (K) & number of temperatures[c]	GROMACS - LABEL of reconstructed OPEP extracted structures[b]	Length of GROMACS simulations (ns)[d]	Total number of atoms[e]	Temperature (K)[f]
3-mer	50×16	03OP1-A	222.5–525 16	03-GR1-A	100	6504	300
		03OP1-B	222.5–525 16	03-GR1-B	100	5997	300
		03OP1-C	222.5–525 16	03-GR1-C	100	5619	300
		03OP1-D	222.5–525 16	03-GR1-D	100	5949	300
		03OP1-E	222.5–525 16	03-GR1-E	100	6621	300
12-mer	125×16	12OP1-A	222.5–525 16	12-GR1-A	100	56271	300
		12OP1-B	222.5–525 16	12-GR1-B	100	40476	300
		12OP1-C	222.5–525 16	12-GR1-C	100	15381	300
		12OP1-D	222.5–525 16	12-GR1-D	100	16134	300
		12OP1-E	222.5–525 16	12-GR1-E	100	17157	300
20-mer OPp	200×20	20OPp-A	234.6–447.6 20	20GRp-A1	100	25620	300
		20OPp-A	234.6–447.6 20	20GRp-A2	100	25620	300
		20OPp-B	234.6–447.6 20	20GRp-B1	100	27816	300
		20OPp-B	234.6–447.6 20	20GRp-B2	100	27816	300
		20OPp-B	234.6–447.6 20	20GRp-B3 (REMD)	10×12	27816	see text
		20OPp-C	234.6–447.6 20	20GRp-C1	100	49674	300
		20OPp-C	234.6–447.6 20	20GRp-C2	100	49674	300
		20OPp-D	234.6–447.6 20	20GRp-D1	100	23746	300
		20OPp-D	234.6–447.6 20	20GRp-D2	100	23746	300
20-mer OP2	400×22	20OP2-A	223.8–425.9 22	20GR2-A	100	21586	300
		20OP2-B	223.8–425.9 22	20GR2-B	100	24640	300
		20OP2-C	223.8–425.9 22	20GR2-C	100	23554	300
		20OP2-E	223.8–425.9 22	20GR2-E	100	26065	300
		20OP2-N	223.8–425.9 22	20GR2-N1	100	32206	300
		20OP2-N	223.8–425.9 22	20GR2-N2 (REMD)	10×12	32206	see text

This table presents simulations done with OPEP (coarse-grained potential) and GROMACS (all-atom potential).
[a]The total simulation time for OPEP REMD simulations in the format time_per_replica x number_of_replicas.
[b]The label of the OPEP/GROMACS structures extracted. The label indicates the number of monomers, the potential used (OP for OPEP and GR for GROMACS), the simulation index (1,2 or p (preliminary)) and the letter ID of the structure.
[c]The range of temperatures (in K) used for OPEP REMD simulations.
[d]The total simulation time for GROMACS simulations. MD simulations are indicated by only one number while, for REMD simulations, the total simulation time is given in the format time_per_replica x number_of_replicas.
[e]The total number of atoms in the system including protein and solvatation water atoms.
[f]The temperature used in GROMACS simulations (in K).
doi:10.1371/journal.pcbi.1002051.t001

Table 4.I: **Details of all simulations run for the trimer, dodecamer and 20-mer systems.** This table presents simulations done with OPEP (coarse-grained potential) and GROMACS (all-atom potential). (a) The total simulation time for OPEP REMD simulations in the format time_per_replica x number_of_replicas. (b) The label of the OPEP/GROMACS structures extracted. The label indicates the number of monomers, the potential used (OP for OPEP and GR for GROMACS), the simulation index (1,2 or p (preliminary)) and the letter ID of the structure. (c) The range of temperatures (in K) used for OPEP REMD simulations. (d) The total simulation time for GROMACS simulations. MD simulations are indicated by only one number while, for REMD simulations, the total simulation time is given in the format time_per_replica x number_of_replicas. (e) The total number of atoms in the system including protein and solvatation water atoms. (f) The temperature used in GROMACS simulations (in K).

4.6.5 All-atom MD analysis of the conformational and stability properties of OPEP-generated, selected oligomeric structures

The initial structures for all-atom, explicit solvent Molecular Dynamics (MD) simulations were built by reconstructing the atomic detail of selected conformations from the OPEP coarse-grained runs. Reconstruction was carried out using the MAXSPROUT server [164]. Refinement of side chain rotameric states was performed using the program IRECS [165, 166], where the prediction is guided by a combination of potential interaction and rotamer scores calculated with probabilities from the backbone dependent rotamer library. Resulting all-atom structures obtained with this procedure were first minimized using the Macromodel package (Schrodinger Incorporated, USA) for 5000 steps with Polak-Ribier Conjugate Gradient method and an energy gradient criterion for convergence set to 0.05 kJ/mol. This minimization protocol was intended to initially remove unphysical contacts between atoms resulting from the reconstruction procedure, and not to optimize structures. At this stage, the C_α atoms were constrained to their positions with the default force constant (25 kcal/mol $Å^2$).

The resulting minimized systems were then solvated in a cubic-shaped box large enough to contain 1.0 nm of solvent around each initial aggregate. The simple point charge (SPC) water model was used [167] to solvate each oligomer in the simulation box. Each system was subsequently energy minimized with a steepest descent method for 5000 steps. The minimization was considered to be converged when the maximum force was smaller than 0.0001 kJ mol^{-1} nm^{-1}. The initial step size for minimization was 0.01 nm. The calculation of electrostatic forces was done with the PME implementation of the Ewald summation method. The LINCS [168] algorithm was used to constrain all bond lengths and the SETTLE algorithm [169] for the water molecules. Simulations were performed with a dielectric permittivity, $\varepsilon = 1$, and a time step of 2 fs. Initial velocities were taken from a Maxwellian distribution at the desired initial temperature of 300 K. The density of the system was adjusted performing the first equilibration runs at NPT condition by weak coupling to a bath of constant pressure (P0 = 1 bar, coupling time τP = 0.5 ps) [95] and the system was weakly coupled to an external temperature bath [95]

with a coupling constant of 0.1 ps. The proteins and the rest of the system were coupled separately to the temperature bath. Table 4.I summarizes the simulation conditions and number of peptides for each simulation. All simulations and analysis were carried out using the GROMACS package (version 3.3) [170–172] and the GROMOS96 43A1 force field [173–176].

For all MD simulations, aggregates were simulated at 300 K for 100 ns. REMD simulations were also used to investigate the stability and the conformational preferences of two 20-mer aggregates. The replica exchange simulations were carried out using the Solute Tempering REMD [177] protocol using the version implemented in GROMACS by de Groot and coworkers [178]. Twelve temperatures between 308 K and 419 K were selected according to [179] for an exchange probability of around 40 %.

4.7 Results/Discussion

The aggregation process for the three types of GNNQQNY oligomers – containing 3, 12 and 20 chains, respectively – was studied by a multi-scale approach consisting in a preliminary, thorough exploration of the phase space through REMD with the OPEP coarse-grained potential, followed by the refinement of the most representative aggregate structures obtained via all-atom MD or REMD simulations in explicit solvent. The initial concentration for the OPEP runs was around 4.15 mM. This concentration is 10 times higher than the concentration at which amyloid GNNQQNY fibrils form in a few hours according to Nelson *et al.* [32] allowing for the formation of ordered structures within our simulation time frame. The diversity in the number of chains allows us to examine possible intermediates and analyze molecular mechanisms of polymorphism in amyloid aggregates.

For clarity, we first present and discuss results for the trimeric and dodecameric systems as they will serve as basis for understanding the results observed for the 20-mer presented in the last part of this section.

4.7.1 Simulations of trimeric systems

4.7.1.1 Coarse-grained simulations

Coarse-Grained REMD simulations were performed with 16 replicas for 50 ns at temperatures discussed in the materials and methods section. Although the system is not fully converged for the very low-temperature replicas, the PTWHAM-generated specific heat computed over two different time intervals shows that the melting temperature, T_m, is well-established at $\sim$294 K (Fig. 4.3). Below this temperature, a clustering analysis shows that GNNQQNY monomers are assembled into ordered structures with high β-sheet content, while above T_m, the system visits mostly disordered structures with very low secondary structure composition. The alignment of individual strands within oligomers, the secondary structures and the configuration types of the aggregates are summarized in Table 4.II.

Structurally, the trimer displays a strong tendency to form ordered planar β-sheets below T_m (Fig. 4.4, left part of the panel). These appear rapidly, within 1 to 8 ns, in a mostly anti-parallel organization. Following trajectories leading to ordered structures, we see that the three-stranded β-sheet is always preceded by the formation of a mostly anti-parallel dimer seed. Averaging over all structures below T_m, we find that only a very small proportion of structures just below T_m consist of a two-stranded β-sheet interacting with one chain in coil conformation (1.9 %) or three random coil chains (1.1 %). The peptides at a temperature just below T_m prefer an anti-parallel β-strand order (87 % at 267 K) over a parallel arrangement (13 % at 267 K), while this proportion falls to 55–60 % at the lowest temperatures. As seen in Table 4.II, the three β-strands prefer to be perfectly aligned or in-registered at the lowest temperatures and are typically shifted by one residue, *i.e.* out-of-registered, at temperatures close to the melting point. As the temperature increases, the population of two-stranded and three-stranded β-sheets becomes very low, amounting to 8 % and 0 % at 333 K and 352 K, respectively. Except for the lowest temperatures, where mixed parallel/antiparallel sheets are most common, there is a clear dominance of fully antiparallel sheets for three-stranded structures while fully parallel sheets are rare, even among the few three-stranded sheets found above T_m,

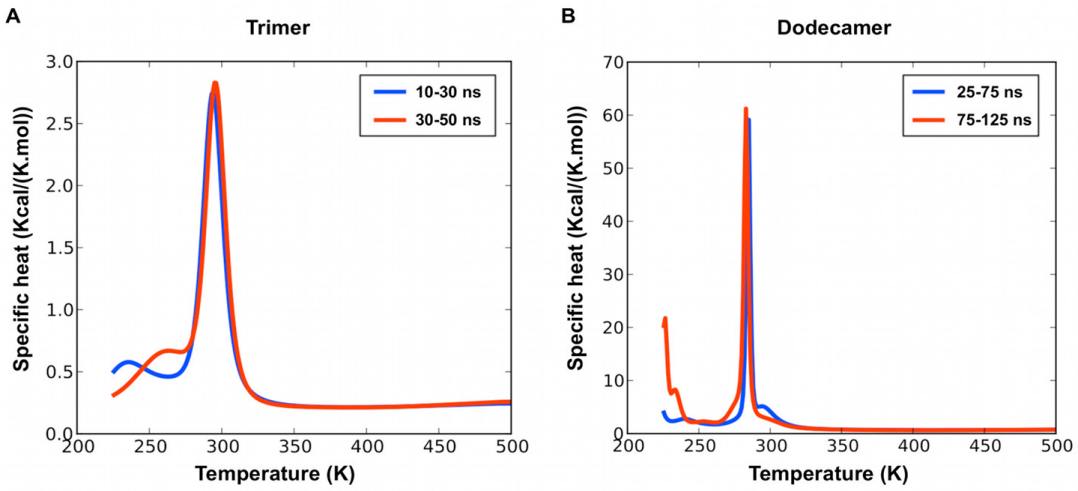

Figure 4.3: **Specific heat as a function of temperature for the trimer and dodecamer systems.** The specific heat is calculated over two time intervals for each system (trimer on the left panel and dodecamer on the right panel). Both systems have converged over the time windows displayed here.

where they reach 21 %, to 68 % for fully anti-parallel.

4.7.1.2 All-atom MD simulations

Five representative OPEP-generated structures, labeled 03OP1-A, 03OP1-B, 03OP1-C, 03OP1-D and 03OP1-E (Fig. 4.4, left side of the panel), were then subjected to all-atom MD simulations as described in materials and methods. These structures can be divided in two sets: 03OP1-A, 03OP1-B, 03OP1-C are characterized by three-stranded β-sheets with mixed parallel/anti-parallel β-strands, while 03OP1-D and 03OP1-E display a fully anti-parallel three-stranded β-sheet.

As seen in the final structures of the all-atom simulations displayed in Fig. 4.4 (right side of the panel), the five structures show different evolutions after the 100 ns all-atom MD. The three structures 03OP1-A, -B and -D tend towards configuration types 2-1, *i.e.* with one chain converted from β-strand to random coil and the two other chains enhancing their β-sheet contents. This inter-conversion is independent on the initial orientation of the strands. In contrast, the other two structures 03OP1-C and -E preserve

			Population						
			222.5 K	235.8 K	250.8 K	266.7 K	283.4 K	300 K	313.8 K
Trimer	Configuration types (%)[a]	3	96.3	100	98.2	98.2	92.6	13	1.9
		2 1	3.7	0	1.8	1.8	1.9	22.2	9.2
		1 1 1	0	0	0	0	5.5	64.8	88.9
	% parallel[b]		45.4	39.8	22.2	12	6.5	7.4	0
	% antiparallel[b]		54.6	60.2	77.8	88	88	24.1	5.6
	% fully parallel sheets[c]		0	0	1.9	0	0	21.1	0
	% fully antiparallel sheets[c]		9.3	20.4	57.4	75.9	86.3	68.4	50
	% mixed sheets[c]		90.7	79.6	40.7	24.1	13.7	10.5	50
	β-sheet content (%)[d]		72.4	70.6	66.6	61.6	55.1	33.3	28.3
	% Strands in-register/out-of-register by 1 residue[e]		51.9/46.3	47.2/48.2	47.2/38.9	38.0/43.5	17.7/45.1	21.1/42.1	33.3/41.7
Dodecamer	Configuration types (%)[a]	7 5	0	4.5	9.8	54.1	75.2	34.6	0
		7 4 1	30.8	33.1	12.8	1.5	0.8	1.5	0
		8 4	0	11.3	38.4	36.8	5.3	1.5	0
		12	66.9	7.5	9	0.8	1.5	0.8	0
		11 1	2.3	15	9	3	1.5	0.8	0
		10 1 1	0	9	6	0	3	1.5	0
	% parallel[b]		32.3	31.3	40.9	47.9	41.9	37.9	17.1
	% antiparallel[b]		60.2	48.4	45.8	46.7	49.5	41.1	16
	% fully parallel sheets[c]		0	0.4	1.2	0.4	0.8	15.3	22.6
	% fully antiparallel sheets[c]		6.3	10.8	6.2	2.3	2.7	11.4	21.8
	% mixed sheets[c]		93.7	88.8	92.6	97.3	96.5	73.3	55.6
	β-sheet content (%)[d]		75.8	59.1	61.7	63.5	61.8	43.8	11.4
	% Strands in-register/out-of-register by 1 residue[e]		52.5/27.1	33.5/46.3	39.2/41.7	44.7/43.4	38.8/54.0	30.6/45.6	29.5/39.7

Temperatures above 313.8 K are not displayed here since they are populated essentially by conformations with random coil monomers with no secondary structure. The percentages are calculated over all the structures obtained in the last 40 ns (trimer) and in the last 100 ns (dodecamer) of the OPEP REMD simulations, where the systems have converged.

[a]The dominant configuration types (as described in the OPEP Analysis and Structure Selection section).

[b]The average amount of parallel and anti-parallel strands in the β-sheets formed. The sum of parallel and antiparallel strands in a structure does not always total 100% if the structure sees strands in an undefined orientation, i.e. attached by only one hydrogen bond.

[c]The average amount of fully parallel, fully antiparallel and mixed sheets.

[d]The average amount of residues in a β conformation.

[e]The average amount of strands in-register and out-of-register (by one residue) in β-sheets.

doi:10.1371/journal.pcbi.1002051.t002

Table 4.II: **Structural characteristics for small aggregates as a function of temperature.** Temperatures above 313.8 K are not displayed here since they are populated essentially by conformations with random coil monomers with no secondary structure. The percentages are calculated over all the structures obtained in the last 40 ns (trimer) and in the last 100 ns (dodecamer) of the OPEP REMD simulations, where the systems have converged. (a) The dominant configuration types (as described in the OPEP Analysis and Structure Selection section). (b) The average amount of parallel and anti-parallel strands in the β-sheets formed. The sum of parallel and antiparallel strands in a structure does not always total 100 % if the structure sees strands in an undefined orientation, *i.e.* attached by only one hydrogen bond. (c) The average amount of fully parallel, fully antiparallel and mixed sheets. (d) The average amount of residues in a β conformation. (e) The average amount of strands in-register and out-of-register (by one residue) in β-sheets.

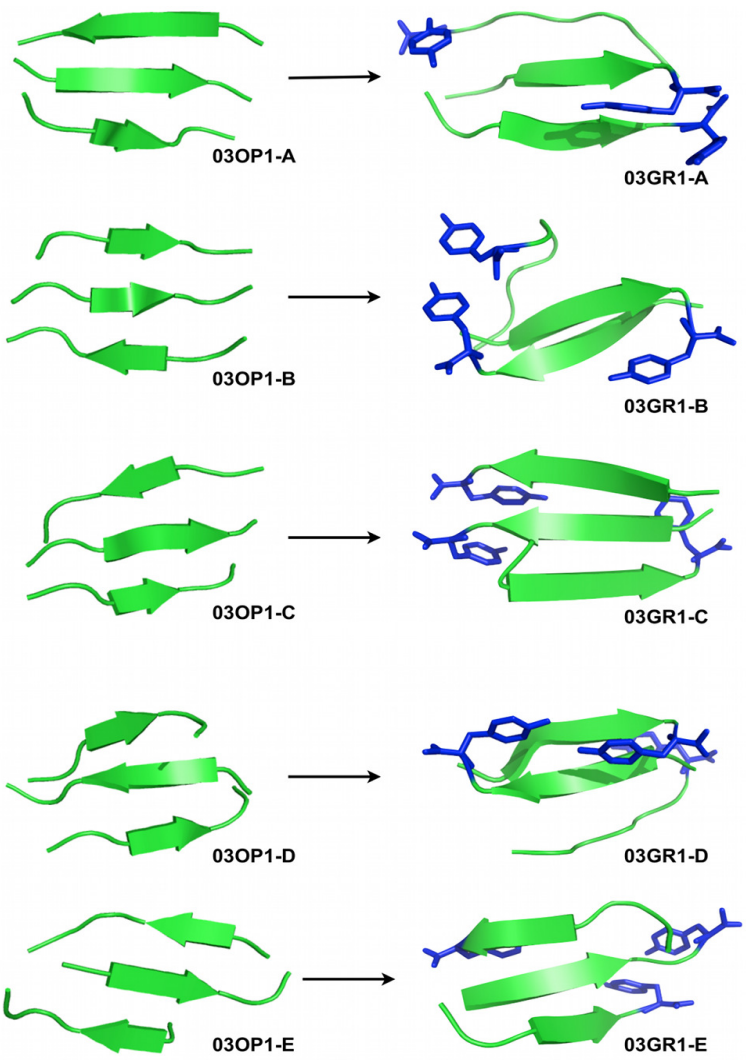

Figure 4.4: **Structures obtained for the trimeric simulations.** We show, on the left-handside panel, representative structures obtained from the OPEP simulations and, on the right-handside panel, the representative structures obtained after all-atom MD refinements. 03OP1-A,-B,-C,-D and -E were extracted respectively at 222.5 K (probability of occurrence for this β-strand organization: 91 %), 235.7 K (80 %), 250.8 K (41 %), 266.7 K (76 %) and 283.4 K (86 %). 03OP1-A to -C are mixed β-sheets while 03OP1-D and -E are fully antiparallel β-sheets. The all-atom structures are represented in secondary structure cartoon and only the tyrosines (most hydrophobic residues in the sequence) are shown in blue sticks (hydrogen atoms are omitted).

their three-stranded β-sheet configurations and enhance their β-sheet contents. Simulation 03GR1-C keeps its starting mixed parallel/anti-parallel configuration of the strands; in the simulation 03GR1-E, one of the peptide flips orientation leading to a perfectly aligned mixed β-sheet from an initial fully anti-parallel sheet.

Even though all-atom simulations cannot capture fully disordered chains within 100 ns at 300 K, the coarse-grained and all-atom simulations indicate that both parallel and anti-parallel arrangements can be found in multiple meta-stable two-stranded and three-stranded structures, with various registers of hydrogen bonds contributing to the structural richness and conformational variability of the trimeric aggregates.

Our trimeric results point to the existence of three minima associated with parallel, antiparallel and mixed parallel/antiparallel β-sheet structures, and are consistent with previous computational studies at the all-atom level on the GNNQQNY trimer [17, 18, 23]. Our conformational distribution for the trimer is not biased, therefore, from the use of the OPEP coarse-grained potential. We emphasize that the population of the fully parallel and antiparallel β-structures in small aggregates vary substantially with the selected force field. Using CHARMM force field and the EEF1 implicit water model, Wales *et al.* predicted equal populations for both states from free energy calculations [19]. Lai *et al.* using multiple MD simulations with the Gromos force field and the SPC explicit water models found many transitions between both states [18], while Reddy *et al.* using the same Gromos force field and the SPC explicit water model predicted a much higher population for the parallel geometry [20].

4.7.2 Simulations of dodecameric systems

4.7.2.1 Coarse-grained simulations

OPEP-REMD was performed with the 16 replicas as in the case of the trimer, but each for 125 ns. Within the first 25 ns, the system converges at low temperature to β-sheet rich structures where the strands prefer an antiparallel orientation, as for the trimer, but with a lower melting temperature of 283 K (see Fig. 4.3) even though the potential energy per monomer in the ordered phase is much lower, reaching -37.0 kcal/mol/monomer

for the 12-mer compared to -18.4 kcal/mol/monomer for the trimer, indicating a clear bias toward aggregation and resulting in a much more marked peak in the specific heat.

Kinetically, the aggregation tendency for the dodecamer is to first form one or two stable four-stranded β-sheets that show little dissociation and that trigger the transient formation of one or two longer β-sheets. The formation of a trimer that precedes the four-stranded β-sheet shows, however, a higher dissociation/association rate. Interestingly, the tendency of the GNNQQNY sequence to form stable tetrameric aggregation nuclei had already been noticed in a previous investigation on the system [17] and was proposed by the Eisenberg group on the basis of entropic and energetic arguments [32]. The final stable ordered structures are shown in Fig. 4.5 (left side of the panel).

As would be expected, a rich set of ordered configurations is visited for the 12-mer (Table 4.II). Regrouping all structures below melting, the dominant conformation, visited 63 % of the time, is a two β-sheet structure with a 7 or 8-strand sheet stabilized by a smaller, 4-5 strand sheet positioned on top (Fig. 4.5, structures 12OP1-B to -E). Single sheets, with 11 or 12 strands also appear with a frequency of 23.3 % below melting (Fig. 4.5, structure 12OP1-A). Surprisingly, strand orientation probabilities vary significantly going from the 3-peptide to the 12-peptide system. As for the 3-peptide system, the anti-parallel orientation is favored below melting for the 12-peptide system especially at the lowest two temperatures where the probability of forming anti-parallel is between 60 % and 45 % compared to 30 % for the parallel. Then, as the temperature is increased, the amount of parallel and anti-parallel orientation becomes almost the same, suggesting that while anti-parallel orientation is energetically favored, it is rapidly overcome by the entropic gain of mixing orientations. The alignment of the β-strands is a mix of perfectly aligned strands and strands misaligned by one residue at all temperatures below the melting point. Because sheets are longer than for the trimer, the 12-mer comprises mostly β-sheets with strands in mixed orientations at low temperatures below T_m with a low probability of forming fully parallel or fully antiparallel sheets (Table 4.II). Interestingly in the few and much smaller sheets observed just above T_m, fully parallel and antiparallel β-sheets form with almost identical probability (data not shown), suggesting that with slower growth, structures visited below T_m could be more ordered.

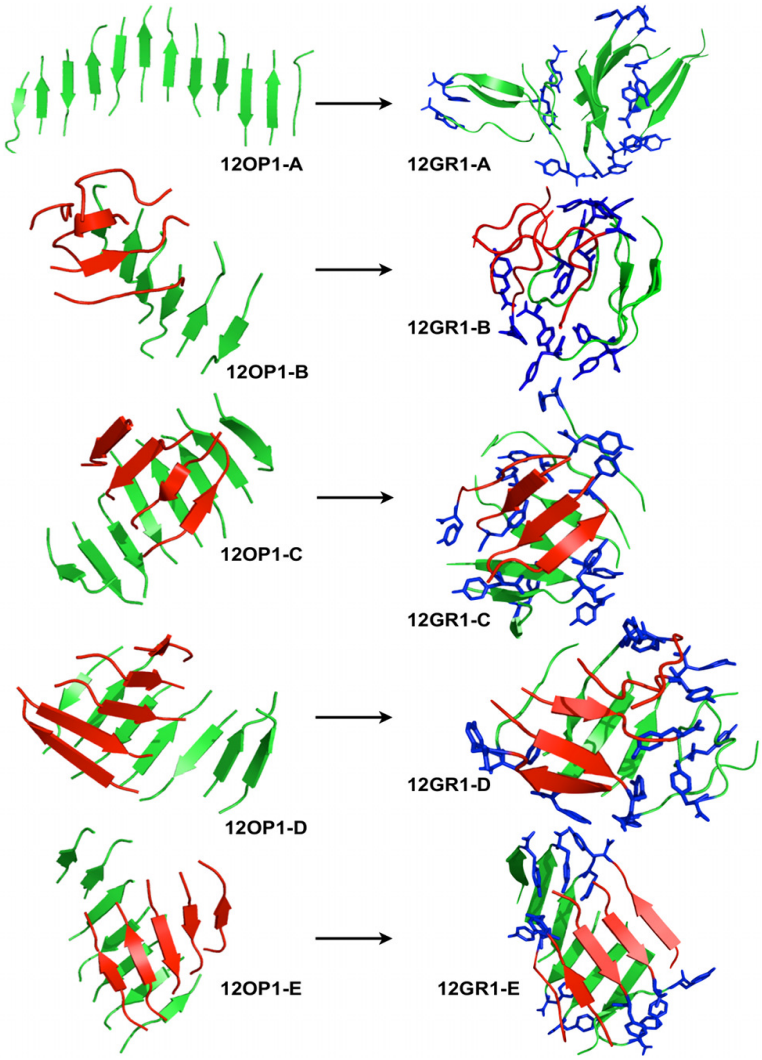

Figure 4.5: **Structures obtained for the dodecameric simulations.** We show, on the left-handside panel, representative structures obtained from the OPEP simulations and, on the right-handside panel, representative structures obtained after all-atom MD refinements. 12OP1-A,-B,-C,-D and -E were extracted respectively at 222.5 K, 235.7 K, 250.8 K, 266.7 K and 283.4 K. 12OP1-A (top left structure) is a long flat beta-sheet. 12OP1-B to -E (second left to bottom left structures) are made of 2 beta-sheets facing each other. Monomers forming β-sheets in the initial state are colored red or green. These colors are kept in the final structure. The tyrosines are shown in blue sticks for the all-atom structures. During the all-atom MD simulation the structures tend to be more globular but the strands see no exchange between the β-sheets, *i.e.* the red and green β-sheets do not dissociate for the 12-mer system.

4.7.2.2 All-atom MD simulations

The 5 most representative structures obtained from OPEP REMD (labeled 12OP1-A to 12OP1-E) were further studied by all-atom MD. Representative structures obtained from the latter simulations are shown in Fig. 4.5 right panel. The 12OP1-A OPEP structure is characterized by the presence of a flat arrangement of β-sheets. It undergoes significant rearrangements during the all-atom evolution in explicit solvent (12GR1-A), as shown by the time evolution of the radius of gyration (Fig. 4.11), with the planar β-sheet breaking into four fragments of two- to four-stranded β-sheets that assemble on top of each other, with two central parallel β-sheets covered on both sides by a perpendicular β-sheet. The overall amount of β-sheet structure is conserved during the all-atom simulation (Table 4.IV).

Structure 12OP1-B is characterized by a mainly parallel twisted β-sheet, with four strands packed on top. This structure is not stable in the all-atom MD setting, simulation 12GR1-B, and evolves towards a compact globular structure as shown by the evolution of the radius of gyration in time (Fig. 4.11). Interestingly, the external side of the final aggregate is lined with hydrophilic Asn and Gln side chains that provide favorable contacts with the solvent. No specific order is observed for contacts among these side chains, although some cases of interdigitation as seen in the final steric zipper are noticed. The interior of the final aggregate is lined with Tyr aromatic side chains.

Such a supramolecular organization of the peptides may be representative of one of the soluble intermediates on the pathway to fibril formation. Solubility is favored by the presence of hydrophilic side chains on the external surface of the aggregate. At the same time, the packing of the interior is not optimal, so that the resulting structure may not be in the most favorable arrangement to ensure lasting stability. Water can also access the interior of the globular aggregate, disrupting inter-strand hydrogen bonds, eventually favoring conformational changes.

Structures 12OP1-C and 12OP1-D are similar to 12OP1-B: the main difference is that four strands pack with their long axis almost perpendicular to the long axis of the extended β-sheet. The main difference between 12OP1-C and 12OP1-D is that the planes

defined by the four strands have different inclinations with respect to the plane of the long extended β-sheet. In the all-atom MD setting – simulations 12GR1-C and 12GR1-D – these structures evolve to less globular, but more compact final arrangements than that observed above, with most of the Tyr side chains in contact with the solvent (Fig. 4.11). The exterior of the aggregates is lined with Asn, while the interior is more compact than for 12GR1-A and 12GR1-B and packed with the side chains of Gln, that form a network of van der Waals and hydrogen bonding contacts.

Finally, structure 12OP1-E is characterized by two orthogonal twisted β-sheets. The OPEP structure is very stable: it does not undergo significant rearrangement during the all-atom MD, contrary to the previous cases, and the β-sheet content remains constant (Table 4.IV). The oligomer is trapped in this conformation by the extensive contacts packing determined by the Tyr side chains in the two sheets. Moreover, the inter-sheet space is filled by Asn and Gln side chains. However no specific packing into the ordered steric zipper is evident.

Table 4.IV recapitulates the conformational heterogeneity and plasticity of the 12-mer aggregates. As a general case, the presence of explicit solvent tend to condense OPEP-generated structures, at the expense of structured β-sheets and the associated parallel-antiparallel structure, strand alignment and register. It must be kept in mind, though, that MD simulations may be affected by sampling limitations associated with the short runs and the presence of solvent.

Overall, the combined results indicate that the configurational richness increases from the trimer to the 12-mer and that the critical nucleus has not yet been found. Though, the strands do not see much exchange between sheets as seen in Fig. 4.5. While ordered 12-mers are energetically much more favorable than the trimers, entropic factors may be considered prevalent, favoring a wide variety of metastable structures. The presence of explicit solvent decreases significantly the stability of elongated β-sheets either by increasing the effective hydrophobic interactions or decreasing entropic gains, favoring rather more compact structures. Different molecular mechanisms may be responsible for the stabilization of different conformations, endowed with different solubility properties. Indeed, we have observed globular-like structures with an external region

decorated with hydrophilic groups that may determine the oligomers to be soluble in aqueous solution. In contrast, more ordered structures with higher β-sheet content appear to expose more hydrophobic area to the contact with the solvent. In turn, the latter may recruit more monomers or preformed oligomers that can aggregate by the juxtaposition of hydrophobic surfaces. The observations on the 12-mer systems also underline the enormous structural diversity that characterizes the aggregation of amyloidogenic peptides, which is reflected at the macroscopic level in a high degree of polymorphism.

4.7.3 Simulations of 20-mer systems

Next, we turned to the study of 20-mers in order to assess the importance of the number of chains on the final supra-molecular organization and determine whether new structural motifs can emerge.

4.7.3.1 Coarse-grained simulations

Three REMD simulations with OPEP were thus generated for the GNNQQNY 20-mer systems: 20OPp, 20OP1 and 20OP2. A preliminary run 20OPp was run to identify the four most common low-energy clusters, from which we extract the central structure for each: 20OPp-A, 20OPp-B, 20OPp-C and 20OPp-D (Fig. 4.6, left panel). These were used as starting points for MD simulations with GROMACS. The first three are two-sheet structures while the fourth is a three-sheet configuration. What is particularly interesting here is that we obtain a protofibril-like structure (20OPp-B) among the most dominant clusters after only 200 ns starting from a random coil configuration. Interestingly, the protofibril-like structure is possible but not dominant in this preliminary simulation.

Following this preliminary run, we have performed two additional simulations 400ns-long 20OP1 and 20OP2 (Fig. 4.7) to attempt to better sample the phase space to determine the degree of preference and the importance of the protofibril-like structure among the morphologies accessible to that sequence for twenty peptides. Even after 400 ns, however, neither simulation is fully converged and the melting temperature is evaluated, from specific heat, to be at 280 K or higher, with ordered structures forming success-

			β-Sheet Contenent (based on the dssp program)[b]				% Parallel - Antiparallel[c]				% Res Align (% 100 Align - % Shift by 1 Res - % Shift by 2 Res...)[d]			
			GROMACS - GR		OPEP - OP		GROMACS - GR		OPEP - OP		GROMACS - GR		OPEP - OP	
	Configuration Type[a]	Structure	First Cluster	Final Structure	CG	Min	First Cluster	Final Structure	CG	Min	First Cluster	Final Structure	CG	Min
Trimers	3	03_1 - A	38%	19%	71% (100%)	67%	50 - 50	50 - 0	50 - 50	50 - 50	50 - 50	50 - 50	50 - 50	50 - 50
	3	03_1 - B	38%	24%	19% (27%)	67%	0 - 50	50 - 50	50 - 50	50 - 50	50 - 50	50 - 0 - 50	50 - 50	50 - 50
	3	03_1 - C	57%	62%	62% (87%)	67%	50 - 50	50 - 50	50 - 50	50 - 50	50 - 50	50 - 50	50 - 50	50 - 50
	3	03_1 - D	48%	57%	52% (73%)	67%	0 - 50	0 - 50	0 - 50	0 - 100	50 - 50	50 - 50	50 - 50	50 - 50
	3	03_1 - E	48%	52%	57% (80%)	53%	50 - 50	0 - 50	0 - 50	0 - 100	50 - 0 - 50	50 - 0 - 50	0 - 50 - 50	0 - 50 - 50
12-mers	12	12_1 - A	49%	35%	64% (90%)	60%	60 - 20	40 - 10	36 - 64	36 - 64	80 - 20	50 - 50	64 - 18 - 18	64 - 18 - 18
	6 4 1	12_1 - B	14%	14%	45% (63%)	44%	13 - 25	20 - 40	22 - 22	33 - 33	35 - 25 - 37	0 - 40 - 60	33 - 56 - 11	33 - 56 - 11
	8 4	12_1 - C	35%	29%	49% (68%)	42%	10 - 20	40 - 20	60 - 40	40 - 30	40 - 40 - 20	50 - 20 - 30	60 - 20 - 20	60 - 20 - 20
	7 5	12_1 - D	33%	24%	54% (75%)	55%	11 - 22	0 - 30	40 - 60	10 - 60	45 - 44 - 11	50 - 50	50 - 50	50 - 50
	7 5	12_1 - E	56%	51%	46%	51%	60 - 40	50 - 20	60 - 40	60 - 40	60 - 30 - 10	50 - 50	60 - 30 - 10	50 - 40 -10
20-mers	11 7 1 1	20_p - A1	43%	38%	38% (53%)	37%	44 - 6	56 - 13	76 - 24	44 - 25	56 - 31 - 13	63 - 19 - 18	34 - 33 - 33	44 - 37 - 19
		20_p - A2	31%	34%	38% (53%)	37%	33 - 0	50 - 0	76 - 24	44 - 25	40 - 40 - 20	63 - 31 - 6	34 - 33 - 33	44 - 37 - 19
	11 7 1 1	20_p - B1	52%	49%	36% (50%)	66%	69 - 19	59 - 12	71 -24	75 - 25	31 - 44 - 25	35 - 47 -18	33 - 39 - 28	31 - 44 - 25
		20_p - B2	47%	38%	36% (50%)	66%	40 - 13	38 - 13	71 -24	75 - 25	34 - 53 - 13	31 - 56 - 13	33 - 39 - 28	31 - 44 - 25
	11 7 1 1	20_p - C1	39%	35%	44% (44%)	59%	44 - 19	47 - 20	72 - 22	69 - 25	31 - 50 - 19	54 - 33 -13	33 - 39 - 28	31 - 44 - 25
		20_p - C2	59%	36%	44% (44%)	59%	69 - 25	69 - 8	72 - 22	69 - 25	27 - 47 - 26	46 - 54	33 - 39 - 28	31 - 44 - 25
	12 7 1	20_p - D1	44%	46%	44% (62%)	64%	24 - 35	24 - 35	35 - 47	41 -47	59 - 35 - 6	65 - 33 -6	59 - 35 - 6	59 - 35 - 6
		20_p - D2	27%	27%	44% (62%)	64%	29 - 24	22 - 22	35 - 47	41 - 47	63 - 31 - 6	61 - 33 -6	59 - 35 - 6	59 - 35 - 6
	9 6 5	20_2 - A	44%	44%	34% (47%)	28%	69 - 13	61 - 11	72 - 22	73 - 13	31 - 44 -25	39 - 39 - 22	56 - 39 - 5	20 - 53 - 27
	8 7 5	20_2 - B	45%	39%	39% (55%)	71%	41 - 41	17 - 22	71 - 18	41 - 59	41 - 18 - 41	56 - 28 - 16	35 - 53 - 12	41 - 24 - 35
	7 7 6	20_2 - C	36%	37%	52% (73%)	65%	44 - 17	41 - 35	76 - 24	59 - 41	50 - 39 - 11	35 - 47 - 18	30 - 41 - 29	41 - 47 - 12
	8 8 4	20_2 - E	32%	33%	51% (72%)	63%	28 - 11	42 - 16	59 - 41	50 - 39	44 - 39 - 17	42 - 37 - 21	54 - 23 -23	39 - 33 - 28
	10 8 2	20_2 - N	41%	39%	42% (59%)	57%	39 - 22	41 - 24	31 - 56	31 - 56	44 - 28 - 28	29 - 47 - 24	44 - 25 -31	44 - 25 - 31

"First Cluster" means the most representative structure of the GROMACS simulations. "Final structure" is the final conformation obtained at the end of the GROMACS simulations. "CG" is the structure extracted at the end of the OPEP simulations before the reconstruction of the side chains. "Min" indicates the structure resulting from the reconstruction of the side chains after a minimization step.

[a]The configuration type (as described in the OPEP Analysis and Structure Selection section).

[b]The average amount (percentage) of residues in a β conformation. For OPEP, the percentage in brackets has been calculated without taking the Glycines into account.

[c]The average amount (percentage) of parallel and anti-parallel strands in a structure. The sum of parallel and antiparallel strands in a structure does not always total 100% if the structure sees strands in an undefined orientation, i.e. attached by only one hydrogen bond.

[d]The average amount of strands in-register and out-of-register (by one residue).

doi:10.1371/journal.pcbi.1002051.t003

Table 4.III: **Structural characteristics for the 20-mer aggregates as a function of temperature below the melting point.** Temperatures above 280.1 K are not displayed here since they are populated essentially by conformations with random coil monomers with no secondary structure. The percentages are calculated over all the structures obtained in the last 200 ns of both OPEP REMD simulations. For details on (a)-(e), see Table 4.IV.II .

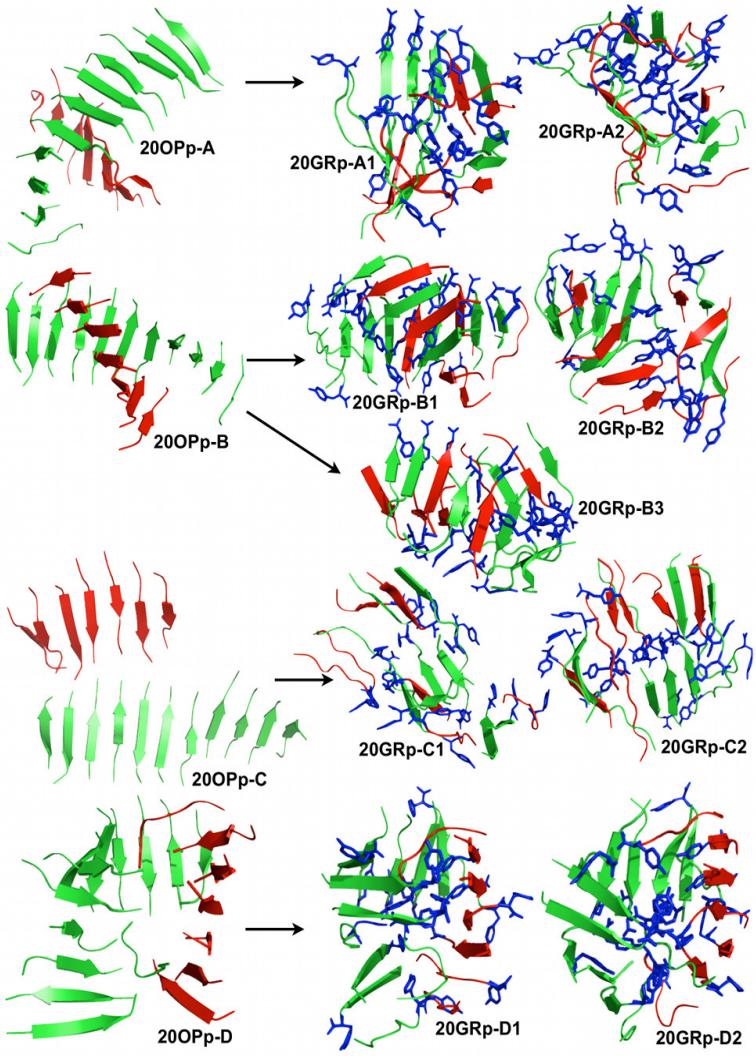

Figure 4.6: **Structures obtained for the 20-mer preliminary simulations.** The stable 20-mer structures obtained from OPEP's preliminary simulation 20OPp are shown on the left-handside panel. The final primary clusters obtained from the OPEP structures with all-atom MD or all-atom REMD are displayed on the right-handside panel. 20OPp-A,-B,-C and -D were extracted at 283.4 K. The color code is the same as in Fig. 4.5. 20OPp-A is composed of 2 perpendicular β-sheets. 20OPp-B is a twisted β-sandwich fibril-like structure. 20OPp-C is made of 2 sheets on top of one another. 20OPp-D consists of a folded sheet (green) facing another shorter sheet (red). During the all-atom MD simulation the structures tend to be more globular with the strands seeing some exchange between the β-sheets, *i.e.* the red and green β-sheets from the OPEP structures dissociate and re-associate during the all-atom MD simulations except for structures 20GRp-D1 and -D2.

fully below this temperature: the melting temperature is likely to continue to increase with the simulation length as the average nucleation time for the density used here appears to be around 1μs based on the fact that slightly more than half the trajectories have not yet visited ordered structures during the 400-ns simulation. In spite of this limitation, we observe significant exchange among the trajectories below melting, suggesting that these achieve some degree of thermodynamic equilibrium.

As for the 12-mer, aggregation is extremely favorable energetically. The melting temperature for 20OP1 varies between 280.4 K and 289.2 K during the last 200 ns of simulation and the energy of ordered structures at the lowest temperature, 223.8 K, is on average -27.8 kcal/mol/monomer for 20OP1, as calculated from the PTWHAM analysis. For 20OP2, the transition is happening between 260.2 K and 290.5 K and the potential energy of aggregated structures at the lowest temperature, 223.8 K, is on average -28.1 kcal/mol, which is comparable to the energies of aggregated structures for 20OP1. Those energies are about 10 kcal/mol/monomer above the dodecamer structures' energies at 222.5 K: clearly, the structures generated for the 20-mer are not as ordered as those found for the 12-mer due to the much longer time needed to sample these energetically-favorable conformations, but also because the entropic loss associated with full-ordering is larger for the 20-mer. For both the 20OP1 and the 20OP2 simulation sets, random coil structures dominate at simulations whose temperature is above 280 K.

Following specific trajectories, as they move through temperatures, it is possible to identify sequences of steps leading to low-energy ordered structures. In the more than 25 such events observed in 20OP1 and 20OP2, the aggregation process is systematically triggered by the formation of a few dimers, trimers and/or tetramers seeds. The conformations obtained from both 20OP1 and 20OP2 are structurally similar in the sense that they are almost always composed of three sheets composed of 5 to 9 strands each either facing each other in a triangle-like or organized in a propeller-like or β-sandwich conformation (Fig. 4.8). Irrespective of the final shape, the system displays a strong tendency to form β-sheets. The five final ordered structures selected from 20OP2 and shown in Fig. 4.8 are representative of all three REMD simulation sets: below melting, the 20-

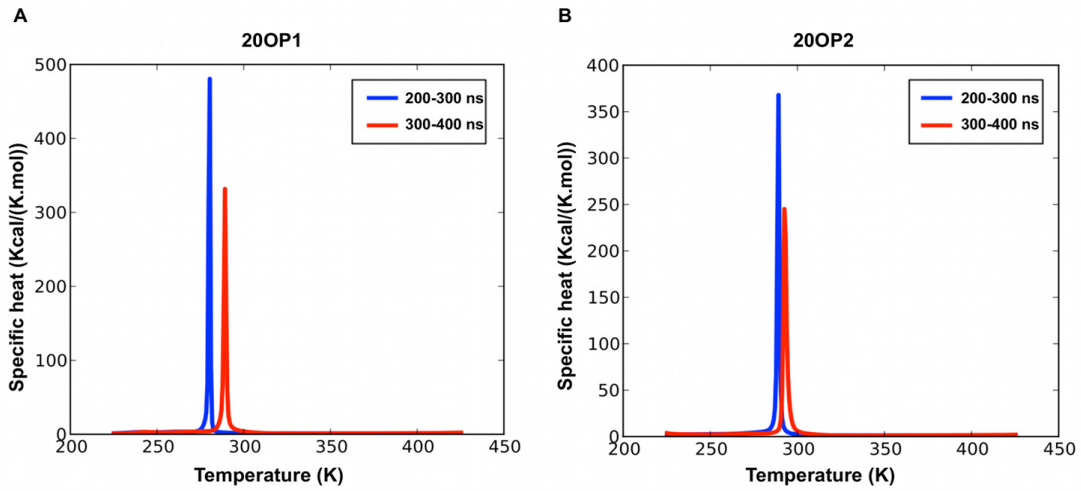

Figure 4.7: **Specific heat as a function of temperature for the two 20-mer simulations sets.** The specific heat is calculated over two time intervals for the systems 20OP1 (left panel) and 20OP2 (right panel) during the last 200 ns.

chain system mostly forms three β-sheets, but can also form two-sheet structures. Looking at the statistics collected for 20OP1 and 20OP2 (Table 4.III), we observe that various three-β-sheet configurations with juxtaposed β-sheets containing 8-7-5 monomers or 9-6-5 monomers are frequent below T_m. Two-β-sheet systems appear to be less frequent but are populated close to the melting temperature as well. Although, sheet lengths differ slightly between simulations 20OP1 and 20OP2 for the dominant structures, the overall results are consistent. We also observe a high number of possible β-sandwich morphologies for the three-sheet configuration among the final structures obtained where some of them see their three sheets facing each other in a triangle or twisted-around-each-other arrangements (Fig. 4.8, structures 20OP2-A, -B, -C and -E). These topologies run from a rather well defined 3-fold symmetry (20OP2-E) to more disordered conformations with little symmetry (20OP2-A, -B and -C). The minimal β-sheet unit contains four strands in 20OP2-E, five strands in 20OP2-A,-B and six strands in 20OP2-C. We note that the structure 20OP2-E is reminiscent of the recently proposed structure of Aβ1-40 fibrils with a three-fold symmetry [5]. Interestingly, ordered two-sheet conformations with one sheet slightly longer than the other, such as 11 9 and 12 8 (not shown, but close

to 20OP2-N, a 8 10 2 configuration in Fig. 4.8), represent a significant fraction of the accessible states either as a β-sandwich –with an occasional insertion of a short β-sheet – or as two perpendicular sheets.

In both REMDs, the β-sheets have a high probability of being in a mixed anti-parallel/parallel orientation state due to their length (over 70 % below T_m) (Table 4.III). As for the dodecamer, as T rises, sheets become shorter and we notice a rise of fully parallel sheets at the five temperatures above T_m (data not shown). Contrary to the 3-mer and 12-mer where the β-strands prefer an anti-parallel orientation below 300 K, the parallel orientation of the strands is preferred in the 20-mer at all temperatures. The extended chains are also dominantly perfectly aligned at the lowest temperatures and the structures see a mix of perfectly aligned strands and strands misaligned by one residue close to the melting point. The dominance of perfectly aligned strands and parallel orientations of the chains is consistent with the experimental observations that GNNQQNY fibrils display parallel β-strands [32].

4.7.3.2 All-atom MD simulations of dominant clusters generated by OPEP-REMD 20OPp

Consistent with the protocols described above for the 3-mers and 12-mers, the dynamical properties of 9 selected oligomeric conformations generated by the OPEP simulations were refined by all-atom MD simulations in explicit solvent (see Table 4.I).

The first set of all-atom MD simulations was run on the structures selected from 20OPp calculations. OPEP runs identified two main types of 3D organization for the 20mer: extended β-sheet and globular-like structures. The former are characterized by the presence of two parallel sheets, while the latter are characterized by a circular organization of the strands, in a mostly parallel arrangement. The major representatives of the extended β-sheet like structures obtained from OPEP are labeled 20OPp-B and 20OPp-C; the globular structures are recapitulated by 20OPp-A (which shows a compact part packed by a more extended sheet) and 20OPp-D, see Fig. 4.6. Two sets of 100-ns MD simulations, starting with different initial velocities, were run for each structure.

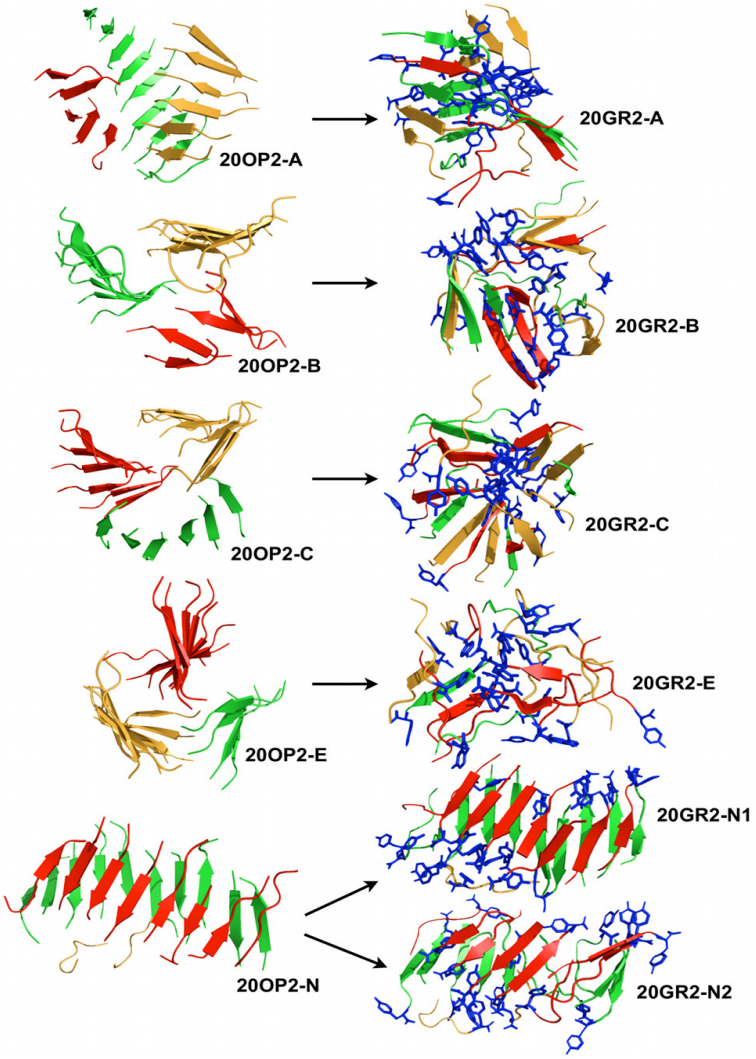

Figure 4.8: **Comparison of the structural properties between OPEP and GROMACS structures for selected stable structures.** The final primary clusters obtained from the OPEP structures with all-atom MD or all-atom REMD are displayed on the right-handside panel. 20OP2-A,-B,-C,-E and -N were extracted respectively at 260.1 K, 249.2 K, 254.7 K, 265.9 K and 292.2 K. The different sheets are distinguished by either a green, red or yellow color and the tyrosines are shown in blue sticks for the all-atom structures. Structures 20OP2-A,-B,-C and -E are composed of 3 sheets twisted around each other while structure 20OP2-N is a 2-sheet fibril-like conformation. During the all-atom MD simulation the structures tend to be more globular with the strands seeing some exchange between the β-sheets, *i.e.* the red, green and yellow β-sheets from the OPEP structures dissociate and re-associate during the all-atom MD simulations except for the fibril-like structures 20GR2-N1 and -N2.

			Population							
			223.8 K	249.3 K	260.1 K	265.9 K	270.3 K	273.8 K	277.1 K	280.1 K
20OP1	Configuration types (%)[a]	8 7 5	93.2	12.4	7.5	7.1	9.8	6.8	7.1	9.4
		8 7 3 2	0.0	44.7	2.6	0.8	0.4	0.8	0.0	0.0
		7 7 6	2.6	12.4	13.9	9.8	12.8	10.5	12.4	7.9
		10 6 4	0.4	0.0	7.5	8.7	5.3	11.3	5.3	3.4
		11 5 4	0.0	0.8	7.5	6.4	6.0	3.4	1.5	0.0
		11 5 2 2	0.0	10.5	0.0	0.0	0.0	0.0	0.0	0.0
		14 6	0.0	0.8	5.3	6.8	7.1	7.5	3.4	4.9
		15 5	0.0	0.0	3.4	7.1	3.8	4.5	6.8	8.3
		10 10	0.0	0.4	4.5	8.7	6.0	5.3	7.9	2.3
		10 5 5	0.0	0.0	1.9	3.8	8.3	7.1	3.8	2.6
		11 9	0.0	0.0	3.8	9.0	11.7	6.8	7.1	4.5
		20	0.0	1.1	3.4	4.9	5.3	5.3	3.4	0.0
	% parallel[b]		73.9	48.8	53.4	55.1	56.2	56.8	58.4	46.8
	% antiparallel[b]		25.1	48.2	41.3	37.6	36.7	35.6	34.5	29.4
	% fully parallel sheets[c]		26.2	22.4	19.7	15.1	11.8	14.5	11.1	16.1
	% fully antiparallel sheets[c]		0.9	7.9	3.1	4.6	1.9	2.5	1.0	6.5
	% mixed sheets[c]		72.9	69.7	77.2	80.3	86.3	83	87.9	77.4
	β-sheet content (%)[d]		59.1	61.3	60.5	60.5	60.5	59.8	57.1	45.7
	% Strands in-register/out-of-register by 1 residue[e]		32.1/38.0	43.1/31.2	42.6/33.8	41.2/36.9	40.1/38.4	38.8/38.6	39.7/38.6	38.7/39.8
20OP2	Configuration types (%)[a]	9 6 5	29.3	29.3	24.4	24.1	18.4	15.4	8.7	8.3
		9 7 4	0.4	1.5	4.5	4.5	7.5	4.9	6.0	3.8
		8 8 4	1.1	0.0	1.9	1.1	3.8	6.0	4.5	4.1
		7 7 6	4.5	6.0	8.7	12.0	12.8	16.2	15.0	12.0
		15 5	12.4	9.4	7.5	2.3	6.8	3.4	1.9	1.5
		13 7	0.8	0.4	0.4	1.9	1.5	2.3	4.9	7.1
		14 6	3.4	3.8	3.0	4.9	10.9	6.4	8.7	6.4
		11 9	9.4	14.7	6.8	3.8	2.3	4.5	5.6	3.8
		11 5 4	16.5	12.4	3.8	0.0	0.4	1.5	0.0	0.0
		8 7 5	0.4	1.9	1.5	1.9	5.6	6.8	6.4	6.0
		10 10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		20	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.4
	% parallel[b]		53.1	56.6	57.1	56.8	56.7	55.7	56.7	55.1
	% antiparallel[b]		38.1	35.7	36.5	37.5	37.9	39.3	38.4	38.5
	% fully parallel sheets[c]		7.4	16.9	15.0	14.2	14.1	14.1	18.7	17.1
	% fully antiparallel sheets[c]		0.6	4.7	3.2	4.2	3.9	4.4	5.1	8.1
	% mixed sheets[c]		92.0	78.5	81.8	81.6	82.0	81.5	76.2	74.8
	β-sheet content (%)[d]		66.5	63.4	63.0	62.5	61.5	60.6	58.7	56.0
	% Strands in-register/out-of-register by 1 residue[e]		51.0/36.3	50.0/34.0	48.1/35.1	44.6/37.0	42.4/37.2	42.3/37.2	39.9/39.1	39.7/39.4

Temperatures above 280.1 K are not displayed here since they are populated essentially by conformations with random coil monomers with no secondary structure. The percentages are calculated over all the structures obtained in the last 200 ns of both OPEP REMD simulations. For details on [a]–[e], see Table 2.
doi:10.1371/journal.pcbi.1002051.t004

Table 4.IV: **Structural characteristics for the 20-mer aggregates as a function of temperature below the melting point.** "First Cluster" means the most representative structure of the GROMACS simulations. "Final structure" is the final conformation obtained at the end of the GROMACS simulations. "CG" is the structure extracted at the end of the OPEP simulations before the reconstruction of the side chains. "Min" indicates the structure resulting from the reconstruction of the side chains after a minimization step. (a) The configuration type (as described in the OPEP Analysis and Structure Selection section) (b) The average amount (percentage) of residues in a β conformation. For OPEP, the percentage in brackets has been calculated without taking the Glycines into account. (c) The average amount (percentage) of parallel and anti-parallel strands in a structure. The sum of parallel and antiparallel strands in a structure does not always total 100 % if the structure sees strands in an undefined orientation, *i.e.* attached by only one hydrogen bond. (d) The average amount of strands in-register and out-of-register (by one residue)

4.7.3.2.1 Aggregates with facing β-sheets The structure of 20OPp-B is characterized by the presence of two perpendicular β-sheets. Each β-sheet consists mainly of twisted parallel strands.

After the first MD run (20GRp-B1) the two sheets are oriented anti-parallel to each other, forming a tight and elongated structure (Fig. 4.6). The Asn side chains from opposite sheets occupy the inter-sheet space with some intertwining of the amide side chains. Significant packing is also provided by the aromatic rings of Tyrosines belonging to adjacent strands in the same sheet. The 100-ns generated aggregate also displays a significant degree of twisting in the strands that make up the two facing antiparallel β-sheets. This final structure is similar to that observed by others [22]. It is important to underline that this twisted configuration forms spontaneously during the MD simulation time starting from a less compact structure.

In the second MD simulation (20GRp-B2), starting from the same initial structure with a different set of velocities, the β-sheet content decreases due a lower degree of packing of the Tyr side chains and interdigitation of the Gln and Asn side chains (Fig. 4.9). Packing interactions, however, still appear to be important in stabilizing the compact structure. During both 20GRp-B1 and 20GRp-B2 MD simulations, the strands are dynamically interchanged between the two β-sheets.

Two elongated sheets may, however, evolve towards very different supramolecular organizations. In 20OPp-C two elongated β-sheets, with mainly parallel β-strands, are in contact through the terminal Tyr aromatic chains in an extended and non-compact structure (Fig. 4.6). In both all-atom MD simulations, 20GRp-C1 and 20GRp-C2, the two β-sheets break up and reorganize in more compact conformations. Though, for 20GRp-C2, the strands do not interchange between sheets while 20GRp-C1 sees some mixing between the strands of the two β-sheets. The compact final structures still show some of the features we have described in the previous case: anti-parallel orientations of facing sheets, compaction of the side chains. One part of the long parallel β-sheet, represented by a four-stranded unit, detaches from the initial complex and packs onto the compact structure described above, minimizing the exposed surface area. The resulting structures correspond to a β-sandwich composed of three sheets. The time evolution of

the radius of gyration is reported in Figs. 4.12 and 4.13.

4.7.3.2.2 Evolution of globular structures All-atom MD simulations were also started from more densely packed structures obtained from OPEP runs, represented by 20OPp-D and 20OPp-A. Each starting structure was subjected to two 100-ns MD simulations, resulting in simulations 20GRp-A1, 20GRp-A2, 20GRp-D1 and 20GRp-D2 (Fig. 4.6). In the case of 20OPp-D both all-atom MD runs show that the starting conformation retains its overall structure with mostly parallel β-strands. Most of the Tyr side chains form a compact hydrophobic core at the interior of the structure. During all-atom MD evolution they reorganize and form small compact clusters that are optimally packed through π-stacking face-to-face interactions, minimizing unfavorable contacts with the solvent. In general, the interior of the aggregate is compact and hydrophobic while the external surface is decorated with the more hydrophilic amide side chains. From the biochemical point of view, this structure could be described as a "soluble" oligomeric state reminiscent of the ones observed for the 12-mer run.

Starting from 20OPp-A, simulation 20GRp-A1 evolves towards a compact globular structure, in which parts of the ordered β-sheets are lost and strands are dynamically interchanged between sheets. The aromatic Tyr side chains cluster in the hydrophobic core of the structure and Asn side chains align on the surface. Most of the conserved sheets are still in parallel orientation.

Strikingly, in simulation 2 (20GRp-A2) the initial structure evolves to form two twisted antiparallel sheets in which the constitutive strands are parallel to each other. This structure resembles the twisted conformation observed for 20OPp-B and was observed in previous simulations [22]. The inter-sheet space is filled by Asn and Gln side chains, which in some case interdigitate (Fig. 4.9). The conformation of this intermediate still shows some Tyr aromatic side chains in the interior, disrupting the optimal interdigitation of amidic side chains and causing the structure to be non-perfectly ordered (Fig. 4.6).

These results suggest that the sheet organization in the twisted antiparallel conformation(s) may be accessible on the aggregation pathway, once two sheets are formed

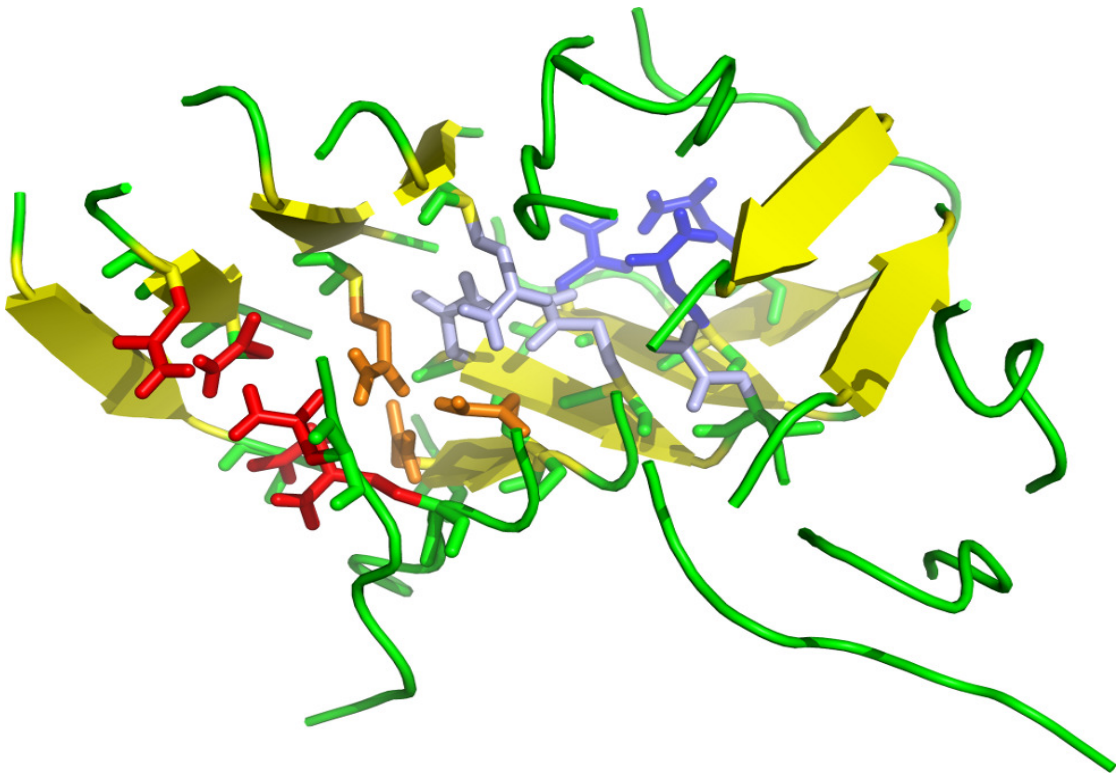

Figure 4.9: **Structure 20-mer showing interdigitation of the side chains.** This structure is representative of the 20GRp-B2 simulation. It shows the inter-sheet space filled by Asn and Gln side chains, which in some case interdigitate.

and docked upon each other. Interestingly, we have observed the formation of elongated, twisted antiparallel structures in MD only in the 20-mer system. The latter appear to evolve preferentially towards globular structures, suggesting that elongated, fibril-like conformations of the oligomers may be accessible only in the presence of a higher number of monomers. At the atomic level, sheet-locking is favored by the packing of Asn and Tyr side chains. The Tyr aromatic packing and the initial formation of steric-zipper-like structures also provide important contributions in determining the ordering and stabilization of the growing aggregate and, possibly, its evolution to a stable fibril.

4.7.3.3 All-atom MD simulations of the dominant clusters generated by OPEP-REMD 20OP2

The detailed role of side chains in determining the conformational characteristics of compact aggregate structures was further evaluated by analyzing at atomic resolution a set of diverse OPEP structures: 20OP2-A, 20OP2-B, 20OP2-C, 20OP2-E and 20OP2-N (Fig. 4.8).

The starting structures of the MD simulations from 20OP2-A, 20OP2-C and 20OP2-E of the aggregates all consist of three extended β-sheets, organized in different tertiary arrangements (see Fig. 4.8). The all-atom evolution of structure 20OP2-A leads to a more compact and globular-like conformation with a mixing of the β-strands and a partial loss of ordered β-structure involving external strands (Fig. 4.8, structure 20GR2-A). Tyr side chains mainly pack in the interior of the globule in the representative structures of MD simulations while most of the hydrophilic side chains (Asn and Gln) are located on the exterior of the oligomer pointing towards the solvent. This arrangement would confer water solubility to the oligomer by sequestering hydrophobic-aromatic side chains to the interior of the aggregate and decorating the external surface with hydrophilic groups.

In the case of 20OP2-C, the starting structure constituted by three β-sheets, which are lined and twisted along a common axis, is not stable in the all-atom MD setting, and immediately evolves to a more compact globular structure that however does not display specific supramolecular properties or preferential orientations of the strands within the

aggregate (Fig. 4.8, structure 20GR2-C).

In the case of 20OP2-E, the evolution of all-atom MD simulations at 300 K determines a large decrease in the degree of ordered β-structure leading to the formation of disordered, amorphous conformations (Fig. 4.8, structure 20GR2-E).

The remaining two representative clusters obtained from OPEP simulations display different three-dimensional arrangements. In the case of 20OP2-B, the structure is characterized by parallel β-sheet motifs that form a less compact conformation than the one observed above. All-atom MD evolution leads to a globular structure with a global reorganization of the β-strands (Fig. 4.8, structure 20GR2-B). A large number of Asn and Gln amidic side chains point towards the interior of the globular structure, and a number of hydrophobic Tyr side chains are aligned in contact with the solvent. The reorganization of the starting structure to this globular intermediate determines a partial loss of ordered secondary structure in some of the constituent peptides. This structure retains a large hydrophobic surface in contact with the solvent. The large hydrophobic area exposed to water may be one of the causes of the insoluble character of these intermediates.

Finally, we simulated the structure of cluster 20OP2-N at all-atom resolution as this aggregate forms an elongated structure with two facing β-sheets. MD evolution at 300 K for this system shows no reorganization of the β-strands (Fig. 4.8, structure 20GR2-N1). In contrast to what we observed for 20OP2-B, the Asn and Gln amidic side chains fill the space between the two sheets, establishing hydrogen bonding interactions, and showing interdigitation of side chains reminiscent of the experimentally observed dry steric zipper. The sheets are not all perfectly parallel, and this might oppose the formation of a perfectly packed steric zipper.

Summarizing, as shown in Table 4.IV, the structural properties of the aggregates cannot be easily rationalized into specific classes: great variability is observed in terms of β-sheet content, percentage of parallel vs. antiparallel arrangements, register and relative orientations of the strands. Interestingly, it appears that the fibril-like structure 20GR2-N1 is stable during the 100-ns all-atom MD simulation, while the other arrangements undergo large reorganization. In general, the consistency of the results in terms of conformational plasticity for all constructs clearly indicates that a wide range of different

structures is accessible to relatively large sized oligomers.

4.7.3.4 All-atom REMD simulations of two elongated 20-mer structures generated by OPEP-REMD

The all-atom simulations of 100 ns starting from 20OPp-B and 20OP2-N showed the possibility for the aggregates to remain elongated partially ordered oligomers whose structures are reminiscent of the arrangements observed by X-rays of micro-crystals. In order to gain more insights into the stability and conformational evolution properties of these structures, we set out to run all-atom REMD simulations starting from OPEP structures 20OPp-B (Fig. 4.6) and 20OP2-N (Fig. 4.8). The resulting all-atom REMD simulations are labeled 20GRp-B3 and 20GR2-N2, respectively.

In the REMD simulation labeled 20GRp-B3, and similarly to what is seen in OPEP REMDs, we observe that structures interconvert between compact and elongated conformations with a pair of sheets facing each other. The main representative structures for simulation 20GRp-B3, and their relative stabilities, are reported in Fig. 4.11(a). In the elongated conformation, the relative orientation of the strands within the sheets tends to be parallel. The two facing strands are oriented antiparallel to each other. The Tyr aromatic side chains form clusters of packed rings that are reminiscent of the arrangements observed in the crystals from the Eisenberg group (Fig. 4.10(a)). Partial ordering of the Gln and Asn side chains into the zipper spine arrangement is also observed. This conformation, labeled as conformation 3 in Fig. 4.10(a), is however not stable enough to be the most populated structure at the lowest temperatures. Representative structures of the most representative clusters are extracted from the trajectories and their relative free energies evaluated with the GB/SA approach implemented in the program MacroModel, and according to what was already reported [17]. These calculations provide an approximated energetic value for the stability of the aggregates in solution, and show that the more elongated structures tend to undergo transitions to more globular like conformations.

Several conformational transitions among different structural clusters are observed, and highlighted by direction arrows in Fig. 4.10(a). Summarizing, the starting confor-

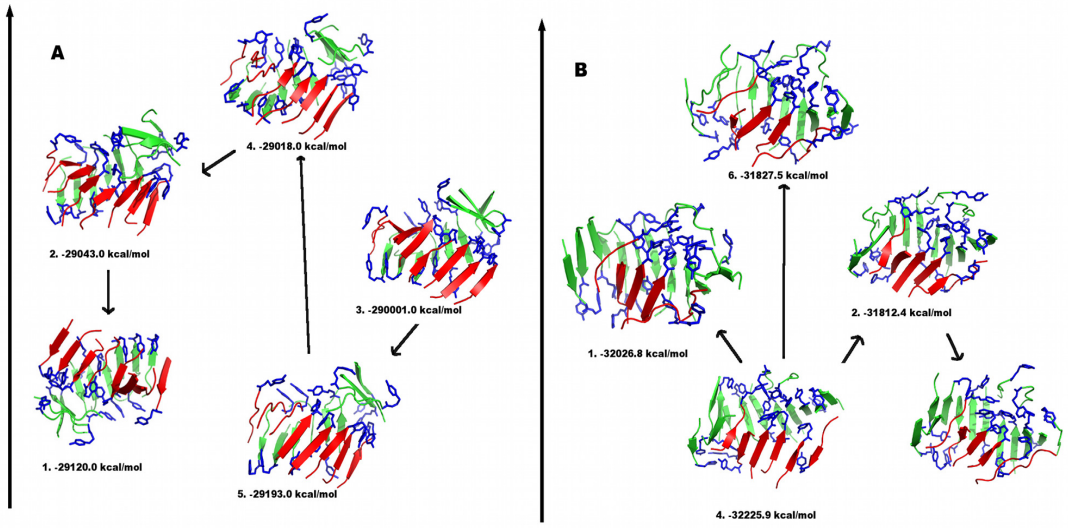

Figure 4.10: **Representative structures of the most populated clusters from all-atom REMD simulations of the 20-mer:** (a) structure 20GRp-B3 and (b) structure 20OP2-N. The number identifying each structure represents the cluster rank (1. being the most populated cluster). The value of the GB/SA energy in water of the complex is reported. Arrows represent transitions between clusters, indicating possible paths between cluster structures.

mation for 20GRp-B3 with two facing sheets antiparallel to each other shows a high tendency to interconvert rapidly into globular, compact structures. The high conformational plasticity and limited stability of elongated conformations had already been noticed in the present OPEP-based simulations. Analysis of the time evolution of REMD trajectories shows that globular conformations form by detachment and re-docking of β-sheet motifs. During the formation of the more globular structures a slight decrease of the amount of ordered β-sheet could also be noticed.

All-atom REMD simulations were also used to analyze the structural evolution of cluster 20OP2-N. The representative structures obtained from the all-atom simulation 20GR2-N2 are reported in Fig. 4.10(b), with their relative stabilities. The packing of the interior of the initial aggregate is not optimal. Consistently with the previous case, the overall aggregate still shows a high degree of structural plasticity undergoing transitions to conformations with a lower degree of ordered β-sheets and more globular-like shapes. Elongated structures featuring two facing β-strands still appear and show partial interdigitation of Asn and Gln side chains in the inter-strand space, with formation of a steric zipper motif. This does not seem sufficient, however, to provide the necessary stabilization for the aggregate to evolve to a protofibril structure. Once more, a number of transitions among different structural families (Fig. 4.10(b)) are observed.

Overall, we observe that the elongated structures reminiscent of the one observed by the Eisenberg group in the fibril microcrystals is accessible at room temperature but is not stable and can easily interconvert into globular or more disordered conformations, even in the presence of 20 monomers (see Table 4.IV). According to our observations, this happens both by detachment-reattachment of strands or small oligomers from existing structures, or through sliding/reptation moves without detaching from the aggregates [86, 180]. This behavior is widely observed across our 20-mer simulations while the 12-mer system does not undergo much reorganization during all-atom MD simulations. Interestingly, when an ordered supramolecular structure forms, with either an elongated or a compact globular shape, there is no space for water in the interior of the aggregate. Water exclusion from the core of the aggregate may represent the first step leading to the formation of a dry interface.

4.8 Conclusions

The self-organization process of peptides and proteins into oligomeric soluble and insoluble aggregates that eventually evolve to fibrils is still difficult, if not impossible, to study at atomic resolution using experimental approaches. In this paper, we have carried out an extensive and comprehensive study of the formation of oligomers of the model peptide GNNQQNY in different conditions combining coarse-grained and all-atom simulation approaches. Different numbers of peptides were used in several simulations. In the smallest systems, composed of three strands, a diverse set of structural motifs is accessible at room temperatures. When bigger systems consisting of 12 chains are analyzed, compact and globular structures begin to appear. Interestingly, in some cases, globular oligomers expose hydrophilic side chains to the contact with the water solvent, providing a viable model for soluble intermediates that have been observed on pathway to the formation of the final fibril. In parallel, at 300 K, globular structures exposing a large amount of hydrophobic surface also appear. These may represent possible nuclei for the growth of bigger supramolecular structures.

In simulations conducted using 20 monomers, we have noticed the appearance of elongated structures characterized by the juxtaposition of two mainly parallel β-sheets with partial interdigitation of amidic side chains reminiscent of the zipper-spine observed in fibril microcrystals. It is important to notice, however, that these structures are not stable in water solvent and evolve towards more globular conformations. This observation suggests that while ordered fibril-like structures are accessible on the energy landscape, they need further stabilization by establishing contacts with multiple copies of similar structures in order to evolve to a fully fibrillar geometry. In this context, the formation of this geometry would require the constructive interplay of many factors and the entropic expense of such process would be clearly very high, explaining the long lag phase times and very slow kinetics of amyloid fibril formation. Moreover, the rich variety of structures and conformational changes observed for the aggregates may also reverberate into the fibril polymorphism observed at the experimental scale.

In summary, our data and structural models represent valid complements to experi-

mental approaches in the attempt to shed light on the supramolecular arrangements of amyloidogenic oligomers, and lead the following conclusions.

First, the 20-mers of GNNQQNY are polymorphic and endowed with a high degree of structural plasticity. Polymorphism of the fibrillar products of amyloid aggregation has been observed for many sequences by X-ray diffraction and solid-state NMR experiments [3, 5, 47] and a recent computational study using a simplified description, where the peptide has a single degree of freedom with two minima [181]. Polymorphism of the low-n oligomers with different registers of inter-peptide hydrogen bonds and orientations of the chains has also been reported using various force fields and computational procedures [17, 19, 147, 158, 182].

Second, the 20-mers of GNNQQNY in explicit water are in dynamic equilibrium, within at least 100 ns, between amorphous structures (high probability) and configurations with three β-sheets in various orientations (medium probability) and two β-sheets (low probability). These two-β sheets, reminiscent of the cross-β structures and the dry steric zipper observed experimentally for mature fibrils, are not parallel, however, suggesting the existence of a free energy barrier preventing the formation of a perfectly packed steric zipper.

Third, there is a reorientation of the β-strands between the GNNQQNY oligomers and fibrils. We find that an anti-parallel β-strand alignment dominates over the parallel one in the 3- and 12-peptide systems. This contradiction with the fibrillar parallel β-strand orientation [32] is however reconciled by the 20-peptide systems, where a significant increase in the amount of parallel strands and fraction of fully parallel β-sheets is observed.

Fourth, a common observation is that short amyloid peptide fragments assume antiparallel β-strand geometries whereas longer peptides, and proteins, often assume parallel geometries. Our simulations along with other recent studies show this geometrical property is more complex and depends strongly on the amino acid composition. The dependence of β-strand orientation with oligomer size occurs in the GNNQQNY (Sup35) peptide and the VQIVYK (PH6) peptide, as reported by another computational study [183], because both short peptide fragments do not contain any opposite charged

amino acids at the extremities as opposed to numerous fragments, such as the KLVFFAE (Aβ(16-22)) and the KFFE peptides. Table 4.II, for example, shows that the free energy difference in favor of the antiparallel sheet decreases from 0.70 kcal/mol for the trimer to only 0.05 kcal/mol for the dodecamer at 300 K. This competition between the two β-strand orientations during polymerization is also supported experimentally by the co-existence of either antiparallel or parallel strands in seeded hIAPP20-29 (SNNFGAILSS) fibrils [143], and the fact that D23N-Aβ40 forms fibrils with a majority having antiparallel β-sheet structures and a minority having parallel β-sheet structures [184]. We hypothesize that both short polar peptides and longer peptides could display this β-strand orientation transition during aggregation.

In addition, antiparallel β-sheets allow a higher potential variability of the inter-chain hydrogen bond geometry [185, 186] accommodating a higher number of possible strand conformations, which is the favored situation in smaller aggregates. A parallel organization could be more favorable for larger aggregates, which can for instance create double-layered sheets that can pack with a parallel arrangement of the strands as indicated by SS-NMR [187]. Moreover, in larger aggregates, the parallel arrangement might be stabilized by the definition and stabilization of large numbers of side chain - side chain contacts. Therefore, our results are in agreement with experiments but also suggest a more complex relationship between the monomers during the assembly process, something that cannot be measured experimentally.

Fifth, in terms of experimental relevance, it is important to note that evidence exists showing that aggregation pathways can be manipulated by the use of molecular chaperones. In the case of the Sup35 prion protein, the chaperone Hsp104 catalyzes the polymerization of seeds that are crucial for efficient amyloid formation [188]. Different chaperones such as Hsp70 and Hsp40, on the contrary, prevent the assembly of aggregating species when added during the polymerization reactions. In these cases, soluble aggregates are formed, showing that chaperones can redirect amyloidogenic polypeptides into non-amyloidogenic species. Finally, the chaperonin TRiC has also been shown to be a modulator of amyloid formation [189]. TRiC, in combination with Hsp70 and 40, stimulates the reassembly of huntingtin oligomers into soluble species, which are non-

toxic. Clearly, the type of binding and structural remodeling determined by different chaperones is dictated by the details of the molecular recognition between the oligomers and the interaction surfaces of the chaperones. In this context, it is worth noting that generating plausible models of possible oligomeric substrates for the chaperones may be of great help in the design of optimized systems aimed at modulating aggregate properties. Based on the structural and surface properties of the oligomers, models of their complexes with chaperones may be generated. On this basis, the hydrophobic-hydrophilic profile of the chaperone interaction surfaces could, for instance, be changed by means of site-directed mutagenesis, affecting their activity and ultimately the properties of the remodeled oligomers. This would allow a rational manipulation of the amyloidogenic pathways, helping to shed light on a very complex biological phenomenon.

A final consideration helpful to put our results in a biological perspective is related to the importance of the knowledge of oligomeric structures in the design of amyloidogenic inhibitors. In this context, we are currently exploring the characterization of the solvent accessible hydrophobic surface area of the 20-mers to guide docking-experiments of small-molecule compounds (Congo Red and EGCG in particular), in order to derive possible rules for the rational selection of aggregation inhibitors. Preliminary data and results show that this could be helpful in alleviating the difficulties associated to drug-design when dealing with amyloid-targets. Indeed, compared to classical drug-design efforts where the target is an active site, with well-defined structure and cavities, the variety of structures, mechanisms and conformational plasticity of oligomers shown here confirm that rational design of aggregation inhibitors is a daunting challenge. However, careful characterization of oligomeric structures provides useful suggestions for the design of possible inhibitors. Selective compounds or peptidomimetics could be designed/selected to target the oligomer conformations characterized by the presence of aromatic groups on their external surface. These compounds would actually target intermediates that are more prone to be insoluble or to favor the addition of monomers through hydrophobic interactions. Interestingly, most of the existing inhibitors of amyloidogenic pathways are small molecules rich in aromatic functionalities, which can target more than one single aggregating species, showing a general mechanism of ac-

tion [190]. Alternatively, one could design peptidomimetic-based or small molecule chaperones that can stabilize soluble species, subtracting them from the amyloidogenic pathway. This would lead to a redirection of otherwise amyloidogenic peptides into non-amyloidogenic species.

4.9 Supporting Material

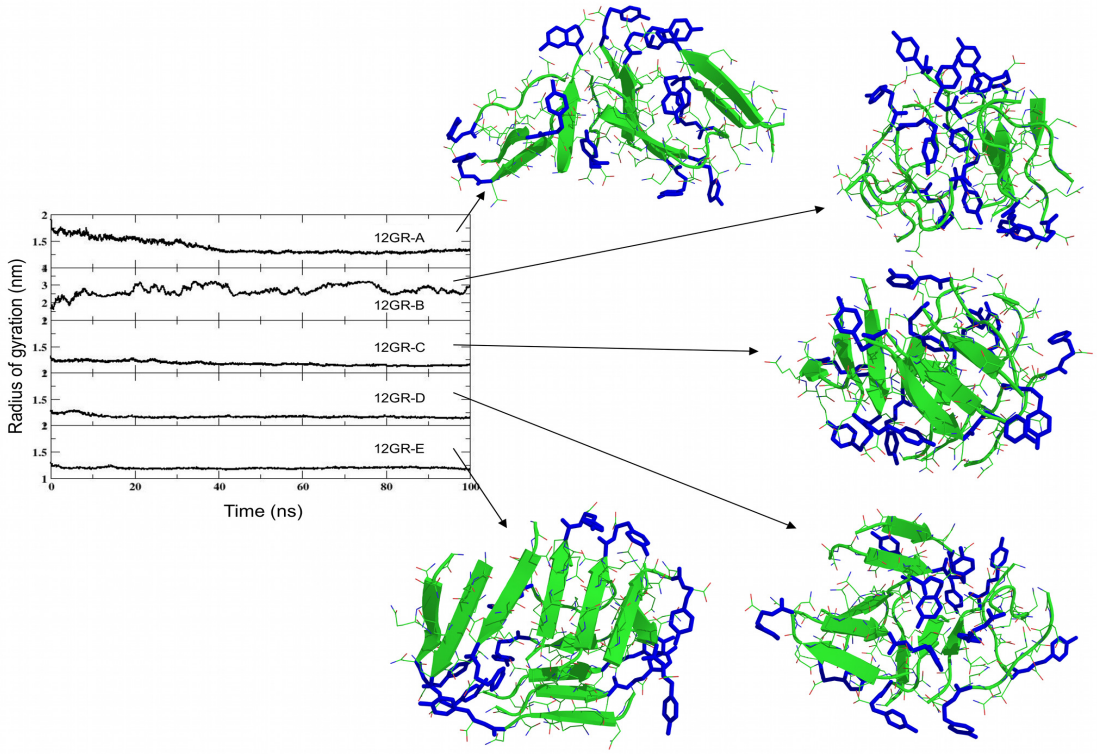

Figure 4.11: **Supporting Figure 1: Time evolution of the radius of gyration of the 12-mer oligomers.** From top to bottom: strctures 12GR1-A, 12GR1-B, 12GR1-C, 12GR1-D and 12GR1-E. The structures shown are the final structures of the all-atom MD simulations with GROMACS.

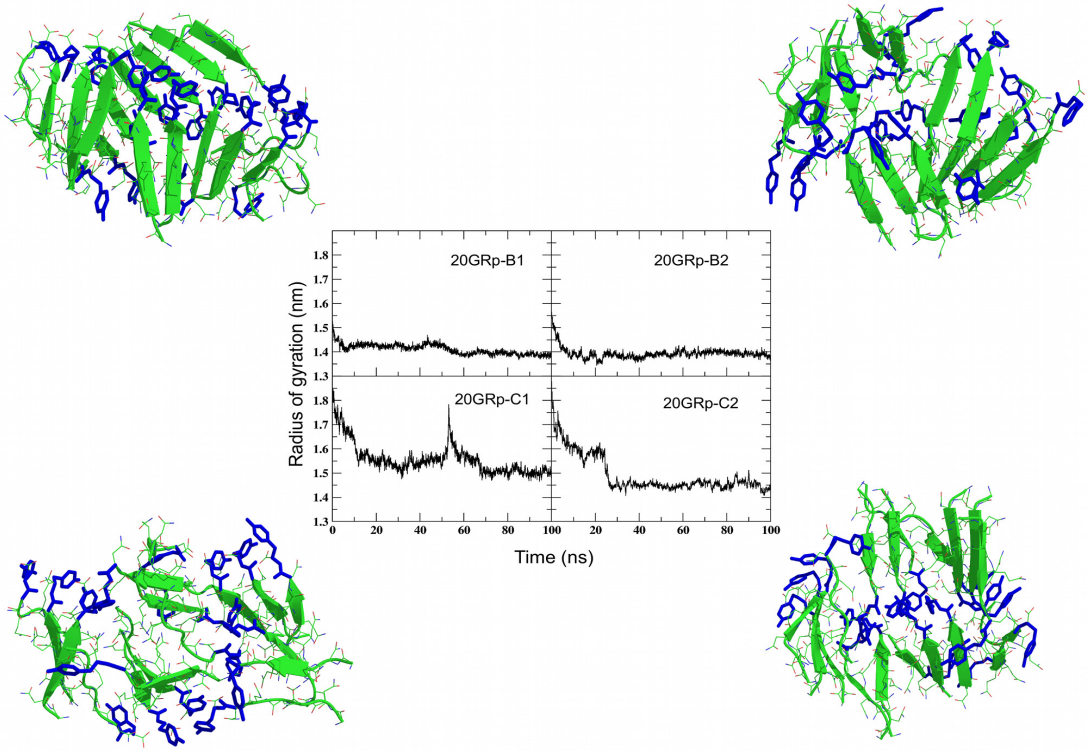

Figure 4.12: **Supporting Figure 2: Time evolution of the radius of gyration of the 20-mer oligomers for the preliminary simulation.** Structures 20GRp-B1, 20GRp-B2, 20GRp-C1 and 20GRp-C2. The structures shown are the final structures of the all-atom MD simulations with GROMACS.

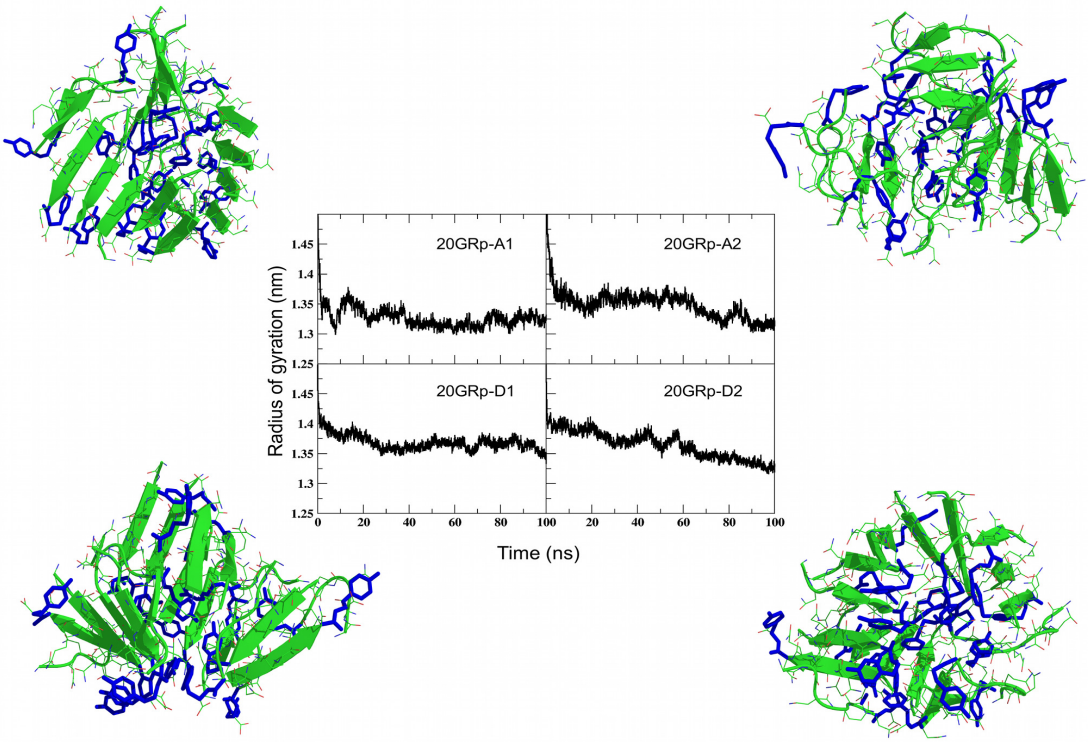

Figure 4.13: **Supporting Figure 3: Time evolution of the radius of gyration of the 20-mer oligomers for the preliminary simulation.** Structures 20GRp-A1, 20GRp-A2, 20GRp-D1 and 20GRp-D2. The structures shown are the final structures of the all-atom MD simulations with GROMACS.

CHAPITRE 5

KINETICS OF AMYLOID AGGREGATION: A STUDY OF THE GNNQQNY PRION SEQUENCE

5.1 Objectifs

Cet article vient compléter l'étude présentée dans le chapitre 3 et explore l'aspect cinétique du processus d'agrégation du système 20-mère du peptide GNNQQNY. Tel qu'exposé dans l'annexe 1, l'assemblage des protéines amyloïdes peut être vu comme un processus complexe de nucléation mais très peu de travaux ont pu en observer le détail des premières étapes à ce jour. Dans cette étude, acceptée pour publication dans le journal *PLoS Computational Biology*, les auteurs explorent les caractéristiques d'agrégation du 20-mère GNNQQNY – ainsi que les évènements microscopiques qui la régissent – et observent, en effet, un phénomène spontané de nucléation (tel que décrit par la théorie physique classique de la nucléation) complexifié par une reorganisation structurelle persistant après la fin de la croissance des oligomères. La compréhension détaillée de la cinétique d'agrégation des protéines vient grandement complémenter de nombreuses études expérimentales, souvent limitées par la résolution de leurs observations et donc limitées dans leur capacité de décrire de tels événements microscopiques.

5.2 Contributions des auteurs

- (Dans l'ordre d'apparition dans l'article) Jessica Nasica-Labouze (JNL) et Normand Mousseau (NM) ont conçu ensemble le plan d'action pour cette étude numérique.
- JNL a réalisé les simulations, l'extraction et l'analyse des données.
- JNL et NM ont contribué à l'élaboration des outils de simulations, d'extraction de données et d'analyse.
- JNL et NM ont participé à la rédaction de l'article.

5.3 Author Summary

Protein aggregation plays an important pathological role in numerous neurodegenerative diseases such as Alzheimer's, Parkinson's, Creutzfeldt-Jakob, the Prion disease and diabetes mellitus. In most cases, misfolded proteins are involved and aggregate irreversibly to form highly ordered insoluble macrostructures, called amyloid fibrils, which deposit in the brain. Studies have revealed that all proteins are capable of forming amyloid fibrils that all share common structural features and therefore aggregation mechanisms. The toxicity of amyloid aggregates is however not attributed to the fibrils themselves but rather to smaller more disordered aggregates, oligomers, forming parallel to or prior to fibrils. Understanding the assembly process of these amyloid oligomers is key to understanding their toxicity mechanism in order to devise a possible treatment strategy targeting these toxic aggregates. Our approach here is to computationally study the aggregation dynamics of a 20-mer of an amyloid peptide GNNQQNY from a prion protein. Our findings suggests that the assembly is a spontaneous process that can be described as a complex nucleation and growth mechanism and which can lead to two classes of morphologies for the aggregates, one of which resembles a protofibril-like structure. Such numerical studies are crucial to understanding the details of fast biological processes and complement well experimental studies.

5.4 Abstract

The small amyloid-forming GNNQQNY fragment of the prion sequence has been the subject of extensive experimental and numerical studies over the last few years. Using unbiased molecular dynamics with the OPEP coarse-grained potential, we focus here on the onset of aggregation in a 20-mer system. With a total of 16.9 μs of simulations at 280 K and 300 K, we show that the GNNQQNY aggregation follows classical nucleation theory (CNT) in that the number of monomers in the aggregate is a very reliable descriptor of aggregation. We find that the critical nucleus size in this finite-size system is between 4 and 5 monomers at 280 K and 5 and 6 at 300 K, in overall agreement with experiment. The kinetics of growth cannot be fully accounted for by the CNT, however.

For example, we observe considerable rearrangements after the nucleus is formed, as the system attempts to optimize its organization. We also clearly identify two large families of structures that are selected at the onset of aggregation demonstrating the presence of well-defined polymorphism, a signature of amyloid growth, already in the 20-mer aggregate.

5.5 Introduction

The aggregation of misfolded amyloid proteins into fibrils is a hallmark of many neurodegenerative diseases such as Alzheimer's, Parkinson's, and Huntington's diseases [25, 26, 136–138] and understanding amyloid aggregation mechanisms is crucial for controlling its destructive consequences. Fibrils are known to be ordered insoluble assemblies with a core cross-β structure. They are not the only aggregated species involved, however, and oligomers, smaller intermediates on or off the fibril formation pathway, have been found to be responsible for amyloid cytotoxicity [41, 190, 191]. Their role in amyloid aggregation is still a matter of debate but significant efforts have gone into better understanding and characterizing their structure and dynamics both experimentally [130, 134, 192] and computationally [193**?** –198]. Oligomers are often found to be precursors to amyloid fibrils. They could also, in some cases, appear as the product of a competition between the ordered fibrillar and amorphous globular morphologies, forming via different assembly pathways. This widespread characteristic of amyloid proteins is described as polymorphism [3, 42, 140] and is under kinetic control [181]. The presence of oligomers is therefore crucial for the fibrillisation process as well as the final morphology of fibrils [199] and understanding their kinetics of formation could be the key to controlling this polymorphism.

The aggregation of amyloid proteins is a highly cooperative self-assembly mechanism, which is often described as a complex nucleation and growth process [200]. The nucleation step, in a supersaturation environment, consists of a series of stochastic events leading to the formation of metastable seeds for the oligomer or fibril to grow on [201]. Nucleation kinetics display two characteristic properties: the presence of 1) a lag time

before aggregates can be detected and 2) a maximum growth rate after nucleation is triggered [107, 111]. Direct experimental observations of nucleation and growth have been reported [103–106] and nucleation was always found to be the rate-limiting step of amyloid formation [111].

The aim of the present work is to investigate the dynamics of amyloid aggregation and the forces driving self-assembly for the 20-mer system of the amyloidogenic GNNQQNY peptide using molecular dynamics (MD) and a coarse-grained potential (OPEP). The nucleation specificity of the N-terminal region (9-39) of the budding yeast prion protein Sup35, GNNQQNY, is well understood. This small heptapeptide alone drives the entire Sup35 protein to self-assemble into amyloid fibrils [202] and, when isolated, displays the same amyloid properties and aggregation kinetics as the full-length Sup35 protein [27]. In addition, its cross-β spine structure has been determined at the atomic level by X-ray crystallography [32]. It is therefore a very good candidate to the study of amyloid aggregation kinetics and numerous computations have been performed on the GNNQQNY sequence to characterize the onset of aggregation for this model [17–20, 203]. This work expands on our previous multi-scale thermodynamic study of different sizes of GNNQQNY systems, where we identified the morphologies accessible to the trimer, dodecamer and 20-mer. Now, we focus on the aggregation kinetics, using long MD simulations of unbiased spontaneous self-assembly. We offer a full analysis of the onset of aggregation for GNNQQNY peptides at a refined coarse-grained level. A total of 16.9 μs of simulations have been collected to allow statistically relevant analyses. Altogether, our results indicate the presence of a nucleated-polymerization process intertwined with oligomer-involving mechanisms, thus leading to a certain degree of polymorphism that is already clearly established for the 20-mer.

Materials and Methods

Following Ref. [64], which showed that the GNNQQNY amyloid aggregates generated with the coarse-grained OPEP forcefield [52] were reasonably preserved in long explicit solvent all-atom MD simulations, we revisit this system focusing, this time, on

the kinetics of the aggregation process.

5.5.1 System description

As in our previous study, we perform implicit solvent coarse-grained molecular dynamics (MD) simulations using the OPEP potential version 3.0 [52]. OPEP is designed for efficient protein folding and structure prediction of large systems over long timescales and is also accurate for studying thermodynamics [59]. In OPEP, all heavy backbone atoms are fully represented (N, H, C_α, C and O). Side chains, for their part, are reduced to a single bead with appropriate geometrical properties and van der Waals radius. The implicit effects of the solvent are included in the interaction parameters of the potential energy function, as detailed elsewhere [52, 53]. OPEP is a well tested potential and has been implemented with a palette of numerical methods such as Monte-Carlo [53–57], the activation-relaxation technique (ART nouveau) [67–72], MD [58–61] and REMD [62–66].

Here, two sets of single temperature MD are performed on a 20-mer of GNNQQNY, with both terminii of each peptide charged, in order to characterize in detail the kinetics of aggregation. The first set consists of a total of 152 100-ns-simulations (76 at 280 K and 76 at 300 K) with configurations saved every 5000 steps. The choice of temperatures is motivated by the fact that 280 K and 300 K are temperatures below and above the transition temperature previously found for the 20-mer of GNNQQNY. As explained below, the initial atomic positions taken for this set are extracted from the simulations reported in Ref. [64]. An additional 10 30-ns-simulations are then carried out from a subset of the starting atomic positions of the previous simulation set (5 at 280 K and 5 at 300 K) with configurations saved every 50 steps to better monitor the detailed evolution of the system during aggregation. All simulations are independent, starting with different random Boltzmann distributed velocities. In every case, we maintain simulation conditions as close as those of Ref. [64], with Berendsen thermostat for temperature control [95], an integration time step of 1.5 fs and periodic boundary conditions applied to a box 200 Å in size containing 20 monomers of GNNQQNY, which represents a constant 4.15 mM concentration.

For simplicity, the starting random structures for our simulations were extracted from the high-temperature set generated in our previous REMD OPEP runs of the GNNQQNY 20-mer [64]. A typical starting structure for our simulations is shown in Figure 1 with all 20 peptides isolated and in random coil conformations. At the start of each run, a minimization procedure is performed using a combination of the steepest descent algorithm and the conjugate gradient method [204], followed by a thermalization of 50 000 steps (0.075 ns) to ensure that all conformations are fully thermalized.

Because of the implicit solvent treatment as well as the peptide's coarse-grained representation, that decrease the number of degrees of freedom, the aggregation kinetics is accelerated. It is therefore not possible to establish a direct connection between the aggregation time observed in the simulation and in experiment. However, as shown in Ref. [59], the thermodynamical properties are, at least qualitatively, maintained. The simulations presented here, therefore, should provide the right qualitative picture for the first steps in the kinetics of aggregation.

5.5.2 Analysis

Most of the analysis on the nucleation and growth kinetics is carried out using a clustering tool [64] adapted to multiprotein assembly and designed to classify β-sheet clusters based on strand attachment. For the purpose of this work, this procedure can also handle the calculation of kinetic association and dissociation rates. To assess strand attachment, the criterion used to define and calculate hydrogen bonds between strands is similar to the DSSP definition [100]. A peptide belongs to a cluster if it is attached to another strand of that cluster by at least two hydrogen bonds. An additional criterion is applied on dihedrals ψ and ϕ angles to determine if a given strand in a cluster has enough amino acids in β-conformation. For each amino acid the ψ and ϕ angles are calculated and if they satisfy the region ψ(in degrees): [-180:-150;0:180], ϕ(in degrees): [-180:0;150:180] (corresponding approximately to the β region of the Ramachandran plot [75]), the amino acid is in a β state. A GNNQQNY peptide is considered in a β state if at least three of its residues are in the β region. If a peptide is not found to be in a β state, it is excluded from the cluster. This determination of secondary

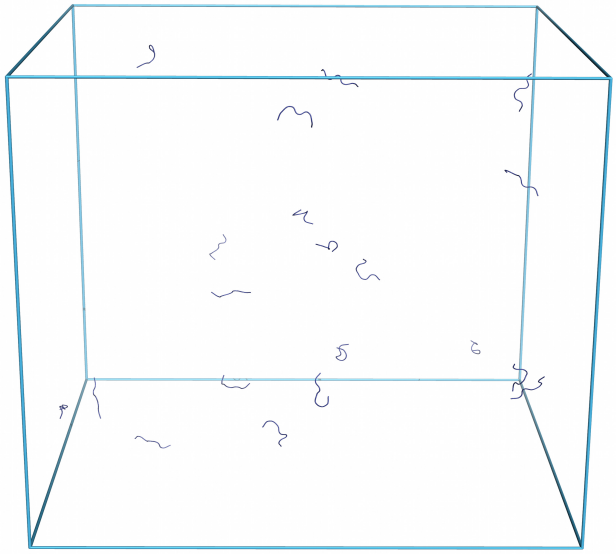

Figure 5.1: **Typical starting structure for our MD simulations.** The 20 peptides are initially in a random coil conformation and placed in a 200 Å box to ensure a 4.15 mM concentration

structure is solely used to determine cluster membership of the strands. The clustering analysis allows us to measure accurately the evolution of clusters over time based on local information and to monitor their properties such as the orientation of strands within β-sheets (i.e., parallel or anti-parallel). For purposes other than cluster determination, secondary structure calculations are made using the STRIDE program [102].

In order to look at the aggregation process in more details, we also consider the association and dissociation rates of the clusters in the following way. With N_g the concentration of $g-$mers, we consider aggregation as a dynamical process involving both association and dissociation that can occur either one monomer or more than one monomer at a time. The former is referred to as growth by monomer addition/monomer loss while the latter is described as being a mix of two processes, oligomer fusion/fragmentation and formation/destruction of oligomers from/into monomers, when involving more than one monomer at a time. We can then define the net rate of creation of $g-$mers as

$$\frac{\Delta N_g}{\Delta t} = C_g^{mono} + D_{g+1}^{mono} - (C_{g+1}^{mono} + D_g^{mono}) + C_g^{oligo} - D_g^{oligo} \tag{5.1}$$

where C_g^{mono} and D_g^{mono} are the creation rate of $(g-1)-$mers into $g-$mers and the destruction rate of $(g+1)-$mers into $g-$mers, C_g^{oligo} and D_g^{oligo} are the creation and destruction rates of $g-$mers either directly from/into monomers, or from the fusion / fragmentation of other sizes of oligomers. All the C and D rates are calculated from our clustering tool and allow us to gather statistics on the microscopic kinetic events and mechanisms.

5.6 Results/Discussion

We present a study of the aggregation kinetics of 20-mer GNNQQNY oligomers under a 4.15 mM concentration, the same concentration that was used in our previous multiscale thermodynamic study of the GNNQQNY 20-mer system [64].

We first present the general results obtained from the 100-ns MD simulations whose configurations were saved every 7.5 ps (5000 simulation steps) with an initial configuration selected as discussed in the materials and methods section. Then we discuss results from the 30-ns MD simulations whose configurations are saved every 75 fs (50 simulation steps) to better study the detailed association and dissociation kinetics of oligomers.

5.6.1 Observed kinetics

At the lowest temperature of 280 K, all 76 100 ns simulations lead to ordered amyloid oligomer formation. In all cases, aggregation is accompanied by a sudden drop of the total potential energy of the system, by over 600 kcal/mol over less than 10 ns, and by an increase in the β-sheet content of 30 %, as calculated with the STRIDE program [102]. While the exact energy value is not significative, due to the implicit-solvent coarse-grained nature of our energy model, its drop corresponds to the formation of a more stable structure . The system then stays in a minimum of energy and both the number of hydrogen bonds and the amount of secondary structure stabilize. As shown in Fig. 5.2(a), which presents a typical aggregation run, the β-content in the structures fluctuates typically around 50 %, near its maximum of 60 %, as the glycines and tyrosines end residues of each 20 peptides do not get involved in the β-sheet hydrogen bonding.

Fig. 5.2(a) also shows the high correlation between the energy drop and the increase in the number of hydrogen bonds as a function of time, suggesting that the cooperativity between hydrogen bonds plays a crucial role in lowering the energy and stabilizing the system.

Aggregation is slower at 300 K and only 68 % of the 76 100-ns simulations lead to ordered amyloids. However, as shown in the typical aggregation run in Fig. 5.2b, the overall ordering follows a trend very similar to that at the lower temperature : a sudden potential energy drop of over 600 kcal/mol over less than 10 ns accompanied by correlated raises in both the number of hydrogen bonds and the β-sheet content. If the final number of hydrogen bonds is very similar to that at 280 K, the secondary structure is less stable and tends to fluctuate around 40 % rather than 50 %.

In order to describe the assembly process we represent the time evolution, the probability density and the orientation of strands in structures as a function of the number of hydrogen bonds and of the number of contacts between side chains as these two coordinates are the least correlated and are the best measure of how ordered the structures are. Figures 5.3 and Supp. Fig. 1 show these quantities for the trajectories plotted in Fig. 5.2 at 280 K and 300 K, respectively.

At 280 K, we observe three distinct kinetic stages over the course of a typical simulation (Fig. 5.3(a)). The first phase is characterized by the nucleation phase, which lasts about 5 ns after the start of the simulation and leads to the formation of the metastable critical nucleus. During this phase, small oligomers form and break under stochastic collisions of the monomers. Seeds below the nucleus size fluctuate considerably, forming and disassembling at a high rate, forming a quasi-equilibrium perfectly reversible process. Once the metastable nucleus forms, the system can move into the aggregation (or growth) phase with a 50 % probability, by definition. In this dynamical phase, almost all of the monomers rapidly assemble around the nucleus to form a partially disordered globular oligomer. In general, this stage is very rapid and typically lasts less than 10 ns. During the third phase, which extends over a timescale of up to 80 ns, the aggregate rearranges itself as monomers explore their local configuration environment within the confines of the oligomers, optimizing the energy and, as a consequence, the secondary

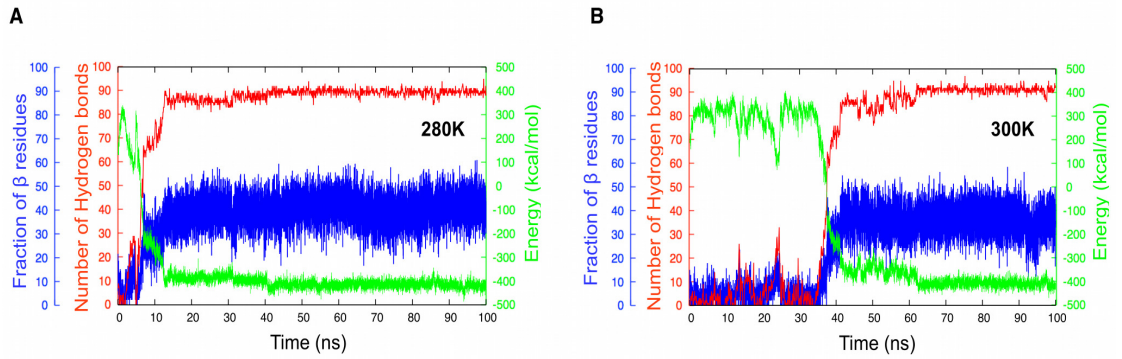

Figure 5.2: **Time evolution of the structural properties of the GNNQQNY 20-mer:** the total potential energy (in green), the total number of hydrogen bonds (in red) and the secondary structure (in blue) for a typical 100-ns simulation at (a) 280 K and (b) 300 K. The discontinuities in the maps (a) and (c) is a plotting artifact in low-density regions where the system is rapidly changing during aggregation and there are therefore not enough points to fill the map regions.

structure and the number of side chain – side chain contacts (see the last 75 ns in Fig. 5.3 (a) and (c)). This process, which we describe as a stabilization phase, is the slowest of the three and accounts for the dense region in Fig. 5.3(b).

This aggregation process is consistent with the “condensation-ordering" mechanism previously observed experimentally [205] and computationally [193–195, 206]. An interesting feature of the kinetics at 280 K is the increase and later dominance of parallel orientation in the structures over time during both the growth and stabilization phases while the structures are mostly antiparallel during the nucleation phase (Fig. 5.3(c)). By looking at the color coding on the right axis, it appears as though the system is loosing some parallel orientation between region 1 and 2 from almost 100 % to ∼80 %. Instead our results indicate that the system continues to evolve and gain some secondary structure between region 1 and 2 of the graph. It is the newly formed β-strands that adopt an antiparallel orientation while the parallel content formed during the growth process remains unchanged. As a whole, 91 % of the MD simulations at 280 K lead to a final assembly dominated by parallel β-sheets, in agreement with recent experimental findings [32, 47, 49] and computational studies [12, 17, 18, 21, 64].

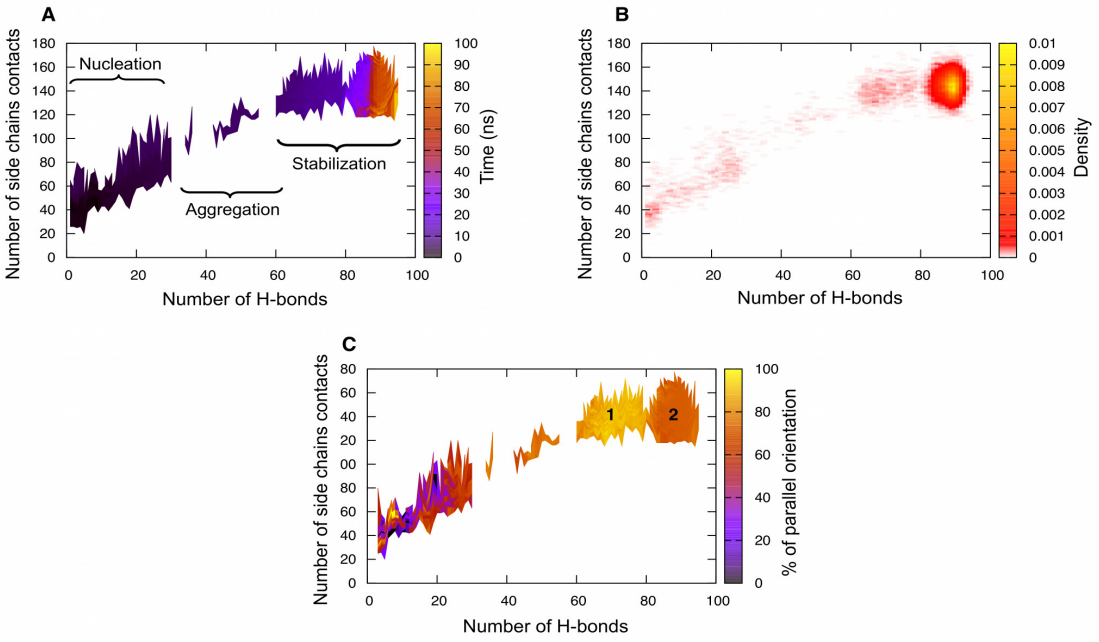

Figure 5.3: **Evolution of the structural properties for the GNNQQNY 20-mer at 280 K as a function of the number of hydrogen bonds and of the number of side chain contacts.** (a) Time evolution map of the system. Black regions indicate the beginning of the simulation while yellow regions indicate the end. (b) Density map representing the probability of having a configuration lie in a specific region. Yellow is the highest density and red the lowest. (c) Proportion of parallel β-strands. Yellow regions indicate that 100 % of the strands are in parallel orientation while black regions indicate that none of the strands are in parallel orientation thus meaning that they all are in antiparallel orientation. Since cluster determination is based on the presence of hydrogen bonds, the percentage of antiparallel orientation of the strands is equal to 1 minus the percentage of parallel orientation. Regions 1 and 2 indicate the two regions of highest parallel orientation.

At 300 K the kinetics globally display the same three phases for nucleation, growth and stabilization of oligomers observed at 280 K, and 95 % of the final aggregated structures display a dominance of parallel orientation of the β-strands (Supp. Fig. 1). The main difference between the two temperatures (Fig. 5.2) is in the lag time associated with the nucleation phase: while the average lag time is found to be $\sim$13 ns at 280 K, it increased to $\sim$56 ns at 300 K, leading to a denser nucleation region on the probability map (Supp. Fig. 1(b)). Mechanistically, this increase in nucleation time can be explained by the presence of bigger thermal fluctuations that destabilize the metastable aggregates, preventing nucleation.

While most simulations at 280 K and 300 K generate a single aggregation event, we observe reversibility for 34 % of aggregation events at 280 K against 40 % at 300 K. In these cases, such as in the example as shown in Fig. 5.4, monomers undergo a complete aggregation process up to and including the stabilization phase before the reverse reaction takes place, leading to a completely or partially random structure. For some simulations, this reversible transition was even observed to occur a few times during the 100 ns run. The presence of reversibility tells us that even though the free energy barrier for forming a 20-mer oligomer is high, the system is not completely biased towards the formation of an ordered oligomer. Thermal fluctuations for this 20-mer are sufficient to destabilize ordered oligomers on a relatively short time scale, a process that cannot be achieved in all coarse-grained aggregation simulations [197, 207] but which is crucial in order to describe aggregation kinetics correctly.

5.6.1.1 Diversity of the trajectories

Among the 76 simulations at 280 K, we find two simulations that display extreme behavior (Fig. 5.5). The first one, R1, (panels (a) to (c)) assembles into a very low-energy structure, reaching as low as -750 kcal/mol in places (lower than the minimum energy of ~ -450 kcal/mol for a typical aggregation process at 280 K) (Fig. 5.5(a)). Unexpectedly, this low energy is associated with the presence of a relatively small number of hydrogen bonds, between 70 to 80 at maximum, while it is usually between 90 to 100 (see Fig. 5.3). This low number of hydrogen bonds is compensated by a very high number of side

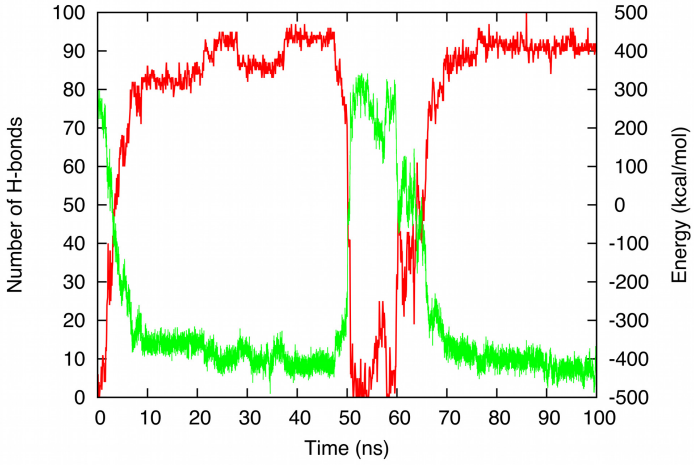

Figure 5.4: **Time evolution of the structural properties of a GNNQQNY 20-mer simulation that shows reversibility in the formation of a structured aggregate:** the total potential energy (in green) and the total number of hydrogen bonds (in red).

chain – side chain contacts that reaches 190 and more, well above the usual maximum of 170 (see the yellow section of Fig. 5.5(b)). In Fig. 5.5(a) the maximum amount of secondary structure is $\sim$45 %, lower than a typical structure in our simulations and seems to fluctuate significantly less than for a typical structure, due to the small number of hydrogen bonds in the structure. The second extreme simulation, R2, (panels (d) to (f) of Fig. 5.5), shows the opposite behavior, with a high number of hydrogen bonds (panel (d)) and a very low number of contacts (between 100 and 120) during the last 20 ns (in yellow in Fig. 5.5(f)). The secondary structure here behaves similarly to a typical simulation and fluctuates around $\sim$50 % (Fig. 5.5(d)). Looking at the final morphologies of the structures in each case (Fig. 5.6 (a) and (b)), we see that the first simulation, with a high number of side chain – side chain contacts and fewer hydrogen bonds, leads to a very compact oligomer composed of several small β-sheets while the second simulation, with a large number of hydrogen bonds and fewer contacts, favors a protofibril-like structure made of two long twisted β-sheets facing each other.

These two simulations demonstrate the existence of a competition between the optimization of the number of contacts and hydrogen bonds. This competition generates a

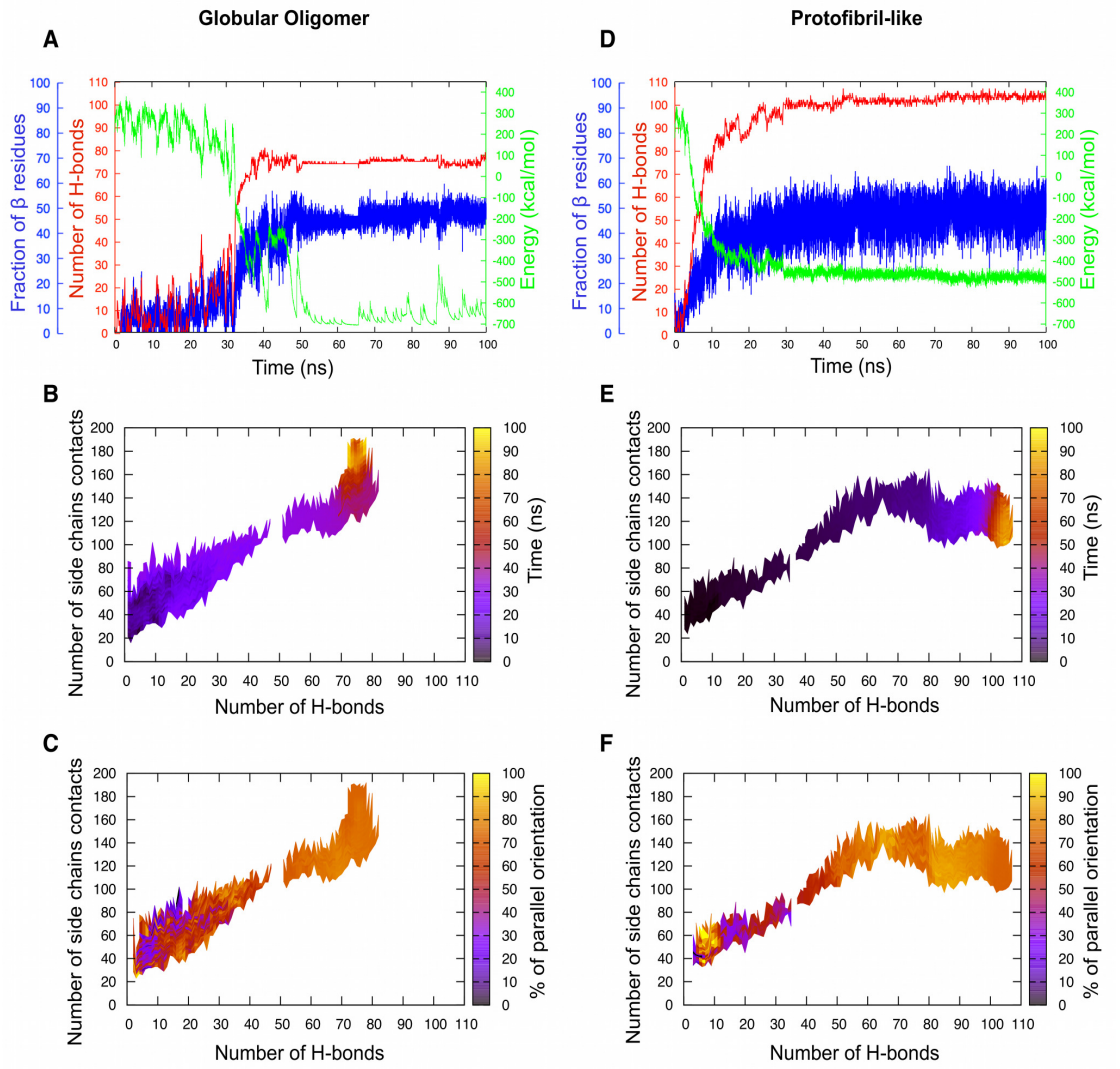

Figure 5.5: **Competition between the globular oligomer (R1) and the protofibril (R2).** (a) Low energy profile displaying a low amount of hydrogen bonds for the globular oligomer; (b) corresponding time map as a function of the number of hydrogen bonds and of the side chain contacts; (c) corresponding amount of parallel β-strands as a function of the same parameters; (d) kinetic profile displaying a particularly high number of hydrogen bonds in the protofibril structure; (e) corresponding time map as a function of the number of hydrogen bonds and of the side chain contacts. f) corresponding amount of parallel beta-strands as a function of the same parameters. These graphs demonstrate the existence of a competition between the globular structure with a low amount of hydrogen bonds and a high amount of contacts and the protofibril structure with a high amount of hydrogen bonds and a low amount of contacts. The actual structures are shown in Fig. 5.6 (a) and (b).

family of formation pathways that can lead to formally different topologies, ranging from a compact globular oligomer to an extended protofibril-like structure [64, 130, 131, 208]. In all cases, the final structures display a very high parallel β content (Fig. 5 (c) and (f)). By comparing these two extreme cases to morphologies of typical simulations at 280 K, we note that, statistically, most of these fall squarely into one of two general morphologies: a class of 3-plus sheet structures that seem rather compact (Fig. 5.6(c)), which resembles the compact oligomer, and a class of 2-sheet structures (panel (d)) similar to the protofibril-like structure (Fig. 6 (b)) but with a distribution of orientations between the two β-sheets. The two extreme structures, which we had already observed in our previous study [64], can therefore be considered as the optimal cases of the two large structure families of ordered amyloid aggregates generated in our simulations.

5.6.2 Details of the aggregation kinetics - the “growth" phase

In this section we present the analysis of the 10 30 ns-MD simulations, five at 280 K and five at 300 K, whose configurations are saved every 75 fs in order to describe the details of the kinetics during the final nucleation and full growth process. Because of the tremendous size of the resulting simulation data, we concentrated our analysis on a 10 ns-window centered around the drop in energy (Fig. 5.7). Panel (a) represents the average energy taken over all five simulations as a function of time at 280 K. Trajectories are aligned, in time, at the point at which they reach -80 kcal/mol, which is roughly the midpoint in the energy drop for all simulations. Most of the energy drop associated with oligomeric growth, on the order of 600 kcal/mol $\pm$ 100 kcal/mol, takes place over 4 ns, in agreement with our earlier observations for a typical aggregation process at 280 K. The relatively small error bars along the energy curve indicate the good reproducibility of the properties over time at 280 K. At 300 K, the growth phase associated with the energy drop, of about 450 kcal/mol $\pm$ 200 kcal/mol, also takes on the order of 4 ns (Fig. 5.7(b)), similar to a 280 K energy drop. The standard deviation on the 300 K curve is, however, greater than at 280 K, demonstrating a greater variability associated with larger thermal fluctuations.

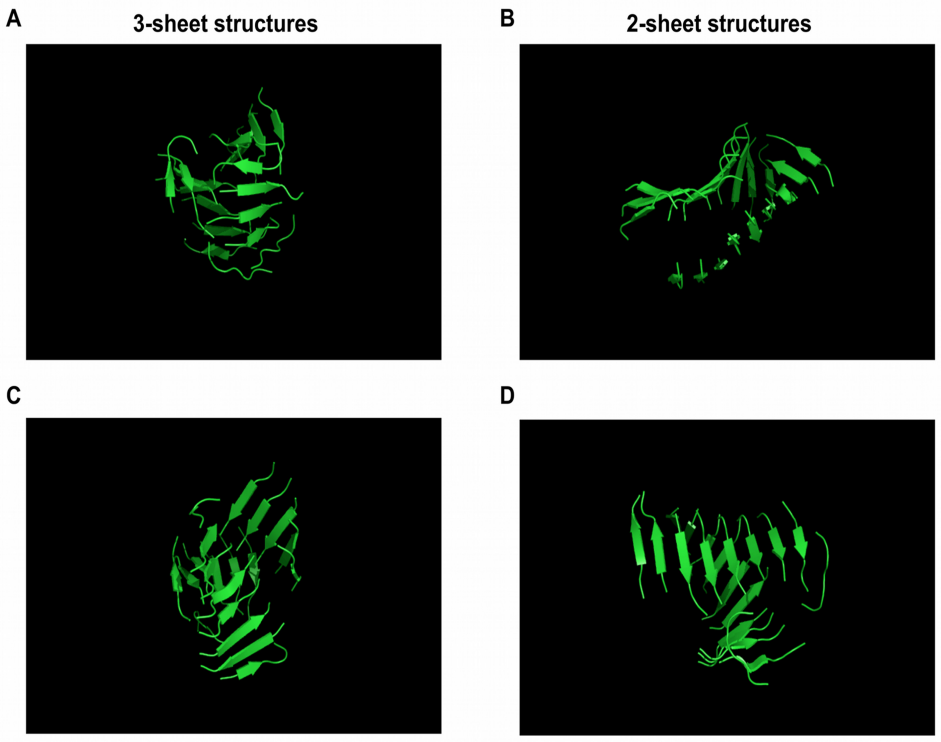

Figure 5.6: **Diversity of the morphologies.** (a) Oligomer displaying an extremely high amount of contacts (simulation R1). (b) Protofibril-like structure displaying a high number of hydrogen bonds (simulation R2); (c) typical 3-plus sheet structure often generated in our simulations; (d) typical 2-sheet structure.

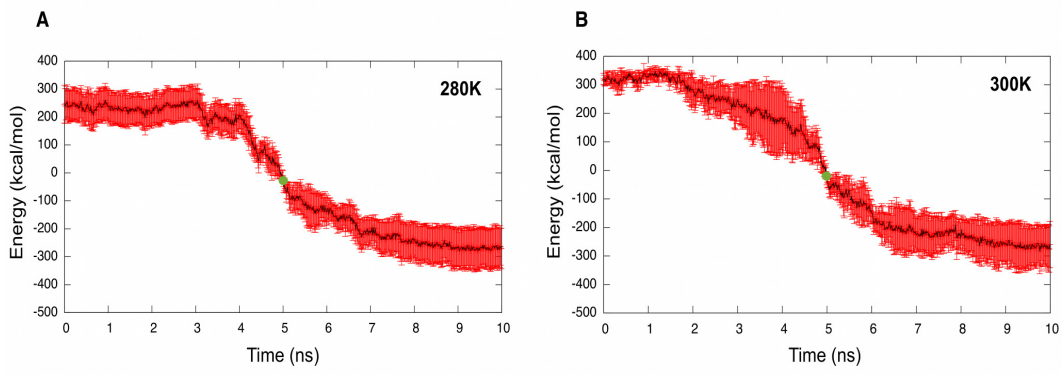

Figure 5.7: **Detailed kinetics of 30 ns MD simulations.** a) Average energy profile at 280 K (5 sets of data were used). The average curve is shown in black while the standard deviation is shown in red. The green dot represents the point around which the profiles were centered before computing the statistics over a 10 ns window (5 ns before and 5 ns after); b) Average energy profile at 300 K (5 sets of data were used).

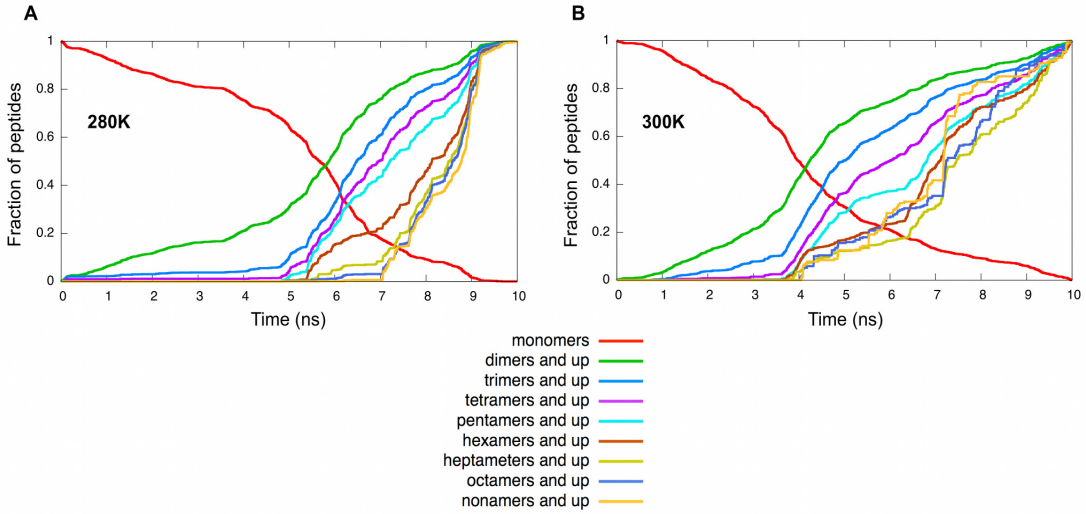

Figure 5.8: **Size evolution.** (a) Fraction of peptides occurring as monomers and cumulative curves for all aggregates size up to 9 monomers at 280 K, i.e., probability of finding an aggregate of at least size n for various n. (b) Same as (a) at 300 K. One simulation run is shown at each temperature, for which data was collected for a 10-ns window centered around the energy drop.

5.6.2.1 Sigmoidal growth and lag time

In Fig. 5.8 (a) and (b), we plot the cumulative size of various aggregates as a function of time over the 10 ns period, i.e., the occurence probability for oligomers of size at least as large as indicated in the graph. Aggregates with a minimal size larger than one show a sigmoid-like growth starting from about the middle of the energy drop, at 5 ns, with an increasing lag time the bigger the species. Sigmoidal aggregation kinetics have been largely observed experimentally [46, 104, 112, 114, 117, 127, 134, 209–211] and numerically [123, 126, 212, 213] and are a well-established characteristic of a nucleated-growth process. Similar cumulative curves have been obtained for Monte-Carlo simulations of large systems of hexapeptides [213] which indicate that the cooperativity between contacts plays a crucial role in the growth and stabilization of all sizes of aggregates.

5.6.2.2 Origin of the energy drop associated with aggregation

Looking at the correlation between the different energy components and the global energy profile at 280 K and 300 K (Fig. 5.9), we observe that the two main contributions to the energy drop come from the hydrogen bonding energy and the hydrophobic/hydrophilic interaction energy. First, the initial collapse of the peptides is accompanied by a decrease in the hydrophobic energy (blue curve, Fig. 5.9) quickly followed by a decrease in the hydrogen bonding energy, which becomes increasingly dominant over the hydrophobic energy, as previously observed in another numerical study of the aggregation kinetics of amyloid peptides [195].

5.6.3 Nucleus characterization

At both temperatures 280 K and 300 K, aggregation is generally triggered by the formation of a small-sized metastable aggregate, which appears to be stable after a certain lag time. This suggests that we are in the presence of an assembly sequence that can be classified as a nucleated-growth process [110, 111, 116, 123, 201, 214–216], i.e., that this small metastable aggregate, which we term nucleus, serves as a nucleation center

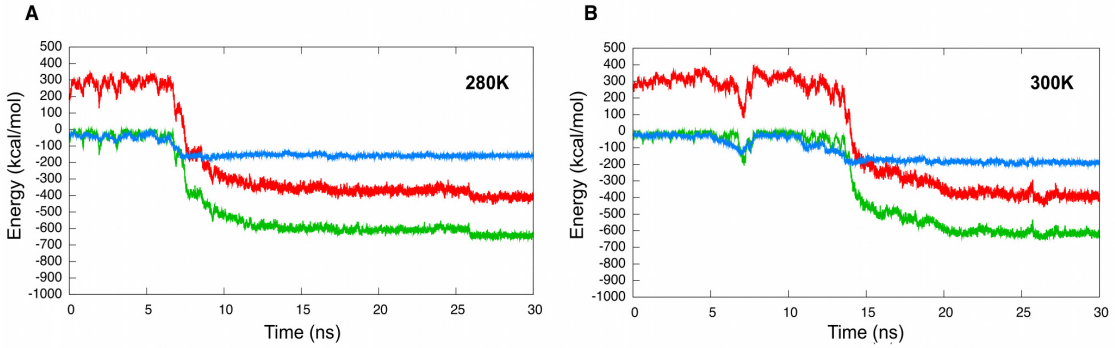

Figure 5.9: **Energy contributions to the total potential energy at (a) 280 K and (b) 300 K.** This figure shows one entire 30 ns run for both panels (a) and (b). The energies shown are the hydrophobic/hydrophilic energy (blue curve) and the hydrogen bonding energy (green curve). The total potential energy is shown in red.

of the aggregation process. The 152 100 ns-MD simulations were divided in 3 sets at both 280 K and 300 K and we computed the free energy as a function of aggregate size and secondary structure for those 3 sets of simulations in order to determine the size and amount of secondary structure of the critical nucleus (Fig. 5.10). Performing this task on different sets of data allows us to have an idea on the order of the fluctuations in the free-energy. At 280 K the nucleus size corresponding to the maximum of free energy is found to be between 4 (Fig. 5.10(a) - green curve) and 5 monomers (Fig. 5.10(a) - red and blue curves) and between 5 (Fig. 5.10(b) - red and blue curves) and 6 monomers (Fig. 5.10(b) - green curve) at 300 K. This result is expected since larger thermal fluctuations require a bigger aggregate to survive and lead to growth. The pentameric critical nucleus identified here is also near the critical size estimated by Nelson *et al.* [32] and by us, in a previous thermodynamic study [64].

As was shown recently [217, 218], the critical nucleus size in a finite-size system is systematically overestimated and it is necessary to correct for this artifact. From the classical nucleation theory (CNT), Grossier *et al.* derive an expression for the total free energy of forming an aggregate of size g monomers in an infinitely large system to be [217]:

$$\Delta G_{\infty}(g) = -gk_BTln(\beta_0) + Ag^{2/3} \tag{5.2}$$

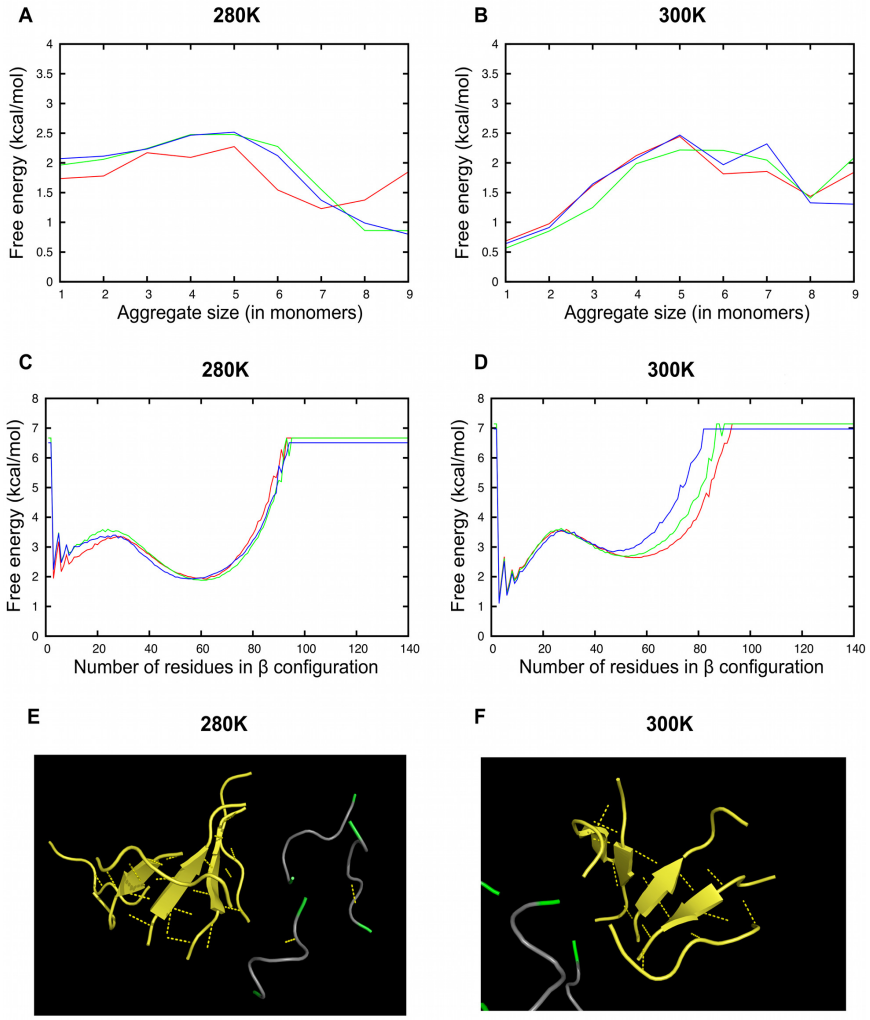

Figure 5.10: **Critical nucleus characterization.** (a) Free energy as a function of the aggregate size at 280 K. (b) Same at 300 K. In both (a) and (b), the maximum of free energy corresponds to a critical nucleus size of $\sim$5 monomers. (c) Free energy as a function of the number of residues in a β conformation at 280 K. (d) Same at 300 K. In both (c) and (d), the first maximum in free energy represents the critical amount of secondary structure necessary for a nucleus to be stable and trigger aggregation. (e) Typical structure at 280 K of a pentamer nucleus with the critical amount of secondary structure shown in panel (c). (f) Typical structure at 300 K of a pentamer nucleus with the critical amount of secondary structure shown in figure (d).

where g is the aggregate size, k_B is the Boltzmann constant, T is the temperature, β_0 is a dimensionless constant and represents the supersaturation and A is the interfacial energy (or surface tension) taken to be a constant in the model. Due to our very small system size, a 20-mer, and the low critical nucleus, it is not possible to obtain a good fit to this continuous equation. However, the overestimation correction could explain the slight difference we observe with respect to the experimentally-derived critical nucleus of four monomers.

Looking at the free energy barrier of forming a certain amount of secondary structure, we find that a viable nucleus requires between 24 and 28 residues in β conformation at 280K while it requires between 27 and 29 residues in β conformation at 300K (Fig. 10C and D). The increase in free energy for 80 residues is due to the finite-size effects of our system. It becomes harder to have 80 residues in β-conformation as no more monomers are available to the system to continue growth. Figure 10 E and F show the dominant pentamer nucleus structure having such amount of secondary structure at 280K and 300K. In both cases, the pentamer seed is partially ordered. In most cases, no more than a dimer is formed beside the nucleus.

5.6.4 Proposed mechanism

To assess the microscopic mechanisms involved in the kinetics, we first identify all types of association and dissociation: growth by monomer addition (and, reversibly, loss by monomer subtraction), growth by fusing two oligomers together (and, reversibly, fragmentation of one oligomer into two smaller oligomers at least 2 monomers in size) and the direct formation/destruction of oligomers from/into monomers. In this section, we refer to any aggregate bigger than one monomer as an oligomer. It is important to point out that there is a wealth of “monomer addition" models for diverse polymer-forming proteins such as actin [124, 125], tubulin [126], the sickle cell hemoglobin [107, 127] and amyloid proteins such as Aβ [103, 128, 129], β2-microglobulin [114] and Sup35 [112]. There also exists numerous “oligomer fusion" models for Aβ [43, 130, 134, 192], α-synuclein [43, 131, 135], the phosphoglycerate kinase protein [133], the lysozyme protein [132] and Sup35 [205, 219], some of which have observed both

processes happening at the same time.

Association and dissociation rates were calculated, with our clustering code, every 75 fs over a 10 ns-window (centered around the energy drop) for the 30 ns-simulations and as described in Eq. (5.1). Then, for each time interval, we calculated the total number of events, originating either from monomer addition/loss, from oligomer fusion/fragmentation or from monomers↔oligomers events across all species such as:

$$\frac{\Delta N_{tot}^{mono}}{\Delta t} = \sum_{g=2}^{20} \frac{\Delta N_{g}^{mono}}{\Delta t} \tag{5.3}$$

and

$$\frac{\Delta N_{tot}^{oligo}}{\Delta t} = \sum_{g=2}^{20} \frac{\Delta N_{g}^{oligo}}{\Delta t} \tag{5.4}$$

where $\Delta N_{g}^{mono}/\Delta t$ and $\Delta N_{g}^{oligo}/\Delta t$ are the “monomer addition/loss" and “oligomer fusion/fragmentation" + monomers↔oligomers components of Eq. (5.1).

Figure 5.11 shows the evolution of these two quantities for both association and dissociation events at 280 K (Fig. 5.11 (a)) and at 300 K (Fig. 5.11 (b)). We differentiate the “fusion/fragmentation" events from the formation/destruction of oligomers (bigger than dimers) directly from/into monomers. At both temperatures, the data clearly shows that the assembly mechanism is dominated by “monomer addition/loss" events. Then when nucleation and aggregation happen, we see a notable increase in the amount of monomer events and a trigger of “oligomer fusion/fragmentation" and “monomers↔oligomers" events. We notice a well-defined increase in the number of “monomer addition/loss" events just before the first “oligomer fusion/fragmentation" events appear. This increase corresponds to the start of nucleation and suggests that once nucleation is triggered and most of the monomers are recruited, they join different sites, or clusters, that will later on fuse together to form a larger oligomer. Later, when the aggregate stops growing in size, we observe no more “single monomer" or “monomers↔oligomers" events and observe, in some cases, the presence of only fusion and fragmentation of oligomers (Fig. 5.11 (a)). This means that further rearrangements in the structure during the stabilization

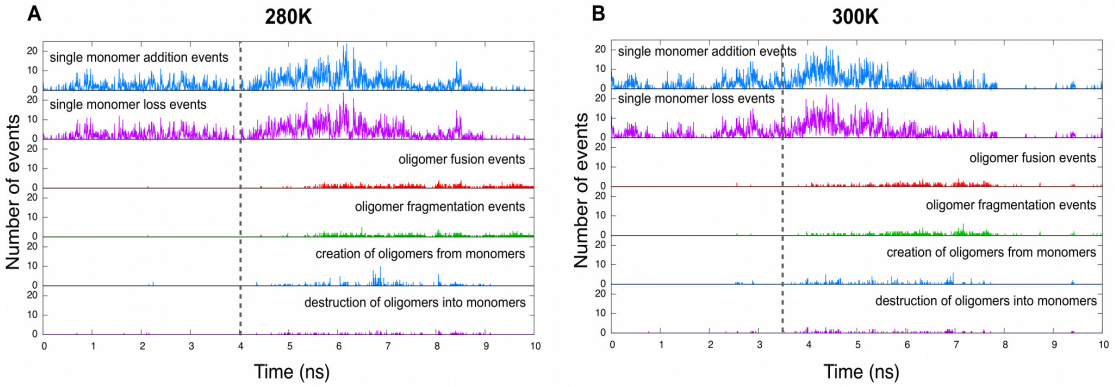

Figure 5.11: **Aggregation mechanism.** Time evolution of association and dissociation events before during and after nucleation by either single monomer events (meaning one at a time), by oligomer fusion/fragmentation, or by formation/destruction of oligomers from/into monomers at (a) 280 K and (b) 300 K. The dashed grey line indicates the beginning of nucleation. In (a), at 280 K, the aggregate stops growing in size after t = 9 ns while in (b), at 300 K, the aggregate stops growing just before t = 8 ns. For ease of reading, each point in the graphs is the sum of events in $5\Delta t$ ($5\Delta t = 375\ fs$).

phase are accomplished mainly through oligomer-involving events, if any.

5.6.5 Conclusion

We presented here a detailed study of the onset of amyloid aggregation for 20-mers of GNNQQNY. Using molecular dynamics with the OPEP coarse-grained force field, we show that nucleation of this small amyloid peptide is dominated by monomer addition/loss events, with very small contributions from larger oligomers, following closely the classical nucleation theory. It is then meaningful to extract a critical nucleus, that can be obtained from the calculation of the free-energy as a function of nucleus size. We find that, at 280 K, this critical size is between 4 and 5 monomers, while it is between 5 and 6 at 300 K, in good agreement with the experimental estimate of 4 monomers [32], especially when taking into account the finite-size bias that tends to overestimate the size critical nucleus [217, 218]. Correspondence with CNT stops there, however, as the kinetic process associated with aggregation and growth differs in two majors from this theory. First, while most of the structural organization takes place during the 4 ns

growth process, aggregates continue to mature by collective motions, slowly dropping in energy as hydrogen bonds and β-sheet content evolve. Second, nucleation does not lead to a single structure, but shows clear polymorphism with a distribution of assemblies that can be classified into two distinct categories: a compact oligomer made of a number of relatively short β-sheets, typically three, and a more extended fibril-compatible two-sheet structure. These structures represent well-separated local basins and the only way to move between them, in our simulations, was through a complete dissociation and reassociation of the monomers. The well-defined polymorphic nature of GNNQQNY is in line with experimental and numerical observations in other amyloid sequences, such as amyloid-β. It was shown there that the protein could adopt multiple fibrillar structures [3, 220], but also off-pathway β-barrel organizations that would be responsible for at least part of the toxicity. [221] For GNNQQNY, the two polymorph families observed here are close enough that they should lead to different fibrillar structures rather than on and off-pathway organizations. Only simulations with a larger number of peptides will be able to tell.

How much of these results can be applied to experimental studies of GNNQQNY? A previous stability study of the structures predicted with OPEP using explicit SPC solvent and all-atom GROMOS96 showed that our simulations are realistic, except for the most extended structures [64]. If the growth time is not directly extendable to all-atom systems, the thermodynamics and, therefore, the critical nucleus size but also the polymorphism, which is a signature of amyloid aggregates, should be valid. Our results suggest that the specific shape, out of a family of structures, is selected very early on and that moving from one to another requires going over a very high barrier, high enough that it was never observed in our simulations, the preferred being going first through a complete dissociation. Such behavior could change with larger aggregates, and the direct rearrangement become more favorable than complete dissociation. Only further work, on larger systems, will show whether new families of structures are possible for GNNQQNY and if the CNT applies when more monomers are in play. Our results on the 20-mer of GNNQQNY are at least compatible with experiments and offer a number of insights into the onset of aggregation and polymorphism for small amyloid peptides.

5.7 Acknowledgements

The authors thank Riccardo Pellarin and Baron Peters for their suggestions. This work was supported in part by the Natural Science and Engineering Research Council of Canada, the Canada Research Chair Foundation and the Fonds québécois de recherche en santé. We are grateful to Calcul Québec / Compute Canada for generous allocation of computer time.

5.8 Supporting Material

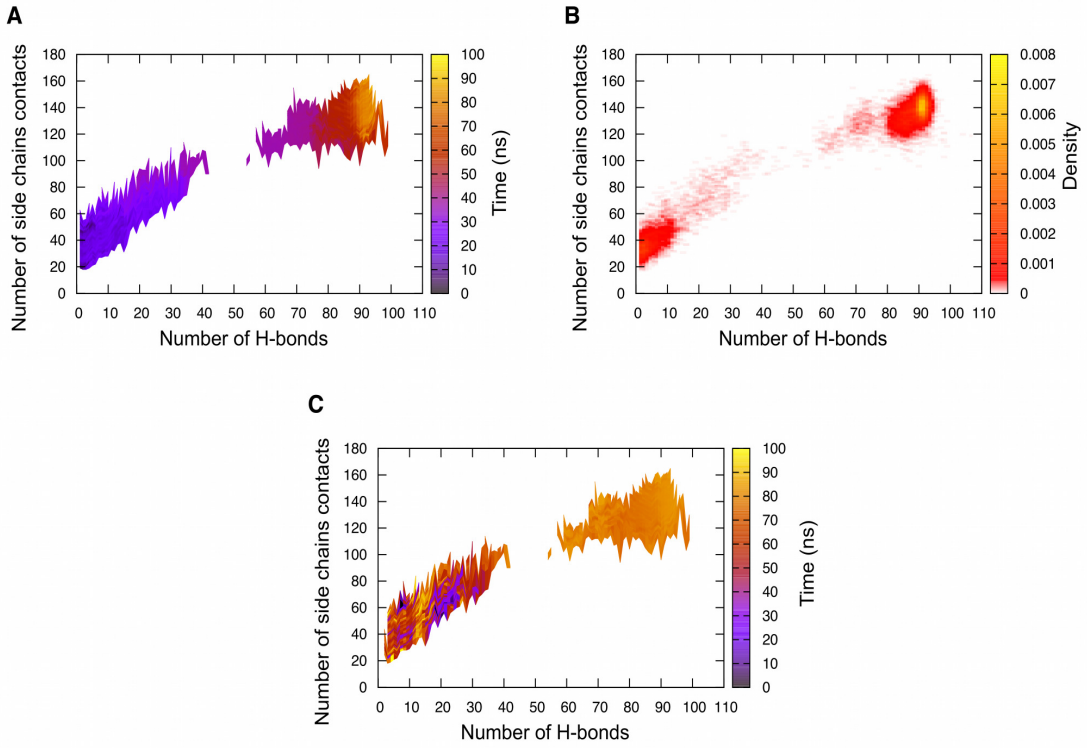

Figure 5.12: **Supplementary Figure 1: Characteristics at 300 K for the GNNQQNY 20-mer as a function of the number of hydrogen bonds and of the number of side chain contacts.** (a) Time evolution map of the system. Black regions indicate the beginning of the simulation while yellow regions indicate the end. (b) Density map representing the probability of having a configuration lie in a specific region. Yellow is the highest density and red the lowest. (c) Proportion of parallel β-strands. Yellow regions indicate that 100 % of the strands are in parallel orientation while black regions indicate that none of the strands are in parallel orientation thus meaning that they all are in antiparallel orientation. In all three plots, the nucleation region is denser, due to bigger thermal fluctuations at 300 K, which destabilize early metastable aggregates.

CHAPITRE 6

THE AGGREGATION OF THE AMYLOID PRION SUP35 GNNQQNY PEPTIDE INTO ORDERED PROTOFIBRILS

6.1 Objectifs

L'objectif de cet article, en préparation et projeté d'être soumis dans le journal PLoS Computational Biology sous peu, est l'étude numérique du système 50-mère de GNNQQNY et des morphologies qui lui sont accessibles. L'étude du 50-mère apporte une autre dimension à la compréhension de l'agrégation des peptides amyloïdes non seulement par rapport aux résultats expérimentaux obtenus pour la séquence GNNQQNY mais aussi par rapport à nos précédents travaux sur le 20-mère de GNNQQNY.

6.2 Contributions des auteurs

- (Dans l'ordre d'apparition dans l'article) Jessica Nasica-Labouze (JNL) et Normand Mousseau (NM) ont conçu ensemble le plan d'action pour cette étude numérique.
- JNL a réalisé les simulations, l'extraction et l'analyse des données.
- JNL et NM ont contribué à l'élaboration des outils de simulations, d'extraction de données et d'analyse.
- JNL a rédigé l'article sous sa forme actuelle et NM a révisé le texte.

6.3 Abstract

The structural characterization of early amyloid aggregates is of substantial physiological importance. Using the OPEP coarse-grained potential and replica-exchange molecular dynamics, we characterize the 50-mer system of the GNNQQNY amyloidogenic peptide from the yeast prion protein Sup35. We find that the 50-mer aggregates are soluble multilayered protofibril-like structures composed of twisted β-sheets, typically

5 or 6. Theses structures can adopt one of two overall shapes, being either elongated or spherical, the former favoring an optimization of the side chain contacts while the latter rather favors the formation of hydrogen-bonds to optimize the secondary structure. We have identified several features of the cross-β structure in the configurations accessible to the 50-mer except for the antiparallel arrangement of the β-sheets, which we found to be parallel. We suggest that there exists a high free-energy barrier of formation for the antiparallel arrangement of the β-sheet, which would render the aggregate unfavorably insoluble at the 50-mer level but more probable for higher order system sizes.

6.4 Introduction

Amyloid proteins are key players in numerous degenerative diseases such as Alzheimer's, Parkinson's, Creutzfeldt-Jakob and Diabetes Mellitus caused by the triggering of their misfolding [25, 26, 136–138]. The byproduct of this misfolding, amyloid fibrils, are kinetically and structurally complex biological systems, which have been very challenging to characterize both experimentally and computationally. A detailed high-resolution structure of the core spine of fibrils from a short amyloid fragment from the yeast prion protein Sup35, GNNQQNY, has been obtained experimentally and has provided great insight into the atomic mechanisms underlying amyloid fibril formation [32]. It reveals a cross-β structure composed of two parallel β-sheets interacting in an antiparallel manner and whose side chains interdigitate to form a stabilizing steric zipper between the two sheets. The aggregation process, however, is still misunderstood and the complex multi-pathway kinetic mechanism leading to the formation of fibrils might involve several intermediates such as disordered oligomers and/or ordered protofibrils [11, 25, 41, 130, 134, 192–198, 222–225], whose structural details are subject to debate.

Computational studies have been a very useful component of recent advances in the improvement of our understanding of the amyloid fibril formation process and they have proven to be good complements to experimental works.

We have successfully characterized structures for the 20-mer system that were also stable in all-atom molecular dynamics simulations and we found that the 20-mer system

was highly polymorphic and had access to two distinct morphologies: one being a compact spherical oligomer structure and the other being an elongated fibril-like structure [64]. We later examined the kinetics of the same 20-mer system and confirmed that the aggregation mechanism was a nucleation process and that the same two classes of morphologies were accessible to the system.

In this paper, we now investigate the aggregation and morphologies accessible to the 50-mer system of GNNQQNY using the well-tested OPEP coarse-grained potential [52, 53] coupled to Replica Exchange Molecular Dynamics (REMD) simulations for accelerated sampling [99] and compare our results to our previous study of the 20-mer GNNQQNY system. While other computational works on GNNQQNY have investigated the first steps of aggregation for aggregates smaller that the 20-mer [12–14, 14–23], we present, to our knowledge, the largest spontaneous self-assembly of GNNQQNY at near-atomic resolution starting from a random conformation. In total, we have accumulated more than 10.6 μs of simulations, allowing relevant statistical analysis. Overall, our results show that protofibril-like structures are accessible at the level of the 50-mer system and whose features are characteristics of experimentally determined amyloid structures.

6.5 Materials and Methods

This study follows our previous numerical work, which explored the thermodynamics of the 3-mer, 12-mer and 20-mer GNNQQNY systems [64]. We therefore apply the same simulation methods and analysis tools as previously used in both of these studies.

6.5.1 The OPEP description

We have carried out implicit solvent coarse-grained simulations with the OPEP potential version 3.2. For each amino acid, the OPEP coarse-grained description consists of all heavy backbone atoms fully represented (N, H, C_α, C and O) and of one single bead, with appropriate geometrical parameters and van der Waals radius, to represent the side chain. The implicit effects of the solvent environment are taken into account in the interaction parameters of the energy function, as detailed elsewhere [52]. OPEP has proven to

be a reliable potential for protein folding and structure prediction [161, 162], the extraction of thermodynamic [62–66], and recently kinetic [226], properties for large systems and has been tested and combined with several numerical methods such as Monte-Carlo [53–57], the activation-relaxation technique (ART Nouveau) [67–72], MD [58–61, 226] and REMD [62–66].

6.5.2 Simulation details

Increasing the size of the system to 50 monomers at the OPEP resolution (2050 particles) was a computational challenge and as a result, the length and number of simulations achieved are shorter compared to the 20-mer simulations previously performed [64, 226]. Our data set is, however, large enough to be statistically significant and our work is, to our knowledge, amongst the first to present such a large amount of statistics for such a large system at near-atomic resolution.

6.5.2.1 Initial structures

For all simulations, the initial structures consisted of 50 random-coiled GNNQQNY monomers in a 275 Å periodic box (Fig. 6.1) to ensure a 4.15 mM concentration, same concentration as our previous simulations of the GNNQQNY system [64, 226]. All simulations were started from the same initial atomic positions but with different random velocities following a Maxwell-Boltzmann distribution.

We use REMD simulations to explore the large folding energy landscape of our 50-mer system. REMD is a widely-used method [99] allowing an efficient sampling of the phase-space accessible to a system by allowing conformations stuck in a local minimum of energy to further explore other phase-space coordinates. In a typical REMD simulation protocol, N MD trajectories (replicas) are run in parallel at N different temperatures.Then, conformational exchanges between adjacent pairs of trajectories are attempted, at regular time intervals, with a Metropolis accept-reject probability:

$$p(i,j) = \min\left\{1.0, \exp\left[\left(\frac{1}{k_B T_i} - \frac{1}{k_B T_j}\right)(E_i - E_j)\right]\right\} \tag{6.1}$$

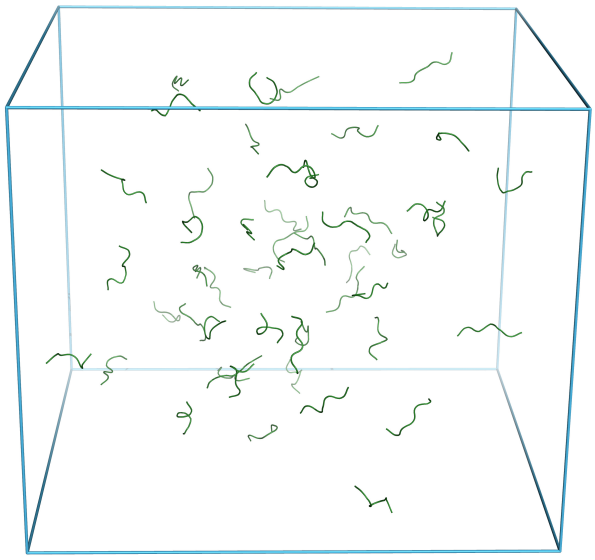

Figure 6.1: **Simulations starting structure.** The 50 peptides are in a random coil conformation and placed in a periodic box 275 Å in size in order to maintain a ∼4.15 mM concentration.

where trajectory i is at temperature T_i with an energy E_i and trajectory j is at temperature T_j with an energy E_j before the exchange. Due to temperature exchanges, all dynamical information is lost while statistical information is retained and only thermodynamic properties, when converged, can therefore be assessed.

We carried out 265-ns-long REMD simulations with exchanges performed every 5000 steps, a simulation time-step of 1.5 fs and configurations saved every 5000 steps. At the beginning of each run, a minimization procedure is performed using a damped MD, followed by a thermalization procedure in 50 000 steps. For our purposes, we chose a gaussian temperature distribution of 40 temperatures centered on 283 K, previously found to be the melting temperature for the 20-mer GNNQQNY [64]. The temperature distribution is detailed in Table 6.I. The gaussian shape of the distribution allows for temperatures to be closer to one another around 283 K, which results in a better sampling at temperatures where the energy is quickly changing for systems undergoing a phase transition at ∼283 K, such as the 20-mer and which we expected to be roughly the same for the 50-mer. Instead, as will be detailed in the "Results/Discussion" section,

the 50-mer system seems to display two transitions, one at ∼280 K, which corresponds to a structural change in the global morphologies and a second, more pronounced, at ∼335 K, which corresponds to a phase transition (or "melting") between ordered and disordered states.

T01	T02	T03	T04	T05	T06	T07	T08	T09	T10
223.8 K	237.7 K	249.2 K	255.9 K	260.1 K	263.1 K	266.0 K	268.2 K	270.2 K	272.1 K
T11	T12	T13	T14	T15	T16	T17	T18	T19	T20
273.8 K	275.5 K	277.1 K	278.6 K	280.1 K	281.6 K	283.0 K	284.5 K	285.9 K	287.4 K
T21	T22	T23	T24	T25	T26	T27	T28	T29	T30
288.9 K	290.6 K	292.2 K	293.9 K	295.7 K	297.8 K	300.1 K	302.7 K	305.9 K	310.1 K
T31	T32	T33	T34	T35	T36	T37	T38	T39	T40
316.8 K	328.3 K	342.2 K	356.2 K	370.1 K	384.1 K	398.0 K	412.0 K	426.0 K	440.0 K

Table 6.I: **Temperature distribution (in Kelvins) for the REMD simulations.**

6.5.3 Analysis

Part of the analysis was performed using a clustering tool able to compute and collect statistics of structural properties such as the number and lengths of β-sheets as well as the orientation and register of β-strands. The algorithm defines clusters based on strand attachment, defined from hydrogen bonds between strands [100]. Discriminative secondary structure criterions on the dihedral angles ϕ and ψ and a cutoff of one hydrogen bond are used to determine if a strand belongs or not to a cluster. The clustering tool was also used to calculate association and dissociation rates between clusters, *i.e.* β-sheets. More details on our clustering tool are provided elsewhere [226].

A PTWHAM analysis was also performed to compute thermodynamic properties of the system for the simulations [101].

Secondary structure calculations (β-sheet content) were made using the STRIDE program [102].

6.6 Results

We first present the data obtained from REMD simulations, assessing the thermodynamic properties of the 50-mer GNNQQNY as well as the morphologies accessible to the system. The results are discussed and compared to our results obtained for the 20-mer system of GNNQQNY [64, 226].

As detailed in the Methods and Materials section, REMD simulations were carried out with 40 replicas for 265 ns. Thanks to our temperature distribution, the exchanges between the replicas are optimized and most low-energy structures exchange well with all temperatures below melting and are therefore not stuck in a local minimum of energy, allowing for a good phase-space sampling. A PTWHAM analysis was first performed to compute and determine the evolution of thermodynamic properties as a function of temperature such as the specific heat, to observe phase transitions, and the radius of gyration, to assess the overall size of the system. Looking at the specific heat, the system appears to be fully converged over the last 210 ns despite a slight shift over time of the specific heat peak (Fig. 6.2 (a)). The system seems to be undergoing two distinct transitions: the first between 275.8 K and 287.8 K and the second, sharper, between 334.8 K and 337.8 K. The first transition is accompanied by a sudden decrease in the radius of gyration R_g of the system of $\sim$6 Å (Fig. 6.2 (b) and Table 6.II). Indeed, we observe that low-temperature ordered structures with high β-sheet content become more globular passed that transition. They however retain most of their secondary structure but the β-sheets rearrange themselves towards a more spherical conformation. The second observed transition corresponds to the "melting" of the system from ordered to disordered states, meaning that for temperatures above the melting point, little or no secondary structure is observed within 265 ns. The structural details of configurations as a function of temperatures below and above melting are given in Table 6.II.

Structurally, the system has a high propensity for forming β-sheets, with a β-content ranging from 45 to 70 % below the melting temperature. Aggregation is extremely favorable and the system forms several – typically 5 or 6 – long β-sheets where typical sheet lengths vary from 4 to 11 strands between 223.8 K and 260.1 K (below the first transition

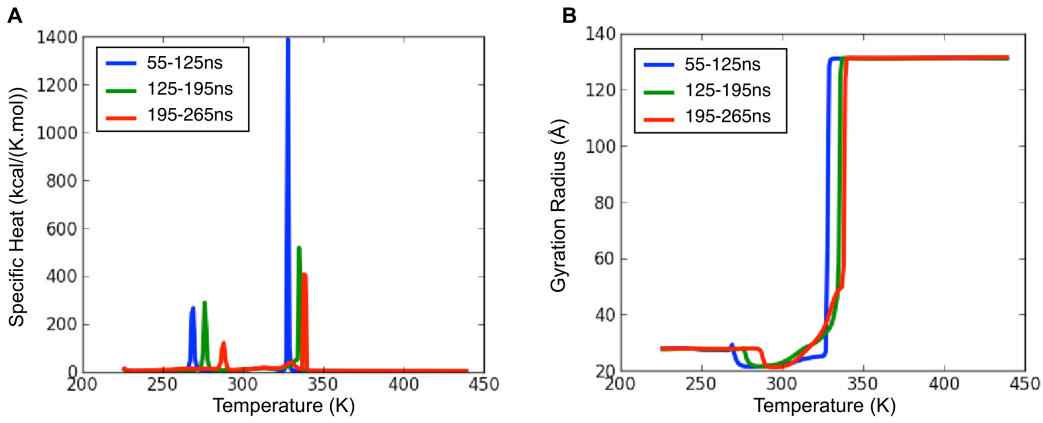

Figure 6.2: **Thermodynamic properties as a function of temperature for the 50-mer GNNQQNY.** The properties are calculated over three converged time intervals: 55-125 ns, 125-195 ns and 195-265 ns. The first 55 ns were discarded for lack of convergence. (a) Specific heat as a function of temperature for the three intervals. (b) Radius of gyration as a function of temperature for the same three intervals.

point) and from 6 to 9 strands between 287.4 K and 328.3 K (below the melting temperature). As for the 20-mer GNNQQNY system, the 50-mer prefers a parallel orientation of the β-strands with 60 to 70 % of the strands in a parallel orientation for all temperatures below melting (with an overall probability of being parallel of 2/3 below melting) with the possibility of forming fully parallel β-sheets of up to 7 strands (Fig. 6.3 (d) and (g)) and on average 5- or 6-strand long, while the fully antiparallel sheet is unfavorable and measures between 2 to 3 strands (Table 6.II). Knowing that the probability of forming parallel strands is 2/3 and consequently the probability of having antiparallel strands is 1/3, the probabilities of randomly obtaining fully parallel pentamer and hexamer sheets would be $(2/3)^4$=19.8 % and $(2/3)^5$=12.5 %, respectively, while the probabilities of obtaining a fully antiparallel sheet of 2 to 3 strands randomly would be $1/3$=33.3 % and $(1/3)^2$=11.1 % respectively. While our data show that the probability of forming the fully antiparallel 2- or 3-stranded sheet is generally far below this random probability, the probability of forming a 5- or 6-stranded fully parallel sheet is greater than 19.8 % and 12.5 % for almost all temperatures, going as high as 28 % (see Table 6.II), indicating a clear natural tendency of the system to form fully parallel sheets beyond pure random-

ness. We also note a preference for the strands to be in-register with each other within sheets with a high probability of having strands out-of-register by one residue, similarly to our results for the 20-mer (Table 6.II).

6.6.1 Diversity in the accessible morphologies

The eight most stable structures, presented in Fig. 6.3, were extracted from our simulations to be the final structures seeing no rearrangement of their β-strands during the last 2.25 ns of each run. Among these eight structures, we observe a fair amount of variability in the overall morphology, even though the different structure families are not as distinct as for our 20-mer system, and the main difference between the structures seems to be the overall shape of the aggregate. We classify structures as either elongated (Fig. 6.3 panels (a), (b), (c) and (d)) or spherical (panels (e), (f), (g) and (h)). All structures are multilayered, made of 5 to 6 β-sheets – with typical sheet lengths ranging from 6 to 11 strands mostly in-register – dominantly parallel, with the possibility of having one or two fully parallel sheets (Fig 6.3 panels (a), (b), (d) and (g)) whose lengths vary from 4 to up to 7 strands. At low temperatures (Fig 6.3 (a) to (d)), the radius of gyration of the structures is varying from 22.3 to 27.8 Å while at higher temperatures (panels (e) to (h)), the structures are more spherical, with a radius of gyration between 20.4 to 20.8 Å. However in each and every case, the structures are soluble and see their hydrophobic Tyr tightly packing in the core of the structure. This tight packing seems to be driving the β-sheets to interact as a first stabilization step and can also explain the preference for a parallel orientation of the β-sheets. To accomodate the Tyr packing, the sheets almost always adopt an overall twisted shape, that allows them to wrap around one another or even closely pack with an inter-sheet distance either in the 8-12 Å range (panels (b) and (c)) – in the order of the cross-β dry interface sheet separation – or in the 10-17 Å range (panels (b), (d), (e), (f) and (g)) – similar to the cross-β wet interface width. However, unlike the cross-β structure [32], the facing sheets are not antiparallel to each other but rather parallel.

6.6.1.1 Protofibril-like structure

Among the low-temperature structures, the morphology shown in panel (b) is of particular interest owing to its symmetry and tube-like morphology, resembling a protofibril structure [13, 197]. Out of the five sheets that compose it, four of them have lengths between 10 and 12 strands – which are on average slightly longer than sheets in other elongated structures – while the fifth sheet is 6-strands long. The wrapping of the four long sheets around each other creates a tube-like structure with the fifth sheet being intercalated at one extremity of the tube. The resulting pattern of Tyr packing is shown in Fig. 6.4 and is highly organized. The fictitious lines (shown in sticks) connecting the Tyr side chains (shown as spheres) demonstrate their alignment and the slightly twisted nature of the path connecting them, following the general shape of the associated β-sheets. In the hydrophobic core, the distance between the Tyr side chains centroids between each line is fluctuating between 6.5 and 7 Å, which is slightly higher than the Tyr-Tyr ring packing distance previously measured to be between 3.7 and 5.2 Å in all-atom simulations of GNNQQNY [23] due to the coarse-grained nature of our representation of the side chains. However, within β-sheets, the average distance between strands is ~ 4.9 Å, which is in accordance with the interstrand distance of 4.87 Å measured by Nelson *et al.* [32].

6.6.1.2 Interplay between side chains - side chains contacts and hydrogen bonding

We then investigate the role of contacts and hydrogen-bonds in maintaining the stability of the structures. For that purpose, we looked at the eight most stable structures and nine additional relatively stable structures, which are not undergoing any rearrangement of their β-strands over the last 1.5 ns of the simulations. Figure 6.5(a) displays the properties of these 17 stable structures in the contacts - hydrogen-bonds space (in green) in addition to the average properties (over the last 210 ns) for all temperatures below melting (in red) (For example T23 in panel (a) represents the average properties of structures at temperature T23 while F24 represents the properties of the final structure at temperature 24.). These average structures follow an 'S'-shaped path across the space

as the temperature increases, suggesting that the conformations first loose contacts and hydrogen-bonds upon heating up – up to T13 = 277.1 K, roughly the first transition – and then regain hydrogen-bonds while still loosing contacts up to T23 = 292.2 K. Because the average structures regain hydrogen-bonds after the first transition, corresponding to a decrease in the radius of gyration (Fig 6.2(b)), we conclude that, upon loosing contacts due to the temperature increase, the structures statistically prefer to adopt a more compact and spherical shape to optimize their secondary structure. In this same space, the 17 stable structures (in green) seem to be divided into two groups (except for structure H): one group with a high amount of contacts and low amount of hydrogen-bonds and a second group with a low amount of contacts and a high amount of hydrogen-bonds. This group division corresponds well to the shape distinction observed for the eight most stable structures, in the sense that elongated structures (A, C and D), with high radii of gyration, lie in the high contacts region while the more spherical structures (E, F and G) lie in the high hydrogen-bonds region. Some final structures (B, F10, F24, E, F, G in green) are lying in the higher hydrogen-bonds region compared to the average structures at the same temperature (namely at T02, T10, T24, T25, T26, T27 in red) and this is due to the fact that the final structures have had time to rearrange and thus increase their secondary structure more over time. The protofibril-like structure B also lies in the high hydrogen-bonds region, despite a high radius of gyration. As the only elongated high-gyration-radius structure with 5 β-sheets, while A, C and D have 6, structure B seems to favor less contacts between sheets and therefore more contacts towards the solvent and rather favors a higher β-sheet content. This can be explained by the fact that structure B has one less sheet than A, C and D and as a result some sheets are further apart and form less contacts as a result.

Figure 6.5(b) then shows the average structures (red) and the same 17 final stable structures (green) as a function of the number of hydrophobic contacts and of the number of hydrogen-bonds. The general trend of the average structures across the temperatures is different from the total contact 'S'-shaped path. From temperature T01 = 223.8 K to temperature T08 = 268.2 K, the average number of hydrophobic contacts in the structures increases despite the fact that the total number of contacts decreases (panel (a)).

This behavior indicates that hydrophilic Asn or Gln contacts uncouple to make room for hydrophobic Tyr-Tyr contacts as the system is heated up. This can be seen as a stability mechanism of the system as a reaction to the temperature increase. Then, from temperature T08 = 268.2 K to temperature T23 = 292.2 K, the hydrophobic contacts decrease, as the total amount of contacts does, letting the hydrogen-bonds slowly increase. Based on the distribution in panel (b), the final stable structures do not seem to be divided in two clear categories, suggesting that the hydrophobic contacts are not a key player in the morphology selection of the system but rather plays an overall role independent of the morphology.

6.7 Discussion

The structure of the 50-mer GNNQQNY is dominantly parallel with 60 to 70 % of the β-strands in a parallel orientation and about 50 % of the strands in-register. While the amount of strands in-register is comparable to the amount observed in our 20-mer GNNQQNY structures [64], the 50-mer has about 10 % more parallel strands than the 20-mer, proportionally, with the favorable presence of substantial fully-parallel β-sheets, demonstrating a natural tendency for the system to prefer a parallel orientation of the β-strands. This increase in parallel orientation from the 20-mer to the 50-mer confirms our previous observations of an increase in the amount of parallel orientation of β-strands when increasing the system's size [64], and we speculate that further increasing the system size would render the β-structure fully parallel, as suggested by the crystal structure of the fully-parallel and in-register GNNQQNY cross-β [32].

The multilayered twisted β-sheet nature of the final stable structures formed, inherent to all eight structures, resembles protofibril morphologies and has been already observed to be accessible to amyloid peptides and stable both experimentally [32, 47, 145] and numerically [13, 15, 22, 193, 197]. We make a distinction between two classes of morphologies for our final stable structures, based on their radii of gyration, and observe an elongated type with more contacts and less hydrogen-bonds and a more spherical type with less contacts and more hydrogen-bonds, the spherical shape being entropically

favored compared to the elongated shape upon temperature increase. Though the overall shape and amount of contacts and hydrogen-bonds separate those two types of structures, the two classes are not as distinct as the two general morphologies observed for the 20-mer system. At the 50-mer level, the structures are rather differentiated by their properties rather than their morphologies.

We also found that all the conformations accessible to the 50-mer are soluble, with their Tyrosine side chains pointing towards the core of the stucture, which is compatible with previous observations of soluble oligomers and protofibrils [11, 25, 41, 222–225]. Though, as we only observe one cluster in each and every conformation, we conclude that we have not yet reached a significant intermediate size to describe full oligomers and protofibrils.

Our attention was also focused on one particular stable structure, structure B, which has a tube-like 5-sheet multilayered morphology that we classify as being a protofibril-like structure and its sheets are longer, on average than most elongated structures. While its geometry allows for most sheets to wrap around each other, two sheets closely interact with an intersheet distance typical of the experimental cross-β dry interface width of 8.5 Å [32], despite being parallel to each other. Also, a typical interstrand distance in structure B is 4.9 Å, as found experimentally from the same study. In addition to these amyloid features, the quasi-symmetric wrapping of the β-sheets in a tube-like fashion creates a highly organized and stable Tyr packing motif in the center of the structure where the average Tyr-Tyr distance corresponds well to the Tyr-Tyr packing distance found by Gsponer *et al.* in all-atom simulations of GNNQQNY [23].

In all cases, the structures find many ways to stabilize under a temperature increase by either increasing the amount of hydrophobic contacts in their core or by increasing their secondary structure through an increase in the amount of hydrogen-bonds. The 6-sheet elongated structure seems optimal to favor contacts and therefore to stabilize the β-sheets together, which suggests a cooperativity between β-sheets upon formation of the structure. However, the 5-sheet protofibril-like structure B finds itself having much less contacts than most elongated structures and rather seems to favor secondary structure with longer sheets in the protofibril axis direction. We therefore conclude that

the protofibril-like structure B is an elongated structure that has overcome the tendency to form more contacts in order to favor growth in the protofibril axis direction while the other 6-sheet elongated structures have not matured enough yet and still need to reduce their contacts in order to grow. This result is in accordance with our previously obtained 2-sheet fibril-like structure for the 20-mer, which favored growth and secondary structures over contacts.

Finally, we did not observe an antiparallel orientation of the β-sheets relative to each other, as suggested by experimental studies of GNNQQNY [32]. This was also the case for the 20-mer and we claim that such a parallel arrangement of the β-sheets is not an artefact from our coarse-grained potential, as these 20-mer structures were also stable in all-atom simulations [64]. We thus argue that the flipping of β-sheets to adopt an antiparallel orientation with respect to one another in our protofibril-like structures would require the overcoming of a high free-energy barrier as the structures would then become insoluble, with some Tyr side chains protuding and contacting the solvent in an unfavorable fashion.

6.8 Conclusion

We have reported the results from our coarse-grained simulation of the 50-mer GNNQQNY starting from random conformations. We found that the 50-mer GNNQQNY oligomers are twisted multilayered protofibril-like structures and can either adopt an elongated or spherical shape depending on their tendency to either optimize contacts or hydrogen-bonds, respectively but the optimal protofibril structure is an elongated structure whose secondary structure has overcome and reduced side chains contacts in order to favor growth along the protofibril axis. They are soluble and see their Tyrosine side chains packing in an hydrophobic core. The structures reveal numerous structural features of the characteristic GNNQQNY amyloid cross-β and fibril structure, which correspond well with experimental observations except for the antiparallel orientation of β-sheets relative to each other. We suggest that the formation of such an antiparallel arrangement is a rare event associated with a high free-energy barrier of formation at

the 50-mer level and would require more monomers to make it a probable event. By comparing our results to a previous study of the 20-mer GNNQQNY, we observe that the aggregates gain parallel orientation of the β-strands as the system's size is increased. Though our system size has not reached a big enough scale to observe significant intermediates, our protofibril-like early aggregate, presenting a lot of features found in amyloid structures, provides great insight into the formation of early species during the first steps of aggregation and complement previous experimental and numerical studies.

6.9 Acknowledgements

This work was supported in part by the Natural Science and Engineering Research Council of Canada, the Canada Research Chair Foundation and the Fonds québécois de recherche en santé. We are grateful to Calcul Québec / Compute Canada for generous allocation of computer time.

	Population													
	223.8 K	237.7 K	249.2 K	255.9 K	260.1 K	272.1 K	280.1 K	287.4 K	295.7 K	310.1 K	316.8 K	328.3 K	342.2 K	356.2 K
Preferred number of β-sheets$^{(a)}$	5	6	6	5	6	6	6	5	5	6	6	6	6	0
4 preferred β-sheet lengths (in strands)$^{(b)}$	9,11,6,8	7,9,4,11	7,9,6,11	6,7,4,5	6,7,2,4	7,4,9,6	7,6,8,9	7,8,9,6	8,7,9,6	7,8,6,9	8,7,9,6	7,8,6,4	5,8,4,7	2
% parallel strands$^{(c)}$	63.6	62.4	63.7	65.7	64.0	67.4	67.0	66.9	66.0	67.9	66.9	66.8	64.8	47.65
% anti-parallel strands$^{(c)}$	36.4	37.6	36.2	34.2	35.8	32.5	32.5	32.9	33.8	32.0	33.0	33.0	35.1	26.75
% fully parallel β-sheets$^{(d)}$	19.4	20.0	18.2	22.8	22.9	26.5	21.2	18.6	20.0	18.9	20.3	19.9	27.6	0.0
Fully parallel β-sheet lengths (in strands)$^{(d)}$	6.4	5.9	5.8	5.4	5.4	4.8	5.2	5.2	5.8	5.1	5.0	4.6	4.6	0.0
% fully antiparallel β-sheets$^{(d)}$	2.8	7.0	9.7	7.1	10.0	7.7	4.1	3.0	2.6	4.4	5.1	5.6	8.9	31.3
Fully antiparallel β-sheet lengths (in strands)$^{(d)}$	3.1	2.7	2.8	3.0	2.6	2.8	2.9	3.1	3.2	3.2	3.3	2.9	3.3	2.0
% mixed β-sheets$^{(d)}$	77.8	73.0	72.1	70.1	67.0	65.8	74.7	78.4	77.4	76.7	74.6	74.5	63.6	68.7
β-sheet content (%)$^{(e)}$	77.1	75.6	74.7	71.9	69.5	63.5	59.9	58.7	55.8	51.7	49.6	45.1	25.2	3.3
% strands in-register/out-of-register$^{(f)}$	49/35	53/33	51/35	48/38	48/37	48/39	49/38	49/38	51/37	50/37	50/37	50/37	46/39	40/37
Radius of gyration R_g (in Å)$^{(g)}$	28.2	28.0	27.9	28.7	27.4	27.4	26.3	23.7	21.6	23.3	24.7	26.4	65.8	131.4

Table 6.II: **Structural characteristics for the 50-mer GNNQQNY as a function of temperature.** Temperatures above 356.2 K are not displayed here since they are disordered conformations populated mainly with random-coiled monomers and with no secondary structure. Not all temperatures are included between 260.1 K and 310.1 K. Because of the similarities of the properties, only one temperature out of 5 is shown for that interval of temperatures. The percentages are averages calculated over all the conformations obtained during the last 210 ns, where the system has converged. $^{(a)}$ The average most probable amount of β-sheets occurring in the structures. $^{(b)}$ The four β-sheet lengths with the highest average probability of occurrence (in strands), the first being the most probable and the fourth being the fourth most probable. $^{(c)}$ The average amount of parallel and antiparallel strands in the β-sheets formed. The sum of parallel and antiparallel strands does not always total 100 % as some strands might be in an undefined orientation, *i.e.* attached only by one hydrogen bond. $^{(d)}$ The average amount of fully parallel, fully antiparallel and mixed sheets and their corresponding average lengths (in strands). $^{(e)}$ The average amount of residues in a β conformation, as calculated by STRIDE. $^{(f)}$ The average amount of strands in-register and out-of-register (by one residue) in β-sheets. $^{(g)}$ The average radius of gyration in Å .

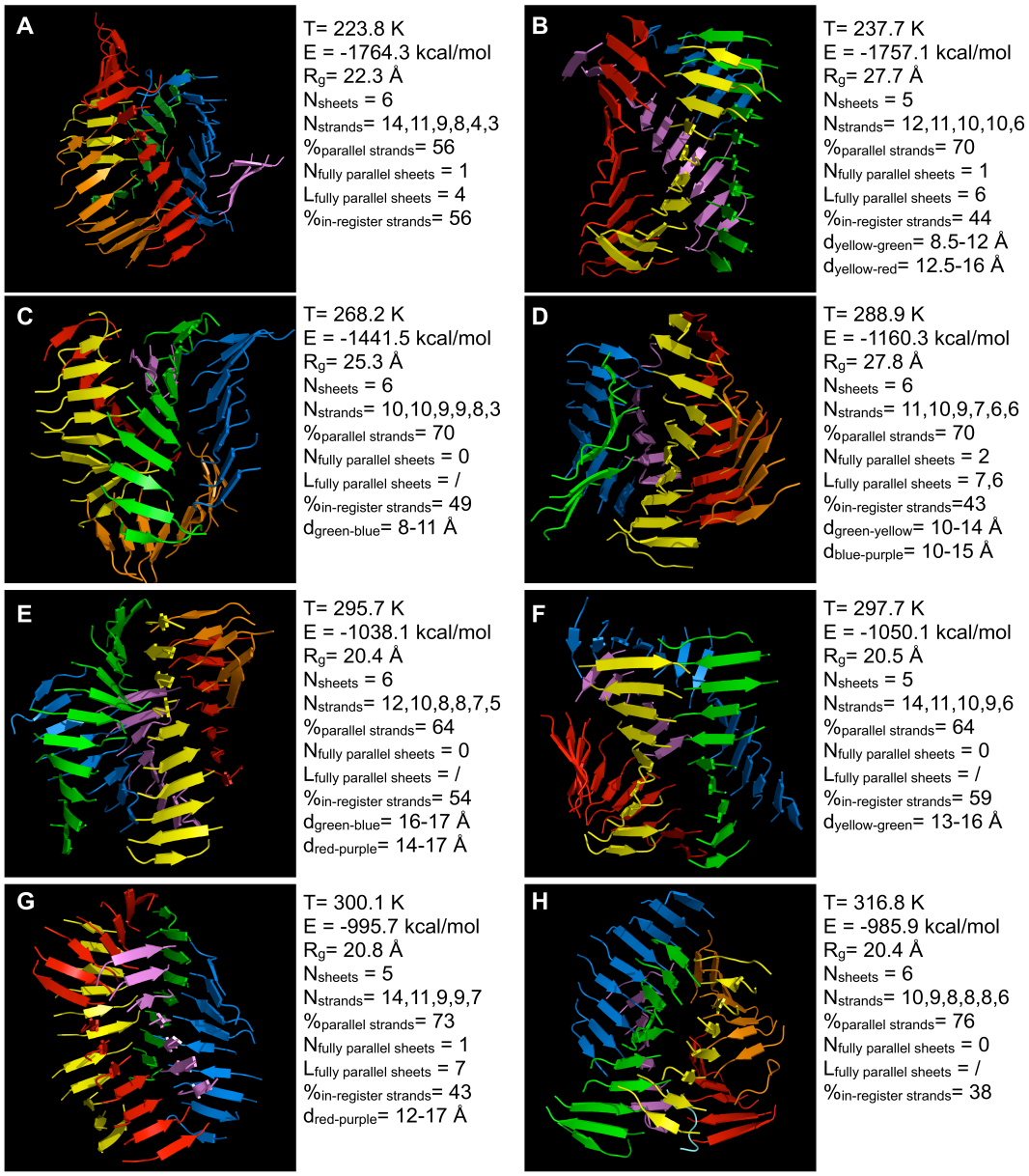

Figure 6.3: **Stable representative morphologies for the 50-mer GNNQQNY.** The legend accompanying each figure indicates: T, the temperature it was extracted from (in Kelvins); E, the energy (in kcal/mol) of the structure evaluated at temperature T01= 223.8 K; R_g, the radius of gyration (in Å); N_{sheets}, the number of β-sheets composing the structure; $N_{strands}$, the lengths of the β-sheets; $\%_{parallel\ strands}$, the percentage of parallel strands in the structure; $N_{fully\ parallel\ sheets}$, the number of fully parallel β-sheets in the structure, $L_{fully\ parallel\ sheets}$, the lengths of these fully parallel sheets. $d_{color1-color2}$, when indicated, is the intersheet distance (in Å) between sheet 1 of color 1 and sheet 2 of color 2.

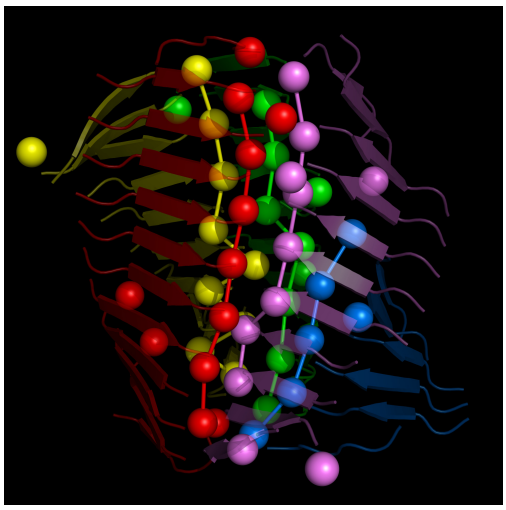

Figure 6.4: **Motif of hydrophobic Tyrosine packing for our protofibril-like structure.** Tyrosine side chains are shown as spheres. Each color corresponds to a different β-sheet (5 in total, shown in semi-transparent cartoon representation). Fictitious lines (shown as sticks) connect the Tyr side chains to highlight the general trace of the Tyr side chains for each sheet inside the hydrophobic core of the protofibril-like oligomer.

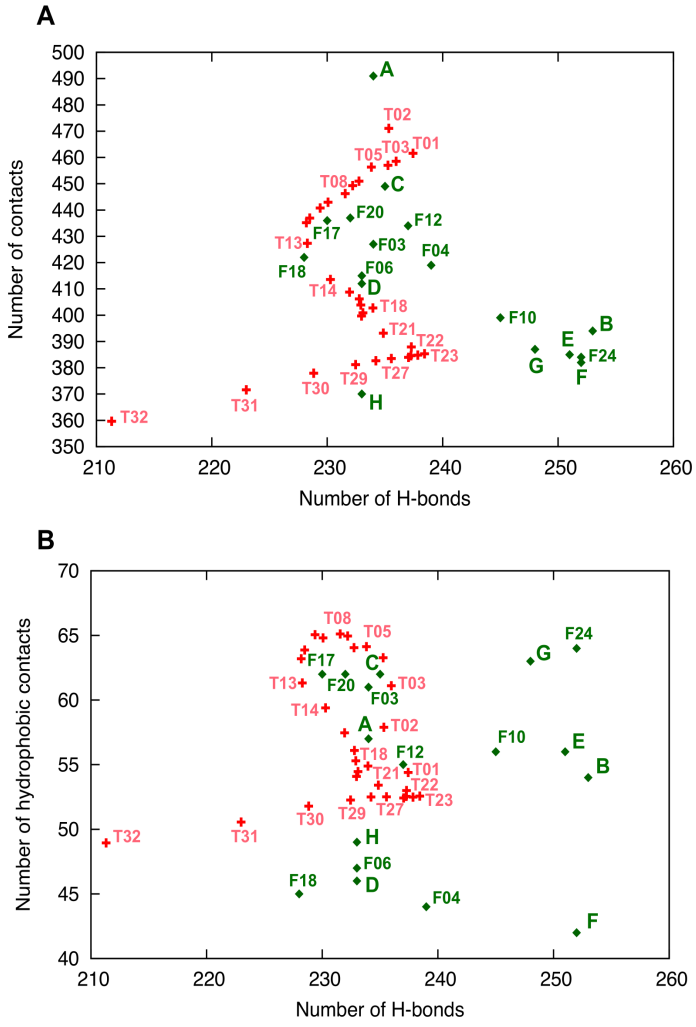

Figure 6.5: **Evolution of the structures across the contacts - hydrogen bonds space.** (a) Final stable structures (green) and average structures (red) as a function of their number of hydrogen bonds and total number of contacts. (b) Same as (a) but as a function of the number of hydrophobic contacts and hydrogen bonds. The one-letter labels in green represent the eight most stable final structures A, B, C, D, E, F, G and H extracted at temperatures T01, T02, T08, T21, T25, T26, T27 and T31 respectively. The FXX green labels represent other stable final structures where the XX indicates the index of the temperature it is extracted from. The TXX red labels represent the average structure (over the last 210 ns) where the XX indicates the index of the temperature it is extracted from (see Table 6.I for correspondance with temperatures in Kelvins). For clarity, not all labels are written.

CHAPITRE 7

CONCLUSIONS

7.1 Synthèse des résultats

Les grandes questions auxquelles cette thèse répond en partie sont: 1) Les oligomères sont-ils sur le chemin de formation des fibres? 2) Comment caractérise-t-on leur structure? 3) Que peut-on apprendre de leur processus d'agrégation? 4) Peut-on caractériser et reproduire la formation des protofibres? 5) Peut-on reproduire la formation de la structure cross-β typique des fibres et protofibres amyloïdes?

Dans leur ensemble, nos résultats répondent à ces questions et démontrent l'existence d'une grande diversité structurelle et mécanistique [2–5] lors de la formation d'agrégats amyloïdes pour la séquence GNNQQNY. Quelle que soit la taille du système, les structures présentent en effet un certain degré de polymorphisme et nous distinguons au moins deux types de chemin de formation pour le 20-mère et pour le 50-mère: un premier chemin par lequel les structures optimisent les contacts entre les chaînes latérales et un autre chemin par lequel les structures favorisent l'optimisation de leur structure secondaire par la formation de liaisons hydrogène. Les agrégats qui adoptent une forme 'protofibrillaire', pour le 20-mère comme pour le 50-mère, semblent favoriser l'optimisation de leur structure secondaire au détriment des contacts entre chaînes latérales. Nous concluons donc, en réponse aux questions 1), 2) et 4), que ces oligomères "protofibrillaires" sont sur le chemin de formation de la fibre amyloïde, alors que les oligomères favorisant les contacts peuvent être sur ou en dehors du chemin de formation de la fibre dépendamment de si, avec l'ajout de plus de monomères, l'élongation devient favorable ou non par rapport à l'optimisation des contacts. Cette compétition entre structure secondaire et contacts entre les chaînes latérales semble être, dans notre cas, l'élément déterminant du chemin de formation.

Nous notons une augmentation graduelle de la proportion de brins parallèles dans les feuillets-β lorsque la taille du système augmente du trimère au 20-mère et jusqu'au

50-mère et prédisons que cette proportion atteindra éventuellement 100 % si la taille du système était augmentée davantage, ce qui est compatible avec les observations expérimentales pour GNNQQNY [32] et plusieurs études numériques [17, 18]. En réponse à la question 5), la formation de structures cross-β a également été observée pour les systèmes 20-mère et 50-mère et les propriétés biochimiques correspondent aux données expérimentales, à l'exception de l'orientation antiparallèle des feuillets-β les uns par rapport aux autres. Nous argumentons cependant que le système préfère, dans un premier temps, cette orientation parallèle des feuillets-β et qu'une barrière d'énergie libre doit être franchie pour permettre l'orientation antiparallèle telle qu'observée par Nelson *et al.* [32], ce qui serait certainement facilité par la présence d'un plus grand nombre de monomères.

Pour répondre à la question 3), l'étude du processus d'agrégation du système 20-mère indique que l'auto-assemblage des peptides GNNQQNY est un phénomène complexe de nucléation dont le noyau critique est composé de 5 monomères, ce qui correspond bien à l'estimation de Nelson *et al.* pour GNNQQNY et qui se produit en trois étapes: d'abord, une phase de nucléation pendant laquelle des noyaux métastables tentent de se former jusqu'au franchissement d'une barrière d'énergie libre après quoi tous les monomères collapsent autour du noyau par le biais de forces hydrophobes pour former un agrégat amorphe désordonné (phase d'agrégation) qui va ensuite subir des réarrangements pour augmenter sa structure secondaire (phase de stabilisation). De tel phénomènes de nucléation ont été observés dans le cas de plusieurs protéines amyloïdes [103, 107, 112, 114, 124–129] et expliquent la croissance exponentielle des agrégats et ainsi leur résistance au système de contrôle cellulaire. En plus de la nucléation, nous observons un mécanisme de fusion et de fragmentation des oligomères. Ce mécanisme, quant à lui, indique que les agrégats formés par la nucléation ne sont pas statiques, mais se réarrangent pour optimiser soit leurs contacts entre chaînes latérales, soit leur structure secondaire, ce qui semble être un consensus dans chacune de nos études numériques.

Ainsi, par le biais d'une approche thermodynamique, mais aussi cinétique de l'étude des premières étapes d'agrégation du peptide GNNQQNY, notre méthode numérique

nous a permis d'accéder à des tailles de systèmes et à des temps suffisants pour observer des phénomènes d'auto-assemblage par nucléation et la possibilité d'au moins deux chemins de formations générant soit des oligomères sphériques, soit des oligomères protofibrillaires pouvant contenir une structure cross-β globalement en accord avec les travaux expérimentaux. Cependant, nos systèmes ne nous permettent de voir qu'un seul agrégat par configuration, ce qui veut dire qu'un seul événement de nucléation a lieu par configuration. De ce fait, nous concluons que nous n'avons pas encore atteint une taille suffisante pour observer des intermédiaires complets, que ce soit pour les oligomères comme pour les protofibres, et nous considérons ainsi que les systèmes que nous observons sont des jeunes oligomères, protofibrillaires ou non. Cependant nos résultats montrent clairement l'importance que la diversité de ces jeunes oligomères peut avoir sur la morphologie finale de la fibre et sur la cinétique d'agrégation et viennent compléter les précédentes études numériques sur l'agrégation de petits agrégats de GNNQQNY [14, 16–21, 23].

7.2 Validité des résultats

Bien que la robustesse du potentiel OPEP ait été démontrée [52, 162], certaines simplifications, comme la représentation des chaînes latérales et un solvant implicite, peuvent entraîner des artéfacts, comme des effets de sur-stabilisation ou une surestimation de certaines interactions, dans nos simulations. Il serait ainsi pertinent de coupler notre étude avec une étude expérimentale, malgré la difficulté actuelle d'observation expérimentale de processus d'agrégation à une si petite échelle. Nos résultats restent à être reproduits et doivent être comparés à d'autres travaux numériques et expérimentaux dans la mesure du possible. Cette même simplification des chaînes latérales nous empêche également de pouvoir caractériser l'interdigitation des chaînes latérales de la structure cross-β, ce qui a motivé un complément de travaux de dynamique tout-atome (Chapitre 4) afin de valider la stabilité des structures 20-mère trouvées avec OPEP et d'explorer la possibilité de formation de la structure cross-β. La rôle des molécules d'eau pourrait aussi être exploré par des méthodes tout-atome afin de déterminer l'importance du

solvant sur l'agrégation et la stabilisation des structures oligomériques amyloïdes et des travaux complémentaires à notre étude serait très pertinents. Un autre point intéressant à soulever est le degré de spécificité de nos résultats par rapport à la séquence GNNQQNY. Avons-nous caractérisé des propriétés morphologiques universelles applicables à tous les types d'oligomères amyloïdes ou avons-nous exploré des propriétés propres à GNNQQNY? Pour répondre à cette question, des modèles plus gros-grain, c'est-à-dire de plus basse résolution mais pour l'étude de systèmes de plus grandes tailles, permettraient de simuler des oligomères dont les constituants seraient démunis de toute information concernant une séquence particulière d'acides aminés et auraient, à la place, des propriétés globales propres aux protéines amyloïdes. Des modèles comme le modèle des tubes de Auer et al. [195] ou encore les "lattice models" comme celui de Li et al. [227] permettent une telle approche universelle.

Nos méthodes de simulations restent cependant compétitives et les conditions initiales de nos simulations, des structures complètement désordonnées, sont non biasées, et les tailles de systèmes simulés sont considérables pour la résolution atteinte, ce qui nous place parmi les leaders dans la communauté scientifique pour la caractérisation des phénomènes d'agrégation amyloïde. Puisqu'une telle résolution n'est pas encore accessible expérimentalement pour l'étude de la cinétique d'agrégation amyloïde, notre méthode numérique nous permet d'apporter une contribution significative en matière de pertinence expérimentale. Néanmoins, le domaine d'étude des phénomènes d'aggrégation amyloïde doit, pour faire des avancées significatives, combiner les résultats d'études expérimentales et numériques à divers échelles afin d'avoir une vision globale permettant de mieux définir les propriétés universelles des oligomères ainsi que leur rôle lors de la fibrillisation. Alors que les études numériques les plus récentes permettent tout juste d'atteindre des tailles de systèmes pertinentes du point de vue expérimental, les techniques expérimentales tendent toujours à améliorer leur résolution à un tel point que l'on peut espérer d'ici une dizaine d'années que les deux champs de recherche pourront bientôt s'entre-valider en ce qui concerne les événements microscopiques qui régissent la cinétique d'assemblage amyloïde à une résolution atomique.

7.3 Recommandations futures

Cette thèse répond en partie à plusieurs questions sur le processus d'agrégation du peptide GNNQQNY, mais entraîne tout autant de nouvelles questions. Notamment, pourrions-nous accéder à des tailles d'agrégats intermédiaires significatives pour un système de 100 monomères ou 200 monomères, qui sont des tailles de systèmes qu'OPEP pourra bientôt être capable de traiter? Comment se compare le processus d'agrégation du peptide GNNQQNY par rapport à d'autres peptides amyloïdogéniques dont les structures cross-β ont été déterminées par Sawaya *et al.* [47]? Une autre question intéressante serait de voir quel est l'effet de l'ajout d'acides aminés, normalement liés de part et d'autre de GNNQQNY dans la protéine Sup35, sur la cinétique d'agrégation et sur les morphologies des structures oligomériques.

À plus long terme, beaucoup d'autres aspects de la formation de structures amyloïdes peuvent être étudiés comme l'effet de l'environnement (pH, ions, glucose, etc.) sur la cinétique d'agrégation, problème auquel plusieurs études expérimentales et quelques études numériques se sont attaquées récemment [199, 228–233]. Des études d'arrimage d'anticorps sur les structures oligomériques stables pourraient aussi être très pertinentes afin de vérifier les propriétés des oligomères déjà observées expérimentalement [11]. Les structures oligomériques pourraient aussi être utilisées en vue de trouver des inhibiteurs de l'agrégation amyloïde ou bien insérées dans des membranes lipidiques pour mesurer leur effets potentiellement toxiques.

Ces recommandations, pour être réalisées avec succès, nécessiteraient une étroite collaboration d'expertises variées aussi bien expérimentales que théoriques et numériques dans le but de mieux comprendre l'agrégation des protéines amyloïdes et leurs effets dévastateurs.

BIBLIOGRAPHIE

[1] D. J. Selkoe, "Folding proteins in fatal ways.," *Nature*, vol. 426, no. 6968, pp. 900–904, 2003.

[2] R. Kodali and R. Wetzel, "Polymorphism in the intermediates and products of amyloid assembly.," *Current opinion in structural biology*, vol. 17, no. 1, pp. 48–57, 2007.

[3] A. T. Petkova, R. D. Leapman, Z. Guo, W.-M. Yau, M. P. Mattson, and R. Tycko, "Self-propagating, molecular-level polymorphism in Alzheimer's beta-amyloid fibrils.," *Science (New York, N.Y.)*, vol. 307, pp. 262–265, Jan. 2005.

[4] R. Wetzel, S. Shivaprasad, and A. D. Williams, "Plasticity of amyloid fibrils.," *Biochemistry*, vol. 46, no. 1, pp. 1–10, 2007.

[5] A. K. Paravastu, R. D. Leapman, W.-M. Yau, and R. Tycko, "Molecular structural basis for polymorphism in Alzheimer's β-amyloid fibrils," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 47, pp. 18349–18354, 2008.

[6] J. I. Guijarro, M. Sunde, J. A. Jones, I. D. Campbell, and C. M. Dobson, "Amyloid fibril formation by an SH3 domain.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 8, pp. 4224–4228, 1998.

[7] M. Stefani and C. M. Dobson, "Protein aggregation and aggregate toxicity : new insights into protein folding, misfolding diseases and biological evolution.," *Journal of molecular medicine (Berlin, Germany)*, vol. 81, no. 11, pp. 678–699, 2003.

[8] V. N. Uversky and A. L. Fink, "Conformational constraints for amyloid fibrillation : the importance of being unfolded.," *Biochimica et biophysica acta*, vol. 1698, no. 2, pp. 131–153, 2004.

[9] M. M. Bucciantini, E. E. Giannoni, F. F. Chiti, F. F. Baroni, L. L. Formigli, J. J. Zurdo, N. N. Taddei, G. G. Ramponi, C. M. C. Dobson, and M. M. Stefani, "Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases.," *Nature*, vol. 416, no. 6880, pp. 507–511, 2002.

[10] S. M. Butterfield and H. A. Lashuel, "Amyloidogenic protein-membrane interactions : mechanistic insight from model systems.," *Angewandte Chemie (International ed. in English)*, vol. 49, no. 33, pp. 5628–5654, 2010.

[11] R. R. Kayed, E. E. Head, J. L. J. Thompson, T. M. T. McIntire, S. C. S. Milton, C. W. C. Cotman, and C. G. C. Glabe, "Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis.," *Audio, Transactions of the IRE Professional Group on*, vol. 300, no. 5618, pp. 486–489, 2003.

[12] J. Park, B. Kahng, and W. Hwang, "Thermodynamic selection of steric zipper patterns in the amyloid cross-beta spine.," *PLoS Computational Biology*, vol. 5, no. 9, p. e1000492, 2009.

[13] J. T. Berryman, S. E. Radford, and S. A. Harris, "Thermodynamic description of polymorphism in Q- and N-rich peptide aggregates revealed by atomistic simulation.," *Biophysical journal*, vol. 97, no. 1, pp. 1–11, 2009.

[14] J. J. Zheng, H. H. Jang, B. B. Ma, and R. R. Nussinov, "Annular structures as intermediates in fibril formation of Alzheimer Abeta17-42.," *The journal of physical chemistry. B*, vol. 112, no. 22, pp. 6856–6865, 2008.

[15] X. Periole, A. Rampioni, M. Vendruscolo, and A. E. Mark, "Factors that affect the degree of twist in beta-sheet structures : a molecular dynamics simulation study of a cross-beta filament of the GNNQQNY peptide.," *The journal of physical chemistry. B*, vol. 113, no. 6, pp. 1728–1737, 2009.

[16] J. Wang, C. Tan, H.-F. Chen, and R. Luo, "All-atom computer simulations of amyloid fibrils disaggregation.," *Biophysical journal*, vol. 95, no. 11, pp. 5037–5047, 2008.

[17] M. Meli, G. Morra, and G. Colombo, "Investigating the mechanism of peptide aggregation : insights from mixed monte carlo-molecular dynamics simulations.," *Biophysical journal*, vol. 94, no. 11, pp. 4414–4426, 2008.

[18] Z. Zhang, H. Chen, H. Bai, and L. Lai, "Molecular dynamics simulations on the oligomer-formation process of the GNNQQNY peptide from yeast prion protein Sup35.," *Biophysical journal*, vol. 93, no. 5, pp. 1484–1492, 2007.

[19] B. Strodel, C. S. Whittleston, and D. J. Wales, "Thermodynamics and kinetics of aggregation for the GNNQQNY peptide.," *Journal of the American Chemical Society*, vol. 129, no. 51, pp. 16005–16014, 2007.

[20] A. S. Reddy, M. Chopra, and J. J. de Pablo, "GNNQQNY-Investigation of Early Steps during Amyloid Formation.," *Biophysical journal*, vol. 98, no. 6, pp. 1038–1045, 2010.

[21] L. Vitagliano, L. Esposito, C. Pedone, and A. De Simone, "Stability of single sheet GNNQQNY aggregates analyzed by replica exchange molecular dynamics : antiparallel versus parallel association.," *Biochemical and biophysical research communications*, vol. 377, no. 4, pp. 1036–1041, 2008.

[22] L. Esposito, C. Pedone, and L. Vitagliano, "Molecular dynamics analyses of cross-beta-spine steric zipper models : beta-sheet twisting and aggregation.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 31, pp. 11533–11538, 2006.

[23] J. Gsponer, U. Haberthür, and A. Caflisch, "The role of side-chain interactions in the early steps of aggregation : Molecular dynamics simulations of an amyloid-forming peptide from the yeast prion Sup35.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 9, pp. 5154–5159, 2003.

[24] E. Herczenik and M. F. B. G. Gebbink, "Molecular and cellular aspects of protein misfolding and disease," *The FASEB Journal*, vol. 22, no. 7, pp. 2115–2133, 2008.

[25] F. Chiti and C. M. Dobson, "Protein misfolding, functional amyloid, and human disease.," *Biochemistry*, vol. 75, pp. 333–366, 2006.

[26] C. A. Ross and M. A. Poirier, "Protein aggregation and neurodegenerative disease," *Nature medicine*, vol. 10, no. 7, pp. S10–S17, 2004.

[27] M. Balbirnie, R. Grothe, and D. S. Eisenberg, "An amyloid-forming peptide from the yeast prion Sup35 reveals a dehydrated beta-sheet structure for amyloid.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 5, pp. 2375–2380, 2001.

[28] C. S. Goldsbury, S. Wirtz, S. Müller, S. Sunderji, P. Wicki, U. Aebi, and P. Frey, "Studies on the in vitro assembly of Abeta1-40 : implications for the search for a beta fibril formation inhibitors.," *Journal of structural biology*, vol. 130, no. 2-3, pp. 217–231, 2000.

[29] R. Jansen, W. Dzwolak, and R. Winter, "Amyloidogenic self-assembly of insulin aggregates probed by high resolution atomic force microscopy.," *Biophysical journal*, vol. 88, no. 2, pp. 1344–1353, 2005.

[30] A. J. Geddes, K. D. Parker, E. D. T. Atkins, and E. Beighton, ""Cross-β" conformation in proteins," *Journal of molecular biology*, vol. 32, pp. 343–358, Mar. 1968.

[31] M. Sunde, "Common core structure of amyloid fibrils by synchrotron X-ray diffraction," *Journal of molecular biology*, vol. 273, no. 3, pp. 729–739, 1997.

[32] R. Nelson, M. R. Sawaya, M. Balbirnie, A. Ø. Madsen, C. Riekel, R. Grothe, and D. Eisenberg, "Structure of the cross-β spine of amyloid-like fibrils," *Nature*, vol. 435, no. 7043, pp. 773–778, 2005.

[33] D. M. Walsh, D. Hartley, Y. Kusumoto, Y. Fezoui, M. Condron, A. Lomakin, G. B. Benedek, D. Selkoe, and D. B. Teplow, "Amyloid β-protein fibrillogenesis," *Journal of Biological Chemistry*, vol. 274, no. 36, pp. 25945–25952, 1999.

[34] G. Habicht, C. Haupt, R. P. Friedrich, P. Hortschansky, C. Sachse, J. Meinhardt, K. Wieligmann, G. P. Gellermann, M. Brodhun, J. Götz, K.-J. Halbhuber, C. Röcken, U. Horn, and M. Fändrich, "Directed selection of a conformational antibody domain that prevents mature amyloid fibril formation by stabilizing Abeta protofibrils.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 49, pp. 19232–19237, 2007.

[35] A. Eckert, S. Hauptmann, I. Scherping, J. Meinhardt, V. Rhein, S. Dröse, U. Brandt, M. Fändrich, W. E. Müller, and J. Götz, "Oligomeric and fibrillar species of beta-amyloid (Abeta42) both impair mitochondrial function in P301L tau transgenic mice.," *Journal of molecular medicine (Berlin, Germany)*, vol. 86, no. 11, pp. 1255–1267, 2008.

[36] S. Chimon, M. A. Shaibat, C. R. Jones, D. C. Calero, B. Aizezi, and Y. Ishii, "Evidence of fibril-like β-sheet structures in a neurotoxic amyloid intermediate of Alzheimer's β-amyloid.," *Nature structural & molecular biology*, vol. 14, no. 12, pp. 1157–1164, 2007.

[37] M. Ahmed, J. Davis, D. Aucoin, T. Sato, S. Ahuja, S. Aimoto, J. I. Elliott, W. E. Van Nostrand, and S. O. Smith, "Structural conversion of neurotoxic amyloid-beta(1-42) oligomers to fibrils.," *Nature structural & molecular biology*, vol. 17, no. 5, pp. 561–567, 2010.

[38] S. Campioni, B. Mannini, M. Zampagni, A. Pensalfini, C. Parrini, E. Evangelisti, A. Relini, M. Stefani, C. M. Dobson, C. Cecchi, and F. Chiti, "A causative link between the structure of aberrant protein oligomers and their toxicity.," *Nature chemical biology*, vol. 6, no. 2, pp. 140–147, 2010.

[39] A. Sandberg, L. M. Luheshi, S. Söllvander, T. Pereira de Barros, B. Macao, T. P. J. Knowles, H. Biverstål, C. Lendel, F. Ekholm-Petterson, A. Dubnovitsky, L. Lannfelt, C. M. Dobson, and T. Härd, "Stabilization of neurotoxic Alzheimer amyloid-beta oligomers by protein engineering.," *Proceedings of the National Academy*

of Sciences of the United States of America, vol. 107, no. 35, pp. 15595–15600, 2010.

[40] M. Fändrich, "Oligomeric intermediates in amyloid formation : structure determination and mechanisms of toxicity.," *Journal of molecular biology*, vol. 421, no. 4-5, pp. 427–440, 2012.

[41] C. Haass and D. J. Selkoe, "Soluble protein oligomers in neurodegeneration : lessons from the Alzheimer's amyloid beta-peptide.," *Nature reviews. Molecular cell biology*, vol. 8, no. 2, pp. 101–112, 2007.

[42] R. Nelson and D. Eisenberg, "Recent atomic models of amyloid fibril structure.," *Current opinion in structural biology*, vol. 16, no. 2, pp. 260–265, 2006.

[43] J. C. Rochet and P. T. Lansbury, "Amyloid fibrillogenesis : themes and variations.," *Current opinion in structural biology*, vol. 10, no. 1, pp. 60–68, 2000.

[44] M. J. Thompson, S. A. Sievers, J. Karanicolas, M. I. Ivanova, D. Baker, and D. Eisenberg, "The 3D profile method for identifying fibril-forming segments of proteins.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 11, pp. 4074–4078, 2006.

[45] P. K. Teng and D. Eisenberg, "Short protein segments can drive a non-fibrillizing protein into the amyloid state.," *Protein engineering, design & selection : PEDS*, vol. 22, no. 8, pp. 531–536, 2009.

[46] A. H. DePace, A. Santoso, P. Hillner, and J. S. Weissman, "A critical role for amino-terminal glutamine/asparagine repeats in the formation and propagation of a yeast prion.," *Cell*, vol. 93, no. 7, pp. 1241–1252, 1998.

[47] M. R. Sawaya, S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. W. Wiltzius, H. T. McFarlane, A. Ø. Madsen, C. Riekel, and D. Eisenberg, "Atomic structures of amyloid cross-beta spines reveal varied steric zippers.," *Nature*, vol. 447, no. 7143, pp. 453–457, 2007.

[48] P. C. A. van der Wel, J. R. Lewandowski, and R. G. Griffin, "Solid-state NMR study of amyloid nanocrystals and fibrils formed by the peptide GNNQQNY from yeast prion protein Sup35p.," *Journal of the American Chemical Society*, vol. 129, no. 16, pp. 5117–5130, 2007.

[49] P. C. A. van der Wel, J. R. Lewandowski, and R. G. Griffin, "Structural characterization of GNNQQNY amyloid fibrils by magic angle spinning NMR.," *Biochemistry*, vol. 49, no. 44, pp. 9457–9469, 2010.

[50] S. B. Andrey, M. L. Chan, and W. P. Power, "HRMAS 1H NMR conformational study of the resin-bound amyloid-forming peptide GNNQQNY from the yeast prion Sup35.," *The journal of physical chemistry. A*, vol. 114, no. 10, pp. 3457–3465, 2010.

[51] K. E. Marshall, M. R. Hicks, T. L. Williams, S. V. Hoffmann, A. Rodger, T. R. Dafforn, and L. C. Serpell, "Characterizing the assembly of the Sup35 yeast prion fragment, GNNQQNY : structural changes accompany a fiber-to-crystal switch.," *Biophysical journal*, vol. 98, no. 2, pp. 330–338, 2010.

[52] J. Maupetit, P. Tuffery, and P. Derreumaux, "A coarse-grained protein force field for folding and structure prediction," *Proteins : Structure, Function, and Bioinformatics*, vol. 69, no. 2, pp. 394–408, 2007.

[53] P. Derreumaux, "From polypeptide sequences to structures using Monte Carlo simulations and an optimized potential," *The Journal of chemical physics*, vol. 111, no. 5, pp. 2301–2310, 1999.

[54] P. Derreumaux, "Generating Ensemble Averages for Small Proteins from Extended Conformations by Monte Carlo Simulations," *Physical review letters*, vol. 85, no. 1, pp. 206–209, 2000.

[55] F. Forcellino and P. Derreumaux, "Computer simulations aimed at structure prediction of supersecondary motifs in proteins," *Proteins : Structure, Function, and Bioinformatics*, vol. 45, no. 2, pp. 159–166, 2001.

[56] P. Derreumaux, "Evidence that the 127-164 region of prion proteins has two equi-energetic conformations with beta or alpha features.," *Biophysical journal*, vol. 81, no. 3, pp. 1657–1665, 2001.

[57] P. Derreumaux, "Insight into protein topology from Monte Carlo simulations," *The Journal of chemical physics*, vol. 117, no. 7, pp. 3499–3503, 2002.

[58] P. Derreumaux and N. Mousseau, "Coarse-grained protein molecular dynamics simulations.," *The Journal of chemical physics*, vol. 126, no. 2, p. 025101, 2007.

[59] P. Derreumaux and N. Mousseau, "Coarse-grained protein molecular dynamics simulations.," *The Journal of chemical physics*, vol. 126, no. 2, pp. 025101–025101, 2007.

[60] W. Song, G. Wei, N. Mousseau, and P. Derreumaux, "Self-assembly of the beta2-microglobulin NHVTLSQ peptide using a coarse-grained protein model reveals a beta-barrel species.," *The journal of physical chemistry. B*, vol. 112, no. 14, pp. 4410–4418, 2008.

[61] Y. Lu, P. Derreumaux, Z. Guo, N. Mousseau, and G. Wei, "Thermodynamics and dynamics of amyloid peptide oligomerization are sequence dependent," *Proteins : Structure, Function, and Bioinformatics*, vol. 75, no. 4, pp. 954–963, 2009.

[62] Y. Chebaro, X. Dong, R. Laghaei, P. Derreumaux, and N. Mousseau, "Replica Exchange Molecular Dynamics Simulations of Coarse-grained Proteins in Implicit Solvent," *The journal of physical chemistry. B*, vol. 113, no. 1, pp. 267–274, 2009.

[63] R. Laghaei, N. Mousseau, and G. Wei, "Effect of the disulfide bond on the monomeric structure of human amylin studied by combined Hamiltonian and temperature replica exchange molecular dynamics simulations.," *The journal of physical chemistry. B*, vol. 114, no. 20, pp. 7071–7077, 2010.

[64] J. Nasica-Labouze, M. Meli, P. Derreumaux, G. Colombo, and N. Mousseau, "A Multiscale Approach to Characterize the Early Aggregation Steps of the Amyloid-Forming Peptide GNNQQNY from the Yeast Prion Sup-35," *PLoS Computational Biology*, vol. 7, no. 5, p. e1002051, 2011.

[65] S. Côté, P. Derreumaux, and N. Mousseau, "Distinct Morphologies for Amyloid Beta Protein Monomer : Aβ 1–40, Aβ 1–42, and Aβ 1–40(D23N)," *Journal of Chemical Theory and Computation*, vol. 7, no. 8, pp. 2584–2592, 2011.

[66] S. Côté, R. Laghaei, P. Derreumaux, and N. Mousseau, "Distinct Dimerization for Various Alloforms of the Amyloid-Beta Protein : Aβ 1–40, Aβ 1–42, and Aβ 1–40(D23N)," *The journal of physical chemistry. B*, vol. 116, pp. 4043–4055, Apr. 2012.

[67] N. Mousseau, P. Derreumaux, G. T. Barkema, and R. Malek, "Sampling activated mechanisms in proteins with the activation-relaxation technique.," *Journal of molecular graphics & modelling*, vol. 19, no. 1, pp. 78–86, 2001.

[68] G. Wei, P. Derreumaux, and N. Mousseau, "Sampling the complex energy landscape of a simple β-hairpin," *The Journal of chemical physics*, vol. 119, no. 13, pp. 6403–6406, 2003.

[69] G. Wei, N. Mousseau, and P. Derreumaux, "Complex folding pathways in a simple β-hairpin," *Proteins : Structure, Function, and Bioinformatics*, vol. 56, no. 3, pp. 464–474, 2004.

[70] G. Wei, N. Mousseau, and P. Derreumaux, "Sampling the self-assembly pathways of KFFE hexamers.," *Biophysical journal*, vol. 87, no. 6, pp. 3648–3656, 2004.

[71] N. Mousseau and P. Derreumaux, "Exploring the Early Steps of Amyloid Peptide Aggregation by Computers," *Accounts of chemical research*, vol. 38, no. 11, pp. 885–891, 2005.

[72] L. Dupuis and N. Mousseau, "Understanding the EF-hand closing pathway using non-biased interatomic potentials," *The Journal of chemical physics*, vol. 136, no. 3, p. 035101, 2012.

[73] Z. Feng, G. Gilliland, T. N. Bhat, and H. Weissig, "The protein data bank," *Nucleic acids Research*, 2000.

[74] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, "The Amber biomolecular simulation programs.," *Journal Of Computational Chemistry*, vol. 26, no. 16, pp. 1668–1688, 2005.

[75] G. N. Ramachandran and V. Sasisekharan, "Conformation of polypeptides and proteins," *Advances in Protein Chemistry*, vol. 23, pp. 283–438, 1968.

[76] L. Gentilucci, R. De Marco, and L. Cerisoli, "Chemical modifications designed to improve peptide stability : incorporation of non-natural amino acids, pseudopeptide bonds, and cyclization.," *Current pharmaceutical design*, vol. 16, no. 28, pp. 3185–3203, 2010.

[77] D. D. Eramian, M.-y. M. Shen, D. D. Devos, F. F. Melo, A. A. Sali, and M. A. M. Marti-Renom, "A composite score for predicting errors in protein structure models.," *Protein science : a publication of the Protein Society*, vol. 15, no. 7, pp. 1653–1666, 2006.

[78] M.-Y. Shen and A. Sali, "Statistical potential for assessment and prediction of protein structures.," *Protein science : a publication of the Protein Society*, vol. 15, no. 11, pp. 2507–2524, 2006.

[79] K. T. Simons, C. Kooperberg, E. Huang, and D. Baker, "Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions.," *Journal of molecular biology*, vol. 268, no. 1, pp. 209–225, 1997.

[80] K. T. Simons, R. Bonneau, I. Ruczinski, and D. Baker, "Ab initio protein structure prediction of CASP III targets using ROSETTA.," *Proteins : Structure, Function, and Bioinformatics*, vol. Suppl 3, pp. 171–176, 1999.

[81] K. M. S. Misura, D. Chivian, C. A. Rohl, D. E. Kim, and D. Baker, "Physically realistic homology models built with ROSETTA can be more accurate than their templates.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 14, pp. 5361–5366, 2006.

[82] H. H. Zhou and Y. Y. Zhou, "Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction.," *Protein science : a publication of the Protein Society*, vol. 11, no. 11, pp. 2714–2726, 2002.

[83] F. Melo, R. Sánchez, and A. Sali, "Statistical potentials for fold assessment.," *Protein science : a publication of the Protein Society*, vol. 11, no. 2, pp. 430–448, 2002.

[84] X. Dong, W. Chen, N. Mousseau, and P. Derreumaux, "Energy landscapes of the monomer and dimer of the Alzheimer's peptide Aβ(1–28)," *The Journal of chemical physics*, vol. 128, no. 12, pp. 125108–125108–10, 2008.

[85] G. Wei, N. Mousseau, and P. Derreumaux, "Exploring the energy landscape of proteins : A characterization of the activation-relaxation technique," *The Journal of chemical physics*, vol. 117, no. 24, p. 11379, 2002.

[86] N. M. P. D. Guanghong Wei, "Computational Simulations of the Early Steps of Protein Aggregation," *Prion*, vol. 1, no. 1, p. 3, 2007.

[87] R. Laghaei, N. Mousseau, and G. Wei, "Structure and thermodynamics of amylin dimer studied by Hamiltonian-temperature replica exchange molecular dynamics simulations.," *The journal of physical chemistry. B*, vol. 115, no. 12, pp. 3146–3154, 2011.

[88] R. Laghaei and N. Mousseau, “Spontaneous formation of polyglutamine nanotubes with molecular dynamics simulations.,” *The Journal of chemical physics*, vol. 132, no. 16, p. 165102, 2010.

[89] L. Dupuis and N. Mousseau, “Holographic multiscale method used with nonbiased atomistic forcefields for simulation of large transformations in protein,” *Journal of Physics : Conference Series*, vol. 341, no. 1, p. 012015, 2012.

[90] R. W. Pastor, “Techniques and applications of Langevin dynamics simulations,” *The Molecular Dynamics of Liquid Crystals*, 1994.

[91] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, “GROMACS 4 : Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation,” *Journal of Chemical Theory and Computation*, vol. 4, no. 3, pp. 435–447, 2008.

[92] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, “A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules : Application to Small Water Clusters,” 1981.

[93] H. C. Andersen, “Rattle : A "Velocity" Version of the Shake Algorithm for Molecular Dynamics Calculations,” *Journal of Computational Physics*, vol. 52, p. 24, 1983.

[94] K. Anton and F. B. Hess, “Improving Efficiency of Large Time-scale Molecular Dynamics Simulations of Hydrogen-rich Systems,” *Journal of Computational Chemistry*, vol. 20, no. 8, pp. 786–798, 1999.

[95] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, “Molecular dynamics with coupling to an external bath,” *The Journal of chemical physics*, vol. 81, no. 8, pp. 3684–3690, 1984.

[96] T. Schlick, “Molecular Modeling and Simulation : An Interdisciplinary Guide,” *Molecular Modeling and Simulation : An Interdisciplinary Guide by Tamar Schlick. New York : Springer*, 2010.

[97] S. Nosé, "A unified formulation of the constant temperature molecular dynamics methods," *The Journal of chemical physics*, vol. 81, no. 1, pp. 511–519, 1984.

[98] K. A. Dill and H. S. Chan, "From Levinthal to pathways to funnels.," *Nature structural biology*, vol. 4, no. 1, pp. 10–19, 1997.

[99] Y. Sugita and Y. Okamoto, "Replica-exchange molecular dynamics method for protein folding," *Chemical Physics Letters*, vol. 314, pp. 141–151, 1999.

[100] W. Kabsch and C. Sander, "Dictionary of protein secondary structure : pattern recognition of hydrogen-bonded and geometrical features," *Biopolymers*, vol. 22, no. 12, pp. 2577–2637, 1983.

[101] L. Yang, Q. Shao, and Y. Q. Gao, "Comparison between integrated and parallel tempering methods in enhanced sampling simulations.," *The Journal of chemical physics*, vol. 130, no. 12, pp. 124111–124111, 2009.

[102] D. Frishman and P. Argos, "Knowledge-based protein secondary structure assignment.," *Proteins : Structure, Function, and Bioinformatics*, vol. 23, no. 4, pp. 566–579, 1995.

[103] A. Lomakin, D. S. Chung, G. B. Benedek, D. A. Kirschner, and D. B. Teplow, "On the nucleation and growth of amyloid beta-protein fibrils : detection of nuclei and quantitation of rate constants.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 3, pp. 1125–1129, 1996.

[104] T. Ban, K. Yamaguchi, and Y. Goto, "Direct observation of amyloid fibril growth, propagation, and adaptation.," *Accounts of chemical research*, vol. 39, no. 9, pp. 663–670, 2006.

[105] Y. Liang, D. G. Lynn, and K. M. Berland, "Direct observation of nucleation and growth in amyloid self-assembly.," *JACS communications*, vol. 132, pp. 6306–6308, May 2010.

[106] T. P. J. Knowles, D. A. White, A. R. Abate, J. J. Agresti, S. I. A. Cohen, R. A. Sperling, E. J. De Genst, C. M. Dobson, and D. A. Weitz, "Observation of spatial propagation of amyloid assembly from single nuclei," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 36, pp. 14746–14751, 2011.

[107] F. A. Ferrone, J. Hofrichter, and W. A. Eaton, "Kinetics of sickle hemoglobin polymerization : Ii. a double nucleation mechanism," *Journal of Molecular Biology*, vol. 183, no. 4, pp. 611 – 631, 1985.

[108] F. Ferrone, "Analysis of protein aggregation kinetics," *Methods in Enzymology*, vol. 309, pp. 256–274, 1999.

[109] T. P. J. Knowles, C. A. Waudby, G. L. Devlin, S. I. A. Cohen, A. Aguzzi, M. Vendruscolo, E. M. Terentjev, M. E. Welland, and C. M. Dobson, "An analytical solution to the kinetics of breakable filament assembly.," *Science (New York, N.Y.)*, vol. 326, no. 5959, pp. 1533–1537, 2009.

[110] N. Ferguson, J. Berriman, M. Petrovich, T. D. Sharpe, J. T. Finch, and A. R. Fersht, "Rapid amyloid fiber formation from the fast-folding WW domain FBP28.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 17, pp. 9814–9819, 2003.

[111] J. T. Jarrett and P. T. Lansbury, "Seeding "one-dimensional crystallization" of amyloid : a pathogenic mechanism in Alzheimer's disease and scrapie ?," *Cell*, vol. 73, no. 6, pp. 1055–1058, 1993.

[112] S. R. Collins, A. Douglass, R. D. Vale, and J. S. Weissman, "PLoS Biology : Mechanism of Prion Propagation : Amyloid Growth Occurs by Monomer Addition," *PLoS biology*, vol. 2, no. 10, p. e321, 2004.

[113] N. Carulla, G. L. Caddy, D. R. Hall, J. Zurdo, M. Gairí, M. Feliz, E. Giralt, C. V. Robinson, and C. M. Dobson, "Molecular recycling within amyloid fibrils.," *Nature*, vol. 436, no. 7050, pp. 554–558, 2005.

[114] W.-F. Xue, S. W. Homans, and S. E. Radford, "Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 26, pp. 8926–8931, 2008.

[115] F. Librizzi and C. Rischel, "The kinetic behavior of insulin fibrillation is determined by heterogeneous nucleation pathways.," *Protein science : a publication of the Protein Society*, vol. 14, no. 12, pp. 3129–3134, 2005.

[116] T. Arvinte, A. Cudd, and A. Drake, "The structure and mechanism of formation of human calcitonin fibrils.," *Journal of Biological Chemistry*, vol. 268, no. 9, pp. 6415–6422, 1993.

[117] S. B. Padrick and A. D. Miranker, "Islet amyloid : phase partitioning and secondary nucleation are central to the mechanism of fibrillogenesis.," *Biochemistry*, vol. 41, no. 14, pp. 4694–4703, 2002.

[118] A. M. Ruschak and A. D. Miranker, "Fiber-dependent amyloid formation as catalysis of an existing reaction pathway.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 30, pp. 12341–12346, 2007.

[119] C. Andersen, H. Yagi, M. Manno, V. Martorana, T. Ban, G. Christiansen, D. Otzen, Y. Goto, and C. Rischel, "Branching in amyloid fibril growth," *Biophysical journal*, vol. 96, no. 4, pp. 1529–1536, 2009.

[120] M. Anderson, O. V. Bocharova, N. Makarava, L. Breydo, V. V. Salnikov, and I. V. Baskakov, "Polymorphism and ultrastructural organization of prion protein amyloid fibrils : an insight from high resolution atomic force microscopy.," *Journal of molecular biology*, vol. 358, no. 2, pp. 580–596, 2006.

[121] M. Calamai, N. Taddei, M. Stefani, G. Ramponi, and F. Chiti, "Relative influence of hydrophobicity and net charge in the aggregation of two homologous proteins.," *Biochemistry*, vol. 42, no. 51, pp. 15078–15083, 2003.

[122] D. M. Walsh, A. Lomakin, G. B. Benedek, M. Condron, and D. B. Teplow, "Amyloid beta-protein fibrillogenesis. Detection of a protofibrillar intermediate.," *Journal of Biological Chemistry*, vol. 272, no. 35, pp. 22364–22372, 1997.

[123] J. Zhang and M. Muthukumar, "Simulations of nucleation and elongation of amyloid fibrils.," *The Journal of chemical physics*, vol. 130, no. 3, p. 035102, 2009.

[124] F. Oosawa and M. Kasai, "A theory of linear and helical aggregations of macromolecules.," *Journal of molecular biology*, vol. 4, pp. 10–21, 1962.

[125] A. Wegner and J. Engel, "Kinetics of the cooperative association of actin to actin filaments.," *Biophysical Chemistry*, vol. 3, no. 3, pp. 215–225, 1975.

[126] H. Flyvbjerg, E. Jobs, and S. Leibler, "Kinetics of self-assembling microtubules : an "inverse problem" in biochemistry.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 12, pp. 5975–5979, 1996.

[127] J. Hofrichter, P. D. Ross, and W. A. Eaton, "Kinetics and Mechanism of Deoxyhemoglobin S Gelation : A New Approach to Understanding Sickle Cell Disease," *Proceedings of the National Academy of Sciences of the United States of America*, pp. 4864–4868, Dec. 1974.

[128] W. P. Esler, E. R. Stimson, J. M. Jennings, H. V. Vinters, J. R. Ghilardi, J. P. Lee, P. W. Mantyh, and J. E. Maggio, "Alzheimer's Disease Amyloid Propagation by a Template-Dependent Dock-Lock Mechanism," *Biochemistry*, vol. 39, no. 21, pp. 6288–6295, 2000.

[129] M. J. Cannon, A. D. Williams, R. Wetzel, and D. G. Myszka, "Kinetic analysis of beta-amyloid fibril elongation.," *Analytical biochemistry*, vol. 328, no. 1, pp. 67–75, 2004.

[130] C. Goldsbury, P. Frey, V. Olivieri, U. Aebi, and S. Müller, "Multiple assembly pathways underlie amyloid-β fibril polymorphisms," *Journal of molecular biology*, vol. 352, no. 2, pp. 282–298, 2005.

[131] G. Bhak, Y.-J. Choe, and S. R. Paik, "Mechanism of amyloidogenesis : nucleation-dependent fibrillation versus double-concerted fibrillation.," *BMB reports*, vol. 42, no. 9, pp. 541–551, 2009.

[132] S. E. Hill, J. Robinson, G. Matthews, and M. Muschol, "Amyloid protofibrils of lysozyme nucleate and grow via oligomer fusion.," *Biophysical journal*, vol. 96, no. 9, pp. 3781–3790, 2009.

[133] A. J. Modler, K. Gast, G. Lutsch, and G. Damaschun, "Assembly of amyloid protofibrils via critical oligomers–a novel pathway of amyloid formation.," *Journal of molecular biology*, vol. 325, no. 1, pp. 135–148, 2003.

[134] N. Benseny-Cases, M. Cócera, and J. Cladera, "Conversion of non-fibrillar β-sheet oligomers into amyloid fibrils in Alzheimer's disease amyloid peptide aggregation," *Biochemical and biophysical research communications*, vol. 361, no. 4, pp. 916–921, 2007.

[135] A. L. Fink, "The Aggregation and Fibrillation of α-Synuclein," *Accounts of chemical research*, vol. 39, no. 9, pp. 628–634, 2006.

[136] A. D. Miranker, "Unzipping the mysteries of amyloid fiber formation.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, pp. 4335–4336, Mar. 2004.

[137] E. Zerovnik, "Amyloid-fibril formation. Proposed mechanisms and relevance to conformational disease.," *The FEBS journal*, vol. 269, no. 14, pp. 3362–3371, 2002.

[138] C. M. Dobson, "Protein Misfolding, Evolution and Disease," *Trends in biochemical sciences*, vol. 24, no. 9, pp. 329–332, 1999.

[139] D. Thirumalai, D. K. Klimov, and R. I. Dima, "Emerging ideas on the molecular basis of protein and peptide aggregation.," *Current opinion in structural biology*, vol. 13, no. 2, pp. 146–159, 2003.

[140] M. Fändrich, J. Meinhardt, and N. Grigorieff, "Structural polymorphism of Alzheimer Abeta and other amyloid fibrils.," *Prion*, vol. 3, no. 2, pp. 89–93, 2009.

[141] M. S. Celej, W. Caarls, A. P. Demchenko, and T. M. Jovin, "A triple-emission fluorescent probe reveals distinctive amyloid fibrillar polymorphism of wild-type alpha-synuclein and its familial Parkinson's disease mutants.," *Biochemistry*, vol. 48, no. 31, pp. 7465–7472, 2009.

[142] J. Meinhardt, C. Sachse, and P. Hortschansky, "A [beta](1-40) Fibril Polymorphism Implies Diverse Interaction Patterns in Amyloid Fibrils," *Journal of Molecular Biology*, 2009.

[143] J. Madine, E. Jack, and P. Stockley, "Structural Insights into the Polymorphism of Amyloid-Like Fibrils Formed by Region 20-29 of Amylin Revealed by Solid-State NMR and X-ray Fiber Diffraction," *Journal of the American Chemical Society*, 2008.

[144] R. Verel, I. Tomka, and C. Bertozzi, "Polymorphism in an Amyloid-Like Fibril-Forming Model Peptide," *Angewandte Chemie International Edition*, 2008.

[145] A. T. Petkova, W.-M. Yau, and R. Tycko, "Experimental constraints on quaternary structure in Alzheimer's beta-amyloid fibrils.," *Biochemistry*, vol. 45, no. 2, pp. 498–512, 2006.

[146] L. Kreplak and U. Aebi, "From the polymorphism of amyloid fibrils to their assembly mechanism and cytotoxicity.," *Advances in Protein Chemistry*, vol. 73, pp. 217–233, 2006.

[147] Y. Mo, Y. Lu, G. Wei, and P. Derreumaux, "Structural diversity of the soluble trimers of the human amylin(20-29) peptide revealed by molecular dynamics simulations.," *The Journal of chemical physics*, vol. 130, no. 12, pp. 125101–125101, 2009.

[148] W. L. Klein, W. B. Stine, and D. B. Teplow, "Small assemblies of unmodified amyloid beta-protein are the proximate neurotoxin in Alzheimer's disease.," *Neurobiology of aging*, vol. 25, no. 5, pp. 569–580, 2004.

[149] G. M. Shankar, S. Li, T. H. Mehta, A. Garcia-Munoz, N. E. Shepardson, I. Smith, F. M. Brett, M. A. Farrell, M. J. Rowan, C. A. Lemere, C. M. Regan, D. M. Walsh, B. L. Sabatini, and D. J. Selkoe, "Amyloid-beta protein dimers isolated directly from Alzheimer's brains impair synaptic plasticity and memory.," *Nature medicine*, vol. 14, no. 8, pp. 837–842, 2008.

[150] K. K. Ono, M. M. M. Condron, and D. B. D. Teplow, "Structure-neurotoxicity relationships of amyloid beta-protein oligomers.," *Audio, Transactions of the IRE Professional Group on*, vol. 106, no. 35, pp. 14745–14750, 2009.

[151] H. A. Lashuel, D. Hartley, B. M. Petre, T. Walz, and P. T. Lansbury, "Neurodegenerative disease : amyloid pores from pathogenic mutations.," *Nature*, vol. 418, no. 6895, pp. 291–291, 2002.

[152] M. S. Goldberg and P. T. Lansbury, "Is there a cause-and-effect relationship between alpha-synuclein fibrillization and Parkinson's disease ?," *Nature cell biology*, vol. 2, no. 7, pp. E115–9, 2000.

[153] P. T. Lansbury, "Evolution of amyloid : what normal protein folding may tell us about fibrillogenesis and disease.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 7, pp. 3342–3344, 1999.

[154] W.-F. Xue, A. L. Hellewell, W. S. Gosal, S. W. Homans, E. W. Hewitt, and S. E. Radford, "Fibril fragmentation enhances amyloid cytotoxicity.," *The Journal of biological chemistry*, vol. 284, no. 49, pp. 34272–34282, 2009.

[155] R. Diaz-Avalos, C. Long, E. Fontano, M. Balbirnie, R. Grothe, D. Eisenberg, and D. L. D. Caspar, "Cross-beta order and diversity in nanocrystals of an amyloid-forming peptide.," *Journal of molecular biology*, vol. 330, no. 5, pp. 1165–1175, 2003.

[156] G. Wei, W. Song, and P. Derreumaux, "Self-assembly of amyloid-forming peptides by molecular dynamics simulations," *Frontiers in bioscience*, 2008.

[157] A. Irbäck and S. Mitternacht, "Spontaneous beta-barrel formation : an all-atom Monte Carlo study of Abeta16-22 oligomerization.," *Proteins : Structure, Function, and Bioinformatics*, vol. 71, no. 1, pp. 207–214, 2008.

[158] A. De Simone and P. Derreumaux, "Low molecular weight oligomers of amyloid peptides display beta-barrel conformations : a replica exchange molecular dynamics study in explicit solvent.," *The Journal of chemical physics*, vol. 132, no. 16, p. 165103, 2010.

[159] W. Chen, N. Mousseau, and P. Derreumaux, "The conformations of the amyloid-β (21–30) fragment can be described by three families in solution," *The Journal of chemical physics*, vol. 125, no. 8, pp. 084911–084911, 2006.

[160] A. A. Melquiond, X. X. Dong, N. N. Mousseau, and P. P. Derreumaux, "Role of the region 23-28 in Abeta fibril formation : insights from simulations of the monomers and dimers of Alzheimer's peptides Abeta40 and Abeta42.," *Current Alzheimer research*, vol. 5, no. 3, pp. 244–250, 2008.

[161] J. J. Maupetit, P. P. Derreumaux, and P. P. Tuffery, "PEP-FOLD : an online resource for de novo peptide structure prediction.," *Audio, Transactions of the IRE Professional Group on*, vol. 37, no. Web Server issue, pp. W498–W503, 2009.

[162] J. Maupetit, P. Derreumaux, and P. Tufféry, "A fast method for large-scale De Novo peptide and miniprotein structure prediction.," *Journal Of Computational Chemistry*, vol. 31, no. 4, pp. 726–738, 2009.

[163] P. Carter, C. A. F. Andersen, and B. Rost, "DSSPcont : continuous secondary structure assignments for proteins," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3293–3295, 2003.

[164] L. Holm and C. Sander, "Database algorithm for generating protein backbone and side-chain co-ordinates from a C alpha trace application to model building and

detection of co-ordinate errors.," *Journal of molecular biology*, vol. 218, no. 1, pp. 183–194, 1991.

[165] C. Hartmann and I. Antes, "IRECS : A new algorithm for the selection of most probable ensembles of side-chain conformations in protein models," *Protein science*, 2007.

[166] R. Dunbrack Jr and F. Cohen, "Bayesian statistical analysis of protein side-chain rotamer preferences," *Protein science*, 1997.

[167] H. Berendsen and J. Grigera, "The missing term in effective pair potentials," *Journal of Physical Chemistry*, 1987.

[168] B. Hess and H. Bekker, "LINCS : a linear constraint solver for molecular simulations," *Journal of Computational Chemistry*, 1997.

[169] S. Miyamoto, "SETTLE : an analytical version of the SHAKE and RATTLE algorithm for rigid water models," *Journal Of Computational Chemistry*, 1992.

[170] H. Berendsen, D. van der Spoel, and R. van Drunen, "GROMACS : A message-passing parallel molecular dynamics implementation," *Computer Physics Communications*, vol. 91, pp. 43–56, 1995.

[171] E. Lindahl, B. Hess, and D. van der Spoel, "GROMACS 3.0 : a package for molecular simulation and trajectory analysis," *Journal of Molecular Modeling*, vol. 7, pp. 306–317, 2001.

[172] D. van der Spoel, E. Lindahl, B. Hess, A. van Buuren, and E. Apol, *Gromacs User Manual version 3.2*. Nijenborg 4, 9747 AG Groningen, 2004.

[173] W. Van Gunsteren and H. Berendsen, *Gromos-87 manual ; Biomos BV, Nijenborgh 4 : 9747 AG Groningen*. The Netherlands, 1987.

[174] W. Van Gunsteren, X. Daura, and A. Mark, "GROMOS force field," *Encyclopedia of Computational Chemistry*, 1998.

[175] W. Scott, P. Hünenberger, and I. Tironi, "The GROMOS biomolecular simulation program package," *The Journal of Physical Chemistry*, vol. 103, pp. 3596–3607, 1999.

[176] W. Van Gunsteren, S. Billeter, A. Eising, P. Huenenberger, and P. Krueger, "Biomolecular Simulation : The GROMOS96 Manual and User Guide (Zurich : Hochschulverlag AG an der ETH Zuerich)," 1996.

[177] P. P. Liu, B. B. Kim, R. A. R. Friesner, and B. J. B. Berne, "Replica exchange with solute tempering : a method for sampling biological systems in explicit water.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 39, pp. 13749–13754, 2005.

[178] M. B. Kubitzki and B. L. de Groot, "Molecular dynamics simulations using temperature-enhanced essential dynamics replica exchange.," *Biophysical journal*, vol. 92, no. 12, pp. 4262–4270, 2007.

[179] D. van der Spoel and M. M. Seibert, "Protein folding kinetics and thermodynamics from atomistic simulations.," *Physical review letters*, vol. 96, no. 23, p. 238102, 2006.

[180] S. Santini, G. Wei, N. Mousseau, and P. Derreumaux, "Pathway complexity of Alzheimer's beta-amyloid Abeta16-22 peptide assembly.," *Structure*, vol. 12, no. 7, pp. 1245–1255, 2004.

[181] R. Pellarin, P. Schuetz, E. Guarnera, and A. Caflisch, "Amyloid Fibril Polymorphism Is under Kinetic Control.," *Journal of the American Chemical Society*, vol. 132, no. 42, pp. 14960–14970, 2010.

[182] S. S. Santini, N. N. Mousseau, and P. P. Derreumaux, "In silico assembly of Alzheimer's Abeta16-22 peptide into beta-sheets.," *Journal of the American Chemical Society*, vol. 126, no. 37, pp. 11509–11516, 2004.

[183] D.-W. Li, S. Mohanty, A. Irbäck, and S. Huo, "Formation and growth of oligomers : a Monte Carlo study of an amyloid tau fragment.," *PLoS Computational Biology*, vol. 4, no. 12, p. e1000238, 2008.

[184] R. R. Tycko, K. L. K. Sciarretta, J. P. R. O. J. Orgel, and S. C. S. Meredith, "Evidence for novel beta-sheet structures in Iowa mutant beta-amyloid fibrils.," *Audio, Transactions of the IRE Professional Group on*, vol. 48, no. 26, pp. 6072–6084, 2009.

[185] F. R. F. Salemme and D. W. D. Weatherford, "Conformational and geometrical properties of beta-sheets in proteins. I. Parallel beta-sheets.," *Journal of molecular biology*, vol. 146, no. 1, pp. 101–117, 1981.

[186] F. R. F. Salemme and D. W. D. Weatherford, "Conformational and geometrical properties of beta-sheets in proteins. II. Antiparallel and mixed beta-sheets.," *Journal of molecular biology*, vol. 146, no. 1, pp. 119–141, 1981.

[187] O. N. O. Antzutkin, J. J. J. Balbach, R. D. R. Leapman, N. W. N. Rizzo, J. J. Reed, and R. R. Tycko, "Multiple quantum solid-state NMR indicates a parallel, not antiparallel, organization of beta-sheets in Alzheimer's beta-amyloid fibrils.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 24, pp. 13045–13050, 2000.

[188] J. J. Shorter and S. S. Lindquist, "Destruction or potentiation of different prions catalyzed by similar Hsp104 remodeling activities.," *Molecular Cell*, vol. 23, no. 3, pp. 425–438, 2006.

[189] C. Behrends, C. A. Langer, R. Boteva, U. M. Böttcher, M. J. Stemp, G. Schaffar, B. V. Rao, A. Giese, H. Kretzschmar, K. Siegers, and F. U. Hartl, "Chaperonin TRiC promotes the assembly of polyQ expansion proteins into nontoxic oligomers.," *Molecular Cell*, vol. 23, no. 6, pp. 887–897, 2006.

[190] D. E. Ehrnhoefer, J. Bieschke, A. Boeddrich, M. Herbst, L. Masino, R. Lurz, S. Engemann, A. Pastore, and E. E. Wanker, "EGCG redirects amyloidogenic

polypeptides into unstructured, off-pathway oligomers.," *Nature structural & molecular biology*, vol. 15, no. 6, pp. 558–566, 2008.

[191] M. Malisauskas, A. Darinskas, V. V. Zamotin, A. Gharibyan, I. A. Kostanyan, and L. A. Morozova-Roche, "Intermediate amyloid oligomers of lysozyme : Is their cytotoxicity a particular case or general rule for amyloid ?," *Biochemistry*, vol. 71, no. 5, pp. 505–512, 2006.

[192] G. Bitan, M. D. Kirkitadze, A. Lomakin, S. S. Vollers, G. B. Benedek, and D. B. Teplow, "Amyloid beta-protein (Abeta) assembly : Abeta 40 and Abeta 42 oligomerize through distinct pathways.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 1, pp. 330–335, 2003.

[193] H. D. Nguyen and C. K. Hall, "Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 46, pp. 16180–16185, 2004.

[194] S. Auer, C. M. Dobson, and M. Vendruscolo, "Characterization of the nucleation barriers for protein aggregation and amyloid formation.," *HFSP journal*, vol. 1, no. 2, pp. 137–146, 2007.

[195] S. Auer, F. Meersman, C. M. Dobson, and M. Vendruscolo, "A generic mechanism of emergence of amyloid protofilaments from disordered oligomeric aggregates.," *PLoS Computational Biology*, vol. 4, no. 11, p. e1000222, 2008.

[196] S. Auer, C. M. Dobson, M. Vendruscolo, and A. Maritan, "Self-templated nucleation in peptide and protein aggregation.," *Physical review letters*, vol. 101, no. 25, pp. 258101–258101, 2008.

[197] M. Cheon, I. Chang, and C. K. Hall, "Spontaneous formation of twisted aβ(16-22) fibrils in large-scale molecular-dynamics simulations.," *Biophysical journal*, vol. 101, no. 10, pp. 2493–2501, 2011.

[198] D. Matthes, V. Gapsys, and B. L. de Groot, "Driving Forces and Structural Determinants of Steric Zipper Peptide Oligomer Formation Elucidated by Atomistic Simulations.," *Journal of molecular biology*, 2012, in press.

[199] B. Morel, L. Varela, A. I. Azuaga, and F. Conejero-Lara, "Environmental Conditions Affect the Kinetics of Nucleation of Amyloid Fibrils and Determine Their Morphology," *Biophysical journal*, vol. 99, no. 11, pp. 3801–3810, 2010.

[200] J. Nasica-Labouze and N. Mousseau, *Kinetics of Amyloid Growth, Alzheimer's disease : Molecular Basis of Amyloid-beta protein aggregation and fibril formation - Insights into low molecular weight and cytotoxic aggregates from computer simulations.* Imperial Press College, in press, scheduled for fall 2012.

[201] P. Hortschansky, V. Schroeckh, T. Christopeit, G. Zandomeneghi, and M. Fändrich, "The aggregation kinetics of Alzheimer's β-amyloid peptide is controlled by stochastic nucleation," *Protein science : a publication of the Protein Society*, vol. 14, no. 7, pp. 1753–1759, 2005.

[202] P. M. Tessier and S. Lindquist, "Prion recognition elements govern nucleation, strain specificity and species barriers.," *Nature*, vol. 447, no. 7144, pp. 556–561, 2007.

[203] X. Qi, L. Hong, and Y. Zhang, "A variational model for oligomer-formation process of GNNQQNY peptide from yeast prion protein Sup35.," *Biophysical journal*, vol. 102, no. 3, pp. 597–605, 2012.

[204] C. Lanczos, *Applied analysis*. Dover Publications, 1988.

[205] T. R. Serio, "Nucleated Conformational Conversion and the Replication of Conformational Information by a Prion Determinant," *Science (New York, N.Y.)*, vol. 289, no. 5483, pp. 1317–1321, 2000.

[206] M. Cheon, I. Chang, S. Mohanty, L. M. Luheshi, C. M. Dobson, M. Vendruscolo, and G. Favrin, "Structural reorganisation and potential toxicity of oligome-

ric species formed during the assembly of amyloid fibrils.," *PLoS Computational Biology*, vol. 3, no. 9, pp. 1727–1738, 2007.

[207] J. Sørensen, X. Periole, K. K. Skeby, S.-J. Marrink, and B. Schiøtt, "Protofibrillar Assembly Toward the Formation of Amyloid Fibrils," *The Journal of Physical Chemistry Letters*, vol. 2, no. 19, pp. 2385–2390, 2011.

[208] W. S. Gosal, I. J. Morten, E. W. Hewitt, D. A. Smith, N. H. Thomson, and S. E. Radford, "Competing pathways determine fibril morphology in the self-assembly of beta2-microglobulin into amyloid.," *Journal of molecular biology*, vol. 351, no. 4, pp. 850–864, 2005.

[209] N. Ferguson, J. Berriman, M. Petrovich, T. D. Sharpe, J. T. Finch, and A. R. Fersht, "Rapid amyloid fiber formation from the fast-folding WW domain FBP28.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 17, pp. 9814–9819, 2003.

[210] M. Sorci, R. A. Grassucci, I. Hahn, J. Frank, and G. Belfort, "Time-dependent insulin oligomer reaction pathway prior to fibril formation : Cooling and seeding," *Proteins : Structure, Function, and Bioinformatics*, vol. 77, no. 1, pp. 62–73, 2009.

[211] V. Foderà, S. Cataldo, F. Librizzi, B. Pignataro, P. Spiccia, and M. Leone, "Self-organization pathways and spatial heterogeneity in insulin amyloid fibril formation.," *The journal of physical chemistry. B*, vol. 113, no. 31, pp. 10830–10837, 2009.

[212] K. C. Kunes, D. L. Cox, and R. R. P. Singh, "One-dimensional model of yeast prion aggregation," *Physical Review E*, vol. 72, no. 5, pp. 051915–051915, 2005.

[213] B. Linse and S. Linse, "Monte Carlo simulations of protein amyloid formation reveal origin of sigmoidal aggregation kinetics.," *Molecular bioSystems*, vol. 7, no. 7, pp. 2296–2303, 2011.

[214] J. T. Jarrett, E. P. Berger, and P. T. Lansbury, "The carboxy terminus of the beta amyloid protein is critical for the seeding of amyloid formation : implications for the pathogenesis of Alzheimer's disease.," *Biochemistry*, vol. 32, no. 18, pp. 4693–4697, 1993.

[215] H. Han, P. H. Weinreb, and P. T. Lansbury, "The core Alzheimer's peptide NAC forms amyloid fibrils which seed and are seeded by beta-amyloid : is NAC a common trigger or target in neurodegenerative disease ?," *Chemistry & biology*, vol. 2, no. 3, pp. 163–169, 1995.

[216] S. Chen, V. Berthelier, J. B. Hamilton, B. O'Nuallai, and R. Wetzel, "Amyloid-like Features of Polyglutamine Aggregates and Their Assembly Kinetics," *Biochemistry*, vol. 41, no. 23, pp. 7391–7399, 2002.

[217] R. Grossier and S. Veesler, "Reaching One Single and Stable Critical Cluster through Finite-Sized Systems," *Crystal Growth & Design*, vol. 9, no. 4, pp. 1917–1922, 2009.

[218] J. Wedekind, D. Reguera, and R. Strey, "Finite-size effects in simulations of nucleation.," *The Journal of chemical physics*, vol. 125, no. 21, p. 214505, 2006.

[219] S. Xu, B. Bevis, and M. Arnsdorf, "The assembly of amyloidogenic yeast sup35 as assessed by scanning (atomic) force microscopy : an analogy to linear colloidal aggregation ?," *Biophysical journal*, vol. 81, no. 1, pp. 446–454, 2001.

[220] R. Zhang, X. Hu, H. Khant, S. J. Ludtke, W. Chiu, M. F. Schmid, C. Frieden, and J.-M. Lee, "Interprotofilament interactions between alzheimer's aβ_{1-42} peptides in amyloid fibrils revealed by cryoem.," *Proc. Natl. Acad. Sci. USA*, vol. 106, no. 12, pp. 4653–4658, 2009.

[221] Y. Miller, B. Ma, C.-J. Tsai, and R. Nussinov, "Hollow core of alzheimer's aβ_{42} amyloid observed by cryoem is relevant at physiological ph.," *Proc. Natl. Acad. Sci. USA*, vol. 107, no. 32, pp. 14128–14133, 2010.

[222] D. W. Peterson, H. Zhou, F. W. Dahlquist, and J. Lew, "A soluble oligomer of tau associated with fiber formation analyzed by NMR.," *Biochemistry*, vol. 47, no. 28, pp. 7393–7404, 2008.

[223] M. R. M. Nichols, M. A. M. Moss, D. K. D. Reed, W.-L. W. Lin, R. R. Mukhopadhyay, J. H. J. Hoh, and T. L. T. Rosenberry, "Growth of beta-amyloid(1-40) protofibrils by monomer elongation and lateral association. Characterization of distinct products by light scattering and atomic force microscopy.," *Biochemistry*, vol. 41, no. 19, pp. 6115–6127, 2002.

[224] M. R. Nichols, "Amyloid-β Protofibrils Differ from Amyloid-β Aggregates Induced in Dilute Hexafluoroisopropanol in Stability and Morphology," *The Journal of biological chemistry*, vol. 280, no. 4, pp. 2471–2480, 2004.

[225] R. Friedman, R. Pellarin, and A. Caflisch, "Soluble Protofibrils as Metastable Intermediates in Simulations of Amyloid Fibril Degradation Induced by Lipid Vesicles," *The Journal of Physical Chemistry Letters*, vol. 1, no. 2, pp. 471–474, 2010.

[226] J. Nasica-Labouze and N. Mousseau, "Kinetics of Amyloid Aggregation : a Study of the GNNQQNY Prion Sequence," *submitted to PLoS Computational Biology on July 20th*, 2012.

[227] M. S. Li, D. K. Klimov, J. E. Straub, and D. Thirumalai, "Probing the mechanisms of fibril formation using lattice models.," *The Journal of chemical physics*, vol. 129, no. 17, p. 175101, 2008.

[228] N. J. Edwin, R. P. Hammer, R. L. McCarley, and P. S. Russo, "Reversibility of β-Amyloid Self-Assembly : Effects of pH and Added Salts Assessed by Fluorescence Photobleaching Recovery," *Biomacromolecules*, vol. 11, no. 2, pp. 341–347, 2010.

[229] S. M. Loveday, J. Su, M. A. Rao, S. G. Anema, and H. Singh, "Effect of calcium

on the morphology and functionality of whey protein nanofibrils.," *Biomacromolecules*, vol. 12, no. 10, pp. 3780–3788, 2011.

[230] S. M. Loveday, X. L. Wang, M. A. Rao, S. G. Anema, and H. Singh, "Effect of pH, NaCl, CaCl2 and temperature on self-assembly of β-lactoglobulin into nanofibrils : a central composite design study.," *Journal of agricultural and food chemistry*, vol. 59, no. 15, pp. 8467–8474, 2011.

[231] E. Monsellier, M. Ramazzotti, N. Taddei, and F. Chiti, "A computational approach for identifying the chemical factors involved in the glycosaminoglycans-mediated acceleration of amyloid fibril formation.," *PloS one*, vol. 5, no. 6, p. e11363, 2010.

[232] V. Castelletto, I. W. Hamley, C. Cenker, and U. Olsson, "Influence of Salt on the Self-Assembly of Two Model Amyloid Heptapeptides," *The journal of physical chemistry. B*, vol. 114, no. 23, pp. 8002–8008, 2010.

[233] A. Magno, A. Caflisch, and R. Pellarin, "Crowding effects on amyloid aggregation kinetics," *The Journal of Physical Chemistry Letters*, vol. 1, pp. 3027–3032, 2010.

[234] M. Volmer and Z. Weber, "Nuclei formation in supersaturated states," *The Journal of Physical Chemistry*, vol. 119, no. 277, 1926.

[235] P. Debenedetti, "Metastable Liquids : Concepts and Principles," 1996.

[236] J. Macdonald, "Homogeneous Nucleation of Vapor Condensation. I. Thermodynamics Aspects.," *The American Journal of Physics*, vol. 30, pp. 870–877, 1962.

[237] J. Macdonald, "Homogeneous Nucleation of Vapor Condensation. II. Kinetics Aspects.," *The American Journal of Physics*, vol. 31, pp. 31–41, 1963.

[238] J. F. Smith, T. P. J. Knowles, C. M. Dobson, C. E. Macphee, and M. E. Welland, "Characterization of the nanoscale properties of individual amyloid fibrils.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 43, pp. 15806–15811, 2006.

[239] U. Baxa, "Structural basis of infectious and non-infectious amyloids.," *Current Alzheimer research*, vol. 5, no. 3, pp. 308–318, 2008.

[240] F. Ferrone, "Kinetics of sickle hemoglobin polymerization : : I. Studies using temperature-jump and laser photolysis techniques," *Journal of molecular biology*, 1985.

[241] A. Lomakin, D. B. Teplow, D. A. Kirschner, and G. B. Benedek, "Kinetic theory of fibrillogenesis of amyloid beta-protein.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 15, pp. 7942–7947, 1997.

[242] W. Yong, A. Lomakin, M. D. Kirkitadze, D. B. Teplow, S.-H. Chen, and G. B. Benedek, "Structure determination of micelle-like intermediates in amyloid beta -protein fibril assembly by using small angle neutron scattering.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 1, pp. 150–154, 2002.

[243] G. Tiana, F. Simona, R. A. Broglia, and G. Colombo, "Thermodynamics of beta-amyloid fibril formation.," *The Journal of chemical physics*, vol. 120, no. 17, pp. 8307–8317, 2004.

[244] E. T. Powers and D. L. Powers, "Mechanisms of protein fibril formation : nucleated polymerization with competing off-pathway aggregation.," *Biophysical journal*, vol. 94, no. 2, pp. 379–391, 2008.

[245] C. F. Lee, "Self-assembly of protein amyloids : a competition between amorphous and ordered aggregation.," *Physical Review E*, vol. 80, no. 3 Pt 1, p. 031922, 2009.

[246] M. A. Roseman, "Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds.," *Journal of molecular biology*, vol. 200, no. 3, pp. 513–522, 1988.

[247] R. Cowan and R. G. Whittaker, "Hydrophobicity indices for amino acid residues as determined by high-performance liquid chromatography.," *Peptide research*, vol. 3, no. 2, pp. 75–80, 1990.

[248] B. M. Broome and M. H. Hecht, "Nature disfavors sequences of alternating polar and non-polar amino acids : implications for amyloidogenesis.," *Journal of molecular biology*, vol. 296, no. 4, pp. 961–968, 2000.

[249] F. Chiti, M. Calamai, N. Taddei, M. Stefani, G. Ramponi, and C. M. Dobson, "Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99 Suppl 4, pp. 16419–16426, 2002.

[250] F. Chiti, M. Stefani, N. Taddei, G. Ramponi, and C. M. Dobson, "Rationalization of the effects of mutations on peptide and protein aggregation rates.," *Nature*, vol. 424, no. 6950, pp. 805–808, 2003.

[251] M. Reches, Y. Porat, and E. Gazit, "Amyloid fibril formation by pentapeptide and tetrapeptide fragments of human calcitonin.," *Journal of Biological Chemistry*, vol. 277, no. 38, pp. 35475–35480, 2002.

[252] W. Wang and M. H. Hecht, "Rationally designed mutations convert de novo amyloid-like fibrils into monomeric beta-sheet proteins.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 5, pp. 2760–2765, 2002.

[253] E. A. Meyer, R. K. Castellano, and F. Diederich, "Interactions with aromatic rings in chemical and biological recognition.," *Angewandte Chemie (International ed. in English)*, vol. 42, no. 11, pp. 1210–1250, 2003.

[254] K. F. DuBay, A. P. Pawar, F. Chiti, J. Zurdo, C. M. Dobson, and M. Vendruscolo, "Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains.," *Journal of molecular biology*, vol. 341, no. 5, pp. 1317–1326, 2004.

[255] F. Bemporad, N. Taddei, M. Stefani, and F. Chiti, "Assessing the role of aromatic residues in the amyloid aggregation of human muscle acylphosphatase.," *Protein science : a publication of the Protein Society*, vol. 15, no. 4, pp. 862–870, 2006.

[256] R. D. Hills and C. L. Brooks, "Hydrophobic cooperativity as a mechanism for amyloid nucleation.," *Journal of molecular biology*, vol. 368, no. 3, pp. 894–901, 2007.

[257] L. Jean, C. F. Lee, M. Shaw, and D. J. Vaux, "Structural elements regulating amyloidogenesis : a cholinesterase model system.," *PloS one*, vol. 3, no. 3, pp. e1834–e1834, 2008.

[258] G. W. Platt, K. E. Routledge, S. W. Homans, and S. E. Radford, "Fibril growth kinetics reveal a region of beta2-microglobulin important for nucleation and elongation of aggregation.," *Journal of molecular biology*, vol. 378, no. 1, pp. 251–263, 2008.

[259] M. G. Wolf, J. A. Jongejan, J. D. Laman, and S. W. de Leeuw, "Quantitative prediction of amyloid fibril growth of short peptides from simulations : calculating association constants to dissect side chain importance.," *Journal of the American Chemical Society*, vol. 130, no. 47, pp. 15772–15773, 2008.

[260] J. Adamcik, J.-M. Jung, J. Flakowski, P. De Los Rios, G. Dietler, and R. Mezzenga, "Understanding amyloid aggregation by statistical analysis of atomic force microscopy images.," *Nature nanotechnology*, vol. 5, no. 6, pp. 423–428, 2010.

[261] F. Oosawa and S. Asakura, "Thermodynamics of the Polymerization of Protein.," 1975.

[262] M. G. Wolf, J. v. Gestel, and S. W. de Leeuw, "Modeling amyloid fibril formation : a free-energy approach.," *Methods in Molecular Biology*, vol. 474, pp. 153–179, 2008.

[263] J. S. Pedersen and D. E. Otzen, "Amyloid-a state in many guises : survival of the fittest fibril fold.," *Protein science : a publication of the Protein Society*, vol. 17, no. 1, pp. 2–10, 2008.

ANNEXE I

KINETICS OF AMYLOID GROWTH

I.1 Objectifs

L'objectif de ce chapitre est de faire une revue de littérature complète des phénomènes spontanés d'assemblage ayant lieu chez les protéines amyloïdes et d'expliquer pourquoi la formation de fibres amyloïdes peut être vue comme un processus complexe de nucléation.

Ce chapitre, présentement sous presse, sera publié sous peu dans un livre scientifique qui s'intitule: *Alzheimer's disease: Molecular Basis of Amyloid-beta protein aggregation and fibril formation - Insights into low molecular weight and cytotoxic aggregates from computer simulations.* – Imperial Press College

I.2 Contributions des auteurs

(Dans l'ordre d'apparition dans la publication) Jessica Nasica-Labouze a rédigé la première version du texte et Normand Mousseau a participé à la révision et à la correction.

I.3 Introduction

The kinetics of amyloid fibril formation are controlled by protein concentration and the appearance of cross-β structures. At high protein concentrations, for example, fibril formation is strongly enhanced by the presence of preformed cross-β seeds[1] despite an important entropic barrier. Once a large enough cross-β seed forms, monomers easily join these nuclei, in a fast growth process that leads to fully-formed fibrils. This mechanism, called growth by nucleation, is the key element in amyloid fibrillization [25].

[1] A cross-β seed is an aggregate whose main structure is a cross-β structure large enough to trigger the formation of fibrils through a nucleation process.

However, the biological nucleation process does not follow the classical physics theory of nucleation due to the biochemical complexity of proteins; numerous models have thus attempted to introduce and explain new elements missing from the classical theory of nucleation, but necessary to understand and characterize the nucleation of amyloid proteins as observed in in vitro experiments.

In this chapter, we focus on the elements missing from the classical theory of nucleation to accurately describe amyloid aggregation and on the models introduced to address them. These models will be compared with the most recent experimental results. To conclude this chapter, we offer a critical discussion of the current state of knowledge.

I.4 Classical theory of nucleation

Volmer *et al.* [234] were the first to affirm that the nucleation rate is proportional to the exponential of the free energy of formation for a nucleus and to provide the first formulation of the classical theory of homogeneous[2] nucleation. Although incomplete, this theory still forms the basis on which the most recent models describing the nucleation process of various physical systems are built [235].

I.4.1 Thermodynamic aspects of the classical homogeneous nucleation

Let us consider a homogeneous phase A in a supersaturated state at temperature T . In order to create a phase B by nucleation from the phase A, the new phase must overcome an important free energy barrier ΔF and condensate. Assuming that the phase B forms small roughly spherical aggregates or embryos of radius r, then the free energy barrier can be expressed as

$$\Delta F(r) = 4\pi r^2\sigma - \frac{4}{3}\pi r^3 \rho RT \ln(S) \tag{I.1}$$

where σ is the tension at the embryo's surface, ρ is the density of the phase B, R is the universal gas constant, and S is a measure of supersaturation [236]. If S > 1, the function ΔF(r) has a maximum, which represents the lowest free energy activation barrier that an

[2] *Homogeneous* here means that the system is considered to be composed of identical monomers with no impurities.

embryo must overcome to become stable and start a spontaneous and irreversible local growth (Fig. I.1(a)). The critical embryo size[3] corresponding to that free energy barrier is defined as

$$r^* = \frac{2\sigma}{\rho RT \ln(S)} \tag{I.2}$$

This relation shows that the higher the supersaturation of the phase, the smaller the critical size of stable embryos needed before they can continue to grow. In general, the embryos' formation is thus unfavorable and can only occur as a result of rare events in a supersaturated phase. In the rare case where the critical size is reached, embryos are short-lived and metastable and have a probability of one-half of launching the growth process.

I.4.2 Kinetic aspects of the classical homogeneous nucleation

The kinetic treatment of the classical theory of nucleation as shown by Macdonald *et al.* [237] expresses quantities in terms of g, the number of molecules in a roughly spherical embryo of radius r. Hence g^* is the number of molecules in a critical nucleus of radius r^*. Let C_g be the rate of creation for embryos made of g molecules and let E_g be the rate of destruction for embryos of the same size. A g-mer can thus be created by adding a monomer (i.e. a molecule) to a (g – 1)-mer or by removing a monomer from a (g + 1)-mer. Concurrently, a g-mer can be destroyed by either growing into a (g + 1)-mer or by losing a monomer. If N_g is the concentration of g-mers at an instant t , the rate of formation of g-mers can be written as

$$\frac{\partial N_g}{\partial t} = \left(C_{g-1}N_{g-1} + E_{g+1}N_{g+1}\right) - \left(C_g N_g + E_g N_g\right) \tag{I.3}$$

If we define $I_g = C_g N_g - E_{g+1} N_{g+1}$ as the net rate for g-mers, then Eq. (I.3) can be rewritten as

$$\frac{\partial N_g}{\partial t} = I_{g-1} - I_g \tag{I.4}$$

[3] An embryo of critical size is defined as a *nucleus*.

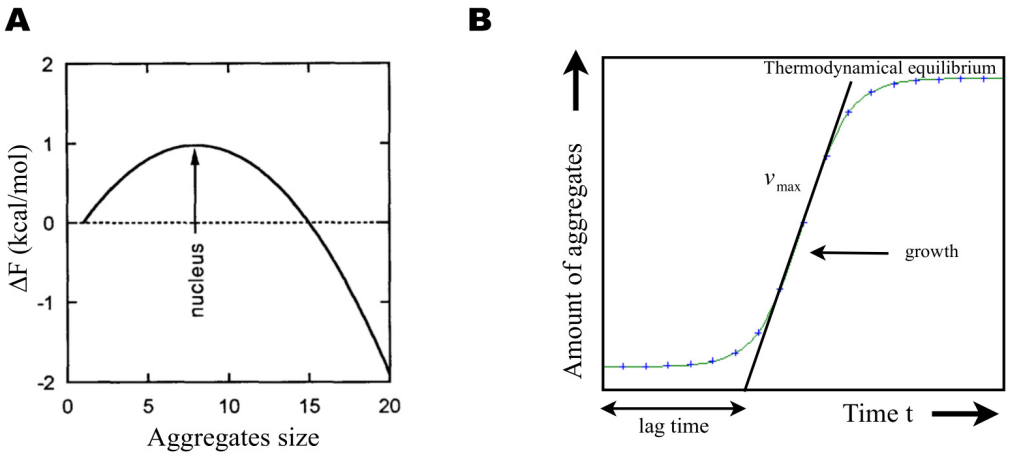

Figure I.1: **Nucleation characteristics.** (a) Typical free energy barrier for the kinetic embryo formation. A nucleus forms when the free energy maximum is overcome by an aggregate (Ferrone, 1999). (b) Sigmoid growth curve characteristic of the presence of a secondary nucleation process. The initial phase is known as the lag phase (i.e. the time necessary to form enough stable nuclei). Then the growth phase is exponential, with a maximum growth rate v_{max} (i.e. the maximum slope). Growth slows down when the monomers start to deplete. The system is then under thermodynamical control. Figure 3.1(a) is reprinted from Ferrone (1999).

Considering the approximation I_g = constant for all g and defining I = $I_g/volume$, the general term for the net formation rate of g-mers per unit time per unit volume [237], we find that

$$I = \frac{C_{g^*} n_{g^*}}{\sqrt{2\pi kT/Q}} \tag{I.5}$$

with $n_{g^*} = n_1 e^{-\frac{\Delta F^*}{kT}}$ and $Q = -[\partial^2(\Delta F(g)/\partial g^2)]_{g^*}$.

n_1 is the initial population of monomers, n_{g^*} is the population of supercritical nuclei of size g^* monomers and k is the Boltzmann constant. In Eq. (I.5), I is the net formation rate for supercritical g^*-mers per unit time per unit volume when dynamic equilibrium is reached. This result indicates that if the supersaturation S increases, then ΔF^* decreases and the rate I consequently increases.

I.4.3 Amyloid fibrillization as a nucleation process and the missing elements to the classical theory to describe it

Based on the classical theory of nucleation, one can argue that the formation of amyloid fibrils displays all of the features of a typical nucleation process. To support this hypothesis, three arguments have been suggested [111, 123].

1. Amyloid aggregation is not possible under a certain critical protein concentration.

2. For a protein concentration slightly above the critical concentration, the fibrillization starts after a lag time, which corresponds to the time necessary to form stable nuclei.

3. The lag time can be reduced or even suppressed by adding preformed nuclei to a solution of monomers.

These three aspects of nucleation have been widely confirmed experimentally for amyloid proteins. Other characteristics of nucleation observed for amyloid proteins also include the presence of metastable nuclei and of an exponential growth during the initial phase of nucleation [111].

Nonetheless, the intrinsic nature of proteins complicates the nucleation process and key elements are missing from the classical theory, as a result, to describe the nucleation and growth of amyloid fibrils. Recent experiments suggest the presence of secondary nucleation processes during the formation of amyloid fibrils and such processes are not described by the classical theory of nucleation. Secondary processes may happen in the form of fragmentation [104, 109, 112–114, 209, 212, 238, 239] or heterogeneous nucleation (occurring at the surface of fibrils) [29, 107, 115–120, 240]. Numerous models have been developed over the last decade in an effort to integrate secondary nucleation processes with the general picture of amyloid aggregation, but the behavior of amyloid proteins can vary greatly from one to another. As a result, each protein has its own mechanism and can undergo one or several secondary processes and it has become a real challenge to try to unify the behavior of various amyloid proteins. An account of the different models proposed to describe the secondary nucleation processes is given in Section 3.5.1.

In addition to secondary processes, a certain degree of cooperativity can exist between intermediary oligomers and mature fibrils that can affect the nucleation process of amyloid proteins [2, 3, 33, 42, 43, 103, 122, 130–135, 140, 192, 241–245]. Models treating that aspect of fibrillization are presented in Section 3.5.2.

Finally, the intrinsic nature of proteins must be taken into account when characterizing the nucleated amyloid fibrillization and considerable effort has been put into developing a better understanding of the significance of the biochemical nature of proteins and peptides in the nucleation of amyloid fibrils [121, 246–260]. This aspect of nucleation is explored in Section 3.5.3.

I.5 The kinetics of amyloid fibrillization

For amyloid proteins, a slow primary nucleation phase produces small aggregates (or critical nuclei) composed of monomers followed by a growth phase, much faster than the nucleation, during which the nuclei are turned into long fibrils. By definition, nucleation is the phase that precedes and includes the appearance of critical nuclei (homogeneous

or heterogeneous, see Section 3.5.1.2). In contrast, growth defines the postnucleation (homogeneous or heterogeneous) phase during which the amount of polymers increases very rapidly after a lag time τ_{lag} . During growth, several species can be present in solution such as embryos, nuclei, oligomers (disordered aggregates that are larger than nuclei), protofilaments (nascent polymeric chains), filaments (long polymeric chains), and fibrils (lateral assemblies of several filaments or protofilaments).

I.5.1 Secondary nucleation processes

All of the models discussed here agree that the growth rate evolution as a function of time for amyloid fibrils is best described by a sigmoid curve whose initial phase corresponds to a nucleation lag phase (Fig. I.1(b)). The lag time here takes into account two levels of nucleation (primary and secondary) and is defined as the time necessary to form a sufficient density of nuclei that are stable enough to generate polymers (Fig. I.1(b)). Experimentally, the lag time is defined as the waiting time necessary for polymers, larger than nuclei, to be first detected after a nucleation process.

In the presence of a classical primary nucleation process only, the polymerization kinetics would be described by a parabolic curve (i.e. the growth rate would be proportional to t^2) [261] and we would witness a linear polymerization with no secondary nucleation processes. Amyloid proteins do not follow such a curve and it has been demonstrated experimentally that the presence of a secondary nucleation[4] phenomenon is necessary to explain the kinetic curves, growth rates, and fibril morphologies observed for a number of amyloid proteins.

Frank Ferrone [108] established a model comparing the kinetic behavior of actin with and without a secondary nucleation process. His conclusion was clear: the inclusion of a secondary nucleation process leads to a growth curve with a much more abrupt slope than the growth curve using a simple primary nucleation process. More precisely, Ferrone found that the incorporation of a secondary nucleation process into a primary nucleation process leads to an exponential growth curve that slows down only when the monomer population starts to significantly deplete. The resulting growth can be described by a

[4] also called “double nucleation”.

sigmoid curve with a steep slope, ν_{max}, a phenomenon that has been extensively observed for amyloid proteins (see the references for secondary processes mentioned above). It is important to note, however, that the impact of double nucleation strongly depends on the quantity of fibrils present in solution. Thus, it contributes little to the total nucleation rate at the beginning of the process, but dominates the reaction after a large amount of fibrils have been formed. As a result, the lag time, which takes into account secondary nucleation, depends weakly on the initial concentration of monomers, and is sensitive rather to the concentration of assembled polymers. In addition, Ferrone's model also predicts that the kinetic growth curve should become convex when the nucleation step is bypassed [109]. Indeed, by adding preformed nuclei to a solution of monomers, the lag time necessary for the formation of stable nuclei is avoided and polymerization may start immediately.

Secondary nucleation is a generic term that includes any mechanism during which the instantaneous formation of new growth sites depends on the amount of fibrils (or filaments) already formed. The two main types of double nucleation are fragmentation (Section 3.5.1.1), which is a breaking mechanism, and lateral nucleation (Section 3.5.1.2), which includes diffusive mechanisms, bifurcation mechanisms, and thickening mechanisms from existing preformed fibrils. Lateral nucleation is often referred to as heterogeneous nucleation as it occurs at preferred sites on the surface of pre-existing fibrils formed by primary nucleation.

I.5.1.1 Secondary nucleation through fragmentation

As discussed above, fragmentation is a common secondary process for many amyloid proteins. This mechanism, which occurs after a primary nucleation phase, accelerates the fibril-forming growth rate: as fibrils break, the number of fibril extremities increases. Those extremities recruit more and more monomers, hence acting as secondary nucleation sites during the growth phase. Numerous models have been proposed in order to describe the influence of fragmentation on the kinetics of amyloid growth.

Knowles *et al.* [109] have developed an analytical model to characterize the fragmentation phenomenon observed for insulin, β-lactoglobulin and the domain WW from

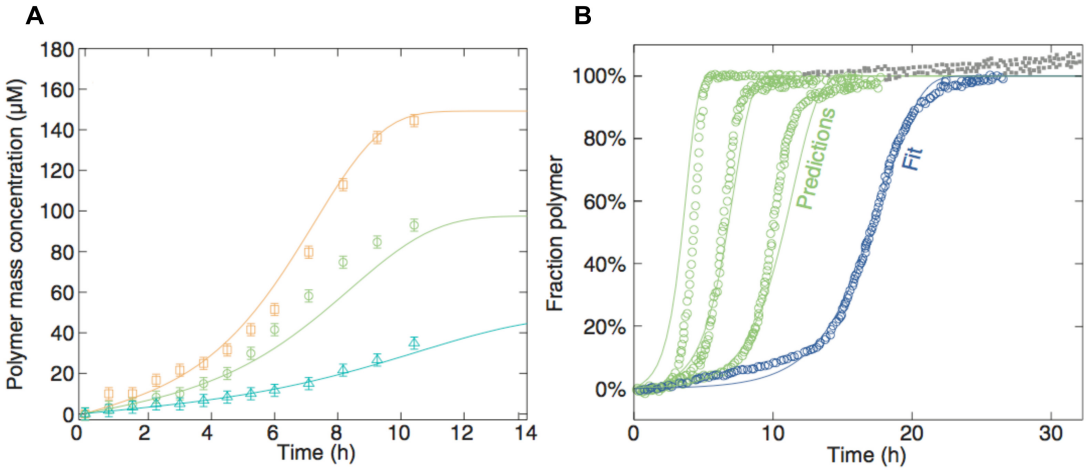

Figure I.2: **Experimental results (Knowles *et al.*, 2009; Ferguson *et al.*, 2003) showing the presence of a fragmentation phenomenon.** (a) Variation of the fibrillization rate as a function of time for insulin for different monomer concentrations m_{tot} (orange: 149 μM, green: 98 μM, cyan: 49 μM) (Knowles *et al.*, 2009). A sigmoid curve characteristic of secondary nucleation processes is apparent for values of m_{tot} sufficiently big (i.e. bigger than the solubility). (b) Polymerization measurements for the WW domain of FBP28 (blue and green circles) (Ferguson et al., 2003). The blue curve represents the best correspondence between Knowles's model and the experimental data for m_{tot} = 50 μM. The green curves are then obtained for different values of m_{tot} (from left to right: 500 μM, 200 μM, 100 μM), keeping the same parameters used for generating the blue curve. The model's predictions agree well with experimental results and the lag time and maximum growth rate are predicted accurately. Grey squares are data not considered for fitting the blue curve and for predicting the green curves. Reprinted from Knowles *et al.* (2009).

FBP28 and compared their analytical results with experimental measurements. In the case of insulin, the typical sigmoid behavior of the growth curve, obtained experimentally, is entirely captured by the analytical model and is characteristic of the presence of a secondary nucleation mechanism, here fragmentation (Fig. I.2(a)). A light-scattering (LS) experiment performed on the WW domain of FBP28 [110] and focusing on the kinetic behavior as a function of the initial concentration of monomers m_{tot} found that the lag time τ_{lag} decreases as m_{tot} is increased. Knowles *et al.* [109] have successfully tuned their fragmentation model to accurately fit these experimental data, predicting the behavior of the WW domain for other values of m_{tot} (Fig. I.2(b)) whose maximum growth rate, under fragmentation, can be analytically expressed as

$$\nu_{max} = \frac{\sqrt{2m_{tot}k_+k_-}}{e} \tag{I.6}$$

where k_+, k_- are the association rate for monomers and the dissociation (fragmentation) rate for polymers. Here, ν_{max} depends only on the association/dissociation rates and the initial concentration of monomers as opposed to the number of formed nuclei and the primary nucleation rate that would be expected for a pure primary nucleation process. For β-lactoglobulin, experiments studying the effect of bypassing the primary nucleation by adding preformed nuclei to a solution of monomers [132] have shown that the growth curve remains sigmoid, which means that there is still another level of nucleation present in the process. A similar behavior has been observed experimentally for other amyloid proteins such as β-amyloid (Aβ), β2-microglobulin, the prion protein, and Phosphatidylinositol 3-kinase's SH3 domain (PI(3)K-SH3) [104, 111–114]. Based on these observations, Knowles *et al.* [109] have analytically set the primary nucleation rate to zero in their model and have successfully reproduced the experimental results for β-lactoglobulin, which reinforces the hypothesis that fragmentation is the dominant secondary process for this protein.

During fragmentation, the growth rate is directly proportional to the concentration of extremities of the available fibrils' segments [112]. It was shown that fibrils under

thermal agitation or sonication[5] undergo breakage and thus a pronounced acceleration of their growth rate along with an increase in their average length after a period of rest following the sonication process [113]. This observation suggests that, in addition to accelerating their growth rate, agitation/sonication leads to longer fibrils than those undergoing a classical polymerization process. The fibril length is then proportional to the fragmentation rate k_- [212]. Another well-observed trait of fragmentation includes the weak dependency of the lag time on the monomer concentration [109, 112, 212]. Experimentally, the lag time is found to be

$$\tau_{lag} \approx (m_{tot})^{-1/2} \tag{I.7}$$

This has been observed for a wide set of amyloid proteins undergoing natural fragmentation such as β2-microglobulin, Sup35 Prion protein, Ure2p protein, insulin, the WW domain of FBP28 and the PrP Prion protein. Analytically, the model of Knowles *et al.* [109] derives the lag time to be

$$\tau_{lag} \approx \frac{1}{\sqrt{2 m_{tot} k_+ k_-}} \approx (m_{tot})^{-1/2} \tag{I.8}$$

The lag time in Eq. (I.8) depends on the fragmentation rate k_- and is consistent with experimental observations. One can thus conclude that τ_{lag} here is essentially defined by the fragmentation mechanism, unlike the lag time from classical polymerization, which depends mainly on the primary nucleation rate k_n and hence on the concentration of monomers present [261]. The model further shows that a minor change in the fragmentation rate can have drastic consequences on the growth kinetics of amyloid fibrils.

Knowles *et al.* have thus successfully reproduced the experimental data and have better characterized the growth mechanism of several amyloid proteins undergoing fragmentation. However, the degree of brittleness of fibrils must be taken into account to better understand and evaluate the importance of the biochemical nature of amino acid sequences on fragmentation and thus on the acceleration of the elongation process for amyloid proteins.

[5] Sonication is the application of ultrasound energy to agitate and promote breakage of fibrils.

I.5.1.2 Heterogeneous secondary nucleation occurring at the surface of existing fibrils

Heterogeneous nucleation (or lateral nucleation) is a nucleation process happening at the surface of pre-existing fibrils (Fig. I.3). Similarly to fragmentation, this phenomenon displays all of the features of a secondary nucleation process: an exponential growth and a lag time weakly dependent on the concentration of monomers before the growth phase. In the heterogeneous nucleation process, fibrils form first by homogeneous (primary) nucleation. Then, as more fibrils are formed, the total accessible fibril surface increases. As a result, more and more potential nucleation sites appear on the fibril surface to recruit monomers and form new nuclei. This leads to a continuous increase of the heterogeneous nucleation rate until depletion of the monomers with a simultaneous cooperativity between the homogeneous and heterogeneous nucleation processes. This mechanism, like fragmentation, depends strongly on the amount of fibrils formed in solution, which explains the presence of an autocatalytic (i.e. exponential) growth rate and of a long lag time before enough polymerized fibrils can be detected.

A mathematical kinetic model was developed by Frank Ferrone [107] to describe heterogeneous nucleation as an explanation for the polymerization of hemoglobin S (HbS). This model offers a good description of the heterogeneous nucleation process occurring at the surface of fibrils and is applicable to amyloid proteins. In this model, the incorporation of monomers can be performed either by homogeneous (primary) or heterogeneous (secondary) nucleation, irreversibly in both cases, under the assumption that the fraction of monomers consumed in the nucleation processes is negligible compared with the amount consumed in fibril growth. In addition, Ferrone mathematically derives an expression for the concentration of heterogeneous nuclei considering that the heterogeneous nucleation happens in two steps.

1. The formation of a m^*-mer nucleus from m^* monomers.

2. The adhesion of this m^*-mer nucleus to the surface of a fibril.

These equations lead to an expression for both the formation rate of fibrils and the

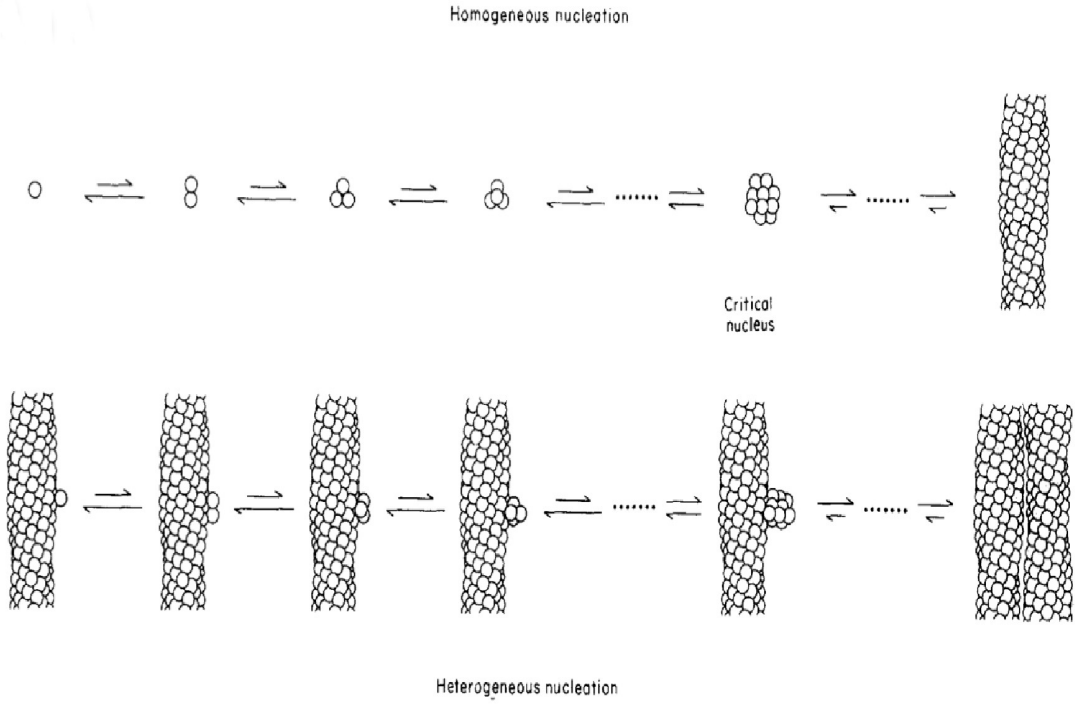

Figure I.3: **Heterogeneous nucleation model.** Reprinted from Ferrone (1985).

disappearance rate of monomers that shows that the evolution of the concentration of polymerized monomers in time is consistent with an exponential growth, typical of a secondary nucleation process. Measurements carried out by Ferrone *et al.* [240] further show a linear relation between the logarithm of the lag time and the logarithm of the monomer concentration.

Heterogeneous nucleation is also called lateral nucleation as it occurs at the surface of fibers. Lateral nucleation has been largely observed experimentally as a natural process for many amyloid proteins such as insulin [29, 115], calcitonin [116], and Islet Amyloid Polypeptide IAPP [117]. Lateral nucleation can also be catalyzed by the presence of a surface as is the case for IAPP20-29, Aβ, α-synuclein, and the τ protein [118]. It has been shown by LS, in the particular case of IAPP20-29, that the presence of a surface facilitates a secondary nucleation by reducing the entropic cost of assembling proteins in nuclei: (i) by allowing proteins to locally aggregate densely on the surface, (ii) by encouraging the formation of ordered structures (i.e. the formation of parallel β-sheets) when in contact with the surface, and (iii) by stabilizing high-energy nuclei in contact with the surface. Lateral nucleation at the surface of fibrils can take three different forms [119]. If the nucleus formed at the surface of a fibril detaches itself after its formation, the lateral nucleation is called diffusive. This phenomenon has been observed for HbS [240] but not for amyloid proteins. If the nucleus remains permanently on the fibril's surface, two scenarios can occur: (i) thickening (the nucleus continues to grow along the pre-existing fibril resulting in a thicker fibril) [29, 115–120], and (ii) branching (the nucleus forms a new branch from the pre-existing fibril). Branching has been observed for the glucagon protein [119] but not for any amyloid proteins and won't be discussed here. Lateral nucleation by thickening, however, has been observed for calcitonin [116] and insulin [29, 115]. During the thickening process for calcitonin, fibrillar centers form from which many fibers radially stretch out, as observed by electron microscopy (EM) (Fig. I.4(a)-(c)) [116]. The process is accompanied by a linear dependency of the logarithm of m_{tot} on the logarithm of τ_{lag} , as mentioned by Ferrone for the secondary nucleation of HbS. As for insulin, Jansen *et al.* [29] have also observed a thickening process by atomic force microscopy (AFM) and it seems that protofilaments are able to

recruit monomers and to act as scaffolds to help form new protofilaments laterally along pre-existing ones (Fig. I.4(d)). Insulin fibers are hence formed by lateral interactions between protofilaments and monomers, by a thickening process, and their aggregation dynamics show a weak dependency of the lag time on the concentration of monomers along with an exponential growth [115].

Thus, observations for various proteins are consistent with the general description of the kinetic characteristics described by Ferrone's mathematical model. However, there remains a great diversity in the observed heterogeneous nucleation processes and each protein follows its own mechanism depending strongly on its amino acid sequence and on the environmental conditions during fibrillization. These aspects must be taken into account in order to better understand the heterogeneous nucleation mechanisms.

I.5.2 Presence of intermediary species during the amyloid nucleation process

Polymorphism is the ability to adopt different morphologies for a fiber made of identical monomers. It is a widespread property of amyloid fibrils [3, 42, 140] that implies that the same protein can form fibrils of different morphologies simply through different mechanisms [130], depending on the environmental conditions. During fibrillization by nucleation, the path taken varies from one mechanism to another, as well as the type of intermediary species involved in the final product, i.e. the fibril. The presence of intermediary oligomers during the formation of fibrils plays an important role in amyloid polymorphism and in fibrillization [2, 33, 122, 130, 134, 192]. For instance, it appears that fibrillization for the protein α-synuclein is augmented when the oligomer's concentration increases [135], which means that oligomers can be intermediary species on the fibril formation pathway.

Numerous models have proposed appealing mechanisms to incorporate oligomers into the picture and to explain their role in the observed polymorphism for amyloid proteins; despite the contradictions between the most accepted models, they explain accurately the experimental observations for individual proteins.

One of the first models constructed by Lomakin *et al.* [241] for the Aβ protein is a kinetic model coupled to experimental results [103]. In this model, the aggregation of free

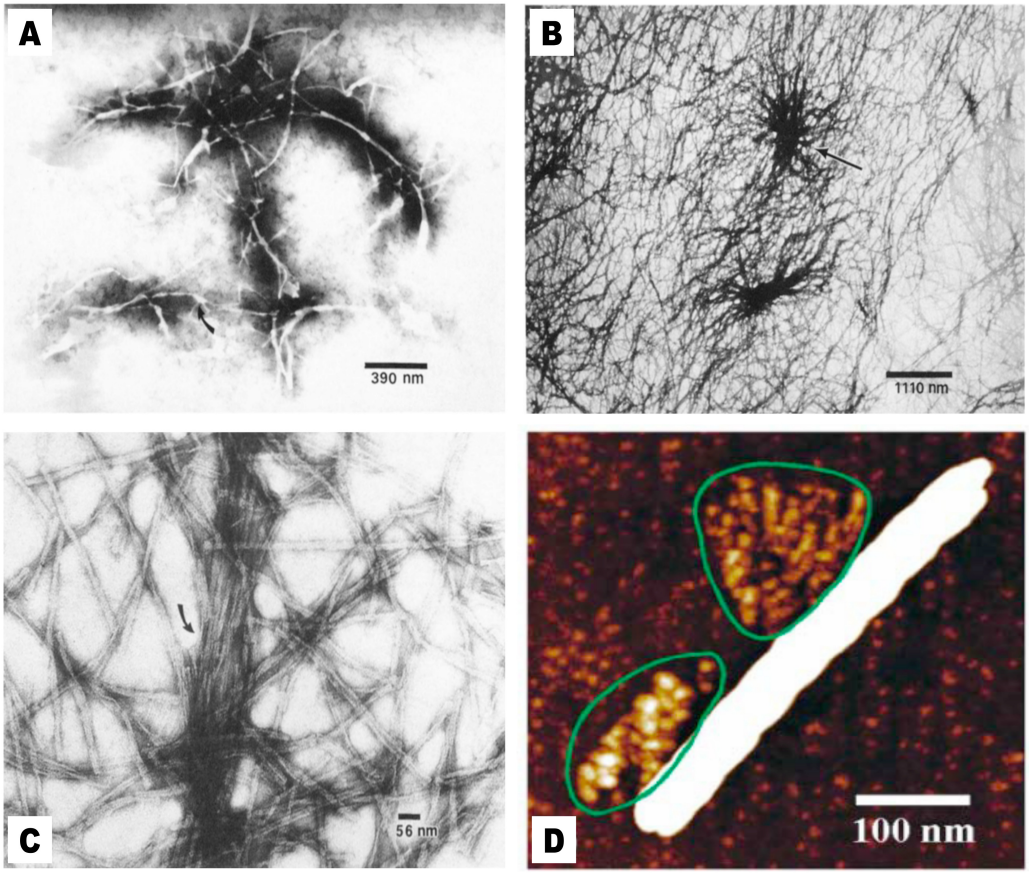

Figure I.4: **Experimental observations of lateral nucleation through thickening.** (a-c) Images obtained for calcitonin by electronic microscopy (EM) (Arvinte *et al.*, 1993). The fibers created by lateral nucleation are arranged in star-like structures spreading radially from a dense center. (c) Close-up of one branch of a star-like structure in (a) and (b). The main core in (c) is made of several protofilaments interacting laterally (arrow). (d) Image of insulin aggregates by atomic force microscopy (AFM) (Jansen *et al.*, 2005). Prefibrillar aggregates accumulate around protofibers through a thickening mechanism. The protofibers serve as scaffolds for low-molecular-weight oligomers and monomers. Reprinted from Arvinte *et al.* (1993) and Jansen *et al.* (2005).

monomers into stable ordered nuclei is considered to happen through the formation of disordered oligomers called micelles. Micelle formation is suggested to be a fast process as it would not require any configurational rearrangement of the monomers, leaving the micelles in a disordered state. Thereafter, stable ordered nuclei would form from these micelles in a very slow process as it would require a rearrangement of the monomers inside the micelles. The model of Lomakin *et al.* [103] also considers the possibility of stable ordered nuclei forming in parallel directly from free monomers according to a process considered to be extremely slow. This model highlights the importance of oligomers in the very first steps of the nucleation process. Hence, oligomers are a species present during the nucleation lag time [134].

The role of micelles for the Aβ protein has also been examined by a thermodynamic model coupled to numerical molecular dynamics simulations [243]. This second model suggests that micelles are an important metastable intermediary species, present at the beginning of nucleation and whose first function is to serve as monomer reservoirs [43, 103, 242].

Other models [33, 43] instead focus on the role of ordered aggregates, or protofibers, acting as important intermediary species in the fibrillization process. Experimental data collected by size exclusion chromatography, LS and EM by Walsh *et al.* [122] for Aβ reveal the existence of ordered intermediaries, or protofibers, and suggest that these protofibers are the product of a nucleation process (Fig. 1.5a). They also suggested that protofibers could interact together to form fibers. The proposed mechanisms for protofiber interactions are depicted in Fig. I.5(b) [33, 43]. Finally, as in the case of micelles, protofibers could serve as monomer reservoirs (Fig. I.5(c)). This hypothesis is supported by the model of Zhang and Muthukumar [123] who suggest that, after the nucleation phase, an Ostwald ripening mechanism dominates. Ostwald ripening consists of a process under which the largest aggregates grow larger at the expense of smaller ones. Hence, in the presence of Ostwald ripening, protofibers would be disintegrated into monomers to serve the cause of larger fibers. AFM results [33], however, suggest that protofibers are forming on the fibril formation pathway for Aβ and as such cannot serve as monomer reservoirs.

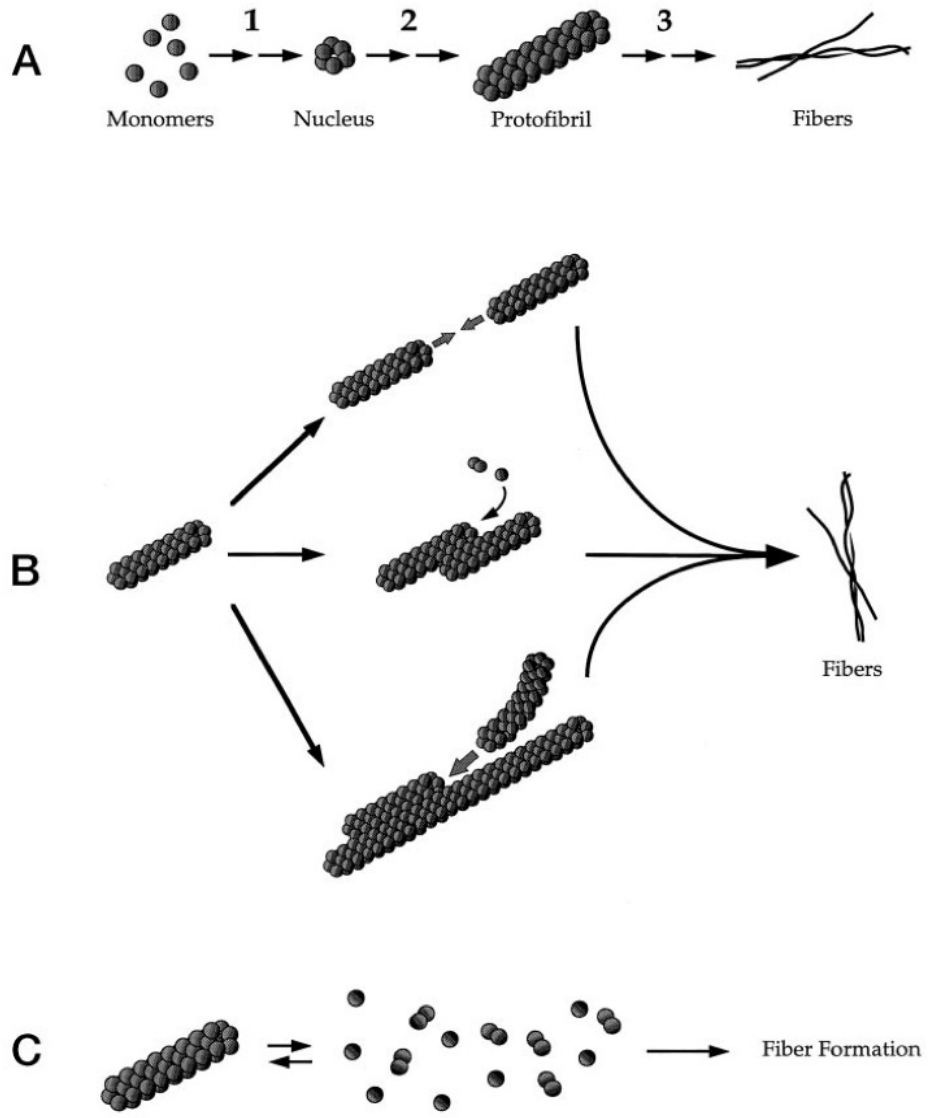

Figure I.5: **Models describing the current theories on fibrillization from protofibers for Aβ (Walsh *et al.*, 1997).** (a) Aβ fibrillization is a nucleation where monomers form nuclei (step 1), growing into protofibers by elongation (step2). Finally, the protofibers interact to form fibers (step 3). (b) Hypotheses on the nature of the association between protofibers, resulting into fibers. The first (top) is an end-to-end association but is very unlikely due to the difficulty protofibers would have to find the right alignment for this type of association. The other two possible associations are similar to an heterogeneous nucleation process as described in Section 3.5.1.2 where protofibers associate laterally and serve as scaffolds for monomers and/or other protofibers in order to form fibrils similarly to Jansen *et al.* (c) Protofibers serve as monomer reservoirs, thus being an off-pathway species. Reprinted from Walsh *et al.* (1997).

More models [43, 130] suggest that protofibers may form by oligomer fusion, through an aggregation pathway that does not involve any nucleation. This phenomenon was observed for the phosphoglycerate kinase protein [133], Aβ [130], Aβ(1-40) [134], α-synuclein [135], and the hen egg-white lysozyme protein [132]. Oligomer fusion implies that, first, disordered oligomers can form easily from free monomers. Once the oligomer concentration is high enough, fusion may occur and give birth to protofibers (Fig. I.6(b)) [132]. Then, protofibers can interact to form fibers. Oligomer fusion is suggested to operate cooperatively with the nucleation process and would hence help amyloid fibril formation [131, 132]. EM and AFM data obtained for Aβ by Goldsbury *et al.* [130] contradict the oligomer fusion hypothesis, with their model instead suggesting that protofibers form through a nucleation mechanism starting from oligomers and by monomer addition.

It is, however, possible that oligomers escape from the nucleation or the fusion process to end up being off-pathway[6] species. A mathematical model [244] predicts that the presence of off-pathway oligomers should slow down the formation of fibrils. This supposes that off-pathway oligomers, thought to be rare but toxic species, have a longer lifetime than on-pathway assemblies. It is also suggested that off-pathway oligomers can rearrange into on-pathway oligomers.

Hence, a large set of experimental data has given rise to multiple scenarios about the role and nature of intermediary species involved in the nucleation process of amyloid proteins and it appears that oligomers, micelles and protofibers are important actors in amyloid aggregation.

I.5.3 Importance of the biochemical properties of amyloid proteins and peptides in the nucleation process

Numerous protein physicochemical parameters – such as the propensity for a sequence to prefer a certain type of secondary structure, the charge, or the hydrophobicity of side chains – come into play to define the aggregation kinetics of amyloid nucleation

[6] *Off-pathway* means here that oligomers do not directly participate to the pathway leading to the formation of fibers.

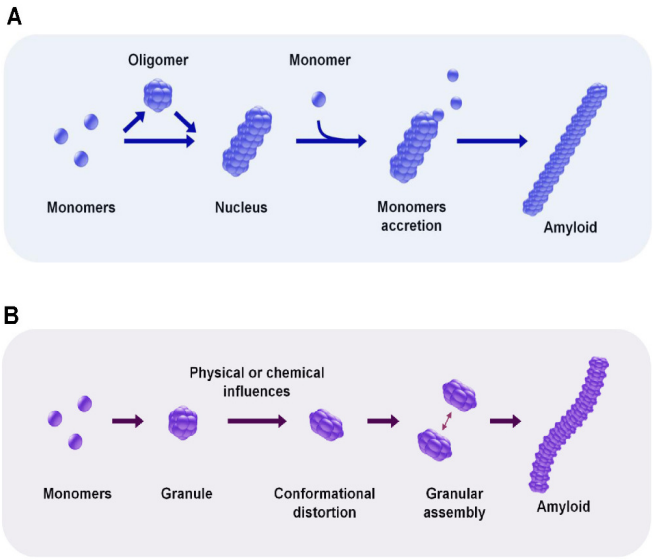

Figure I.6: **Comparison between models describing fibril formation** as (a) a nucleation-growth process and (b) an oligomer fusion process. Reprinted from Rochet and Lansbury (2000).

[25]. Hence, multiple hypotheses have been formulated in order to better characterize the respective role of the biochemical properties of proteins in fibrillization.

The model developed by DuBay *et al.* [254] proposes an explanation for the role of physicochemical properties of proteins in the formation of fibrils and describes nucleation by means of a phenomenological equation based on experimental measurements

$$\log(k) = \alpha_0 + \alpha_{hydr} I^{hydr} + \alpha_{pat} I^{pat} + \alpha_{ch} I^{ch} + \alpha_{pH} E^{pH} + \alpha_{ionic} E^{ionic} + \alpha_{conc} E^{conc} \quad \text{(I.9)}$$

Eq. (I.9) relates the aggregation rate k to various intrinsic (I) and extrinsic (E) physicochemical factors such as the hydrophobicity of the sequence (I^{hydr} , the normalized sum of the hydrophobic contributions from each residue in the sequence as measured experimentally) [246, 247], the presence of sequences alternating hydrophobic and hydrophilic amino acids (I^{pat}, a +1 factor is assigned for each pattern of more than five amino acids alternating between hydrophobic and hydrophilic), the absolute net charge (I^{ch}), the pH

(E^{pH}), the ionic strength (E^{ionic}), and the protein concentration (E^{conc}). The α coefficients were obtained by linear regression over a set of experimental data. Using this equation along with experimental data collected for 15 proteins and their variants, the relative influences of intrinsic[7] properties have been deduced.

1. The increase of a sequence's total hydrophobicity leads to an increase in the aggregation rate [121]. Given that a native protein (*i.e.* correctly folded) exposes very few hydrophobic side chains to the solvent, whereas a misfolded protein exposes many more, this effect is consistent with the fact that misfolding promotes amyloid fibrillization.

2. The presence of patterns which rarely occur naturally [248], alternating hydrophobic and hydrophilic amino acids, is the most influential factor for the aggregation rate. The higher the frequency of appearance of such patterns within a protein, the higher the aggregation rate. This relation has been confirmed by the work of Wang and Hecht [252].

3. The aggregation rate is inversely proportional to the total absolute net charge of a sequence. This effect has been noted experimentally [249].

Experimental data obtained by Chiti *et al.* [250] by amino acid mutagenesis on the human muscle acylphosphatase (AcP) protein further confirm the observations 1-3 extracted from the model. These researchers further observe that the rate of aggregation increases with the increased propensity of a sequence to go from an α to a β state. Moreover, a thermodynamic study of the equilibrium between monomers, oligomers, and fibrils confirms that the parameters 1-3 do favor amyloid fibrillization through nucleation and also predicts that (4) an increase in the number of aromatic amino acids in a sequence should favor fibrillization [245]. This is observed for the β2-microglobulin protein [251, 255, 258].

[7] The effects of extrinsic factors are not described here.

I.5.3.1 Hydrophobicity, patterns alternating hydrophobic and hydrophilic residues, and aromatic residues

Amyloid nucleation cannot occur without interactions between β-sheets. Hence, the interdigitation of hydrophobic side chains, which allows cross-β structures to interact laterally, is crucial in the polymorphism and stability of the product fiber and depends strongly on the amino acid sequence. It has been shown by Wolf *et al.* [259] that the tendency of short sequences to form fibers depends on their ability to gather the hydrophobic side chains, coupled to the presence of patterns alternating hydrophilic and hydrophobic side chains. Such a pattern allows a stable packing of hydrophobic side chains inside the cross-β structure and allows a stable packing of hydrophilic side chains outside contacting the solvent. In addition, it seems that there exists a certain degree of hydrophobic cooperativity between the side chains, which is the basis of the nucleation mechanism [256]. Also, it is thought that the increase of hydrophobicity could shift the equilibrium towards the formation of a larger amount of oligomers [245].

The importance of aromatic amino acids comes from the fact that they are not only hydrophobic but they can also form π-π interactions with other aromatic residues that are directed along the fibrillar axis and thus contribute to the global stability of fibrils [253, 257].

I.5.3.2 Net charge

The increase of the net charge of a protein slows down fibrillization. This can be explained by recent observations [260] on the effect of the total charge of monomer proteins on the morphology of the resulting fibers. Indeed, it seems that the fiber twisted morphology comes from a high mutual repulsion effect between the charged neighboring constituent monomers. This effect of the total charge could slow down fibrillization.

DuBay *et al.* [254], in their model, suggest that other intrinsic factors may be of interest, such as the stability of the native protein, to evaluate the global reconfiguration of monomers prior to nucleation, as misfolding is the key to fibril formation. Hence, the biochemical nature of proteins strongly influences the amyloid nucleation process

and should be taken into account in order to better understand polymorphism and the multiple assembly mechanisms.

I.6 Critical review of the models

Models developed over the last two decades have been, for the most part, constructed from experimentally-derived parameters and observations. For that reason, most models are, unfortunately, designed to describe specific proteins, *in vitro*, under the very specific conditions used in the experiments that the model is based on. As a result, many models describe specific phenomena peculiar to a given sequence without providing a unified picture of amyloid behaviors. This is not necessarily a drawback as these models are particularly relevant for identifying and describing accurately the elements missing from the classical theory of nucleation.

I.6.1 Kinetic models against thermodynamic models

Of the growth models presented in this chapter, most are based on the kinetics of formation as opposed to thermodynamics. This is justifiable as thermodynamic (or statistical-mechanical) models can only describe the equilibrium properties of a system, whereas kinetic (or dynamic) models consider the time evolution of the system's properties. As most experimental studies focus on the time evolution of fibrillization, kinetic models have tended to be used for better validation. As pointed out by Wolf *et al.* [262], kinetic models describe the formation of amyloid fibers as a series of assembly reaction steps, focusing primarily on the stoichiometry and the reversibility (or irreversibility) at each step. These dynamic models can thus be seen as coarse-grained models as they do not take the molecular details of reactions into account.

A decisive factor in the choice of model to use (thermodynamic or kinetic) is whether the process is observed experimentally to be reversible or irreversible [262]. Indeed, if a process is observed to be reversible, the system can reach a dynamic equilibrium given enough time. In this case, a statistical or thermodynamic model is clearly preferable. If, however, a process is observed to happen irreversibly, the system is under kinetic control

as all quantities evolve in time through reaction steps more or less irreversibly[8] and a kinetic model is required. Of course, a correct kinetic model should be able to lead to the right equilibrium allowing the evaluation of all thermodynamic properties, but this adds to the difficulty of constructing a model.

I.6.2 Discussion of the models

All models describing secondary nucleation processes predict an exponential growth curve as well as a lag time depending weakly on the initial monomer concentration. These predictions are consistent with experimental observations for both fragmentation models and heterogeneous nucleation models. Nonetheless, some elements are lacking from the models. In the case of fragmentation, the brittleness of fibers must be taken into account in order to evaluate the potential for a protein to undergo fragmentation. A measure of brittleness could be derived from the biochemical composition of proteins and could be integrated to current models describing fragmentation. The presence of oligomers may also affect the relation between the observed lag time and the initial monomer concentration. As suggested by Lomakin *et al.* [103, 241] and collaborators, a weak dependency could be due to the rapid equilibration between monomers and non-fibrillar oligomers. It would hence be interesting to characterize fully the relation between oligomer concentration and lag time under fragmentation. In the case of heterogeneous nucleation, it has been suggested that not only do heterogeneous nuclei form from fibers, but also protofibers might undergo a heterogeneous nucleation process for some proteins [29, 122]. Hence, a model combining the oligomerization of proteins coupled to the theory of heterogeneous nucleation would be required.

The presence of oligomers, although unusual in a classical nucleation process, is a well-accepted fact for amyloid proteins. Multiple theories have been proposed, but all of the models presented here agree that the presence of intermediary species during the fibrillization process is one cause for amyloid polymorphism. However, many contradic-

[8] In a kinetic model, reaction steps are not necessarily irreversible, For instance, the nucleation reaction is rarely seen as a reversible process in models, while the elongation process forming fibers or other intermediary species can be reversible.

tions remain between the various oligomerization models because each sequence seems to display a unique mechanism, making generalization very difficult.

The physicochemical properties of proteins also play an important role in the amyloid nucleation process and many studies have investigated the role of various biochemical factors on the efficiency of fibril formation. What is fascinating about amyloid proteins is the fact that completely different amino acid sequences can contribute positively to the nucleation process of fibrillization, providing that they follow a pattern alternating hydrophobic and hydrophilic amino acids. Experimental observations support the hypothesis that hydrophobicity and the presence of aromatic residues facilitate fiber formation, whereas charged residues slow it down. Models based on experimental results efficiently describe these phenomena and shed light on the consequences of mutagenesis on the formation of fibrils and on polymorphism. However, aggregation efficiency and aggregate morphologies do not depend solely on the intrinsic properties of proteins. Extrinsic properties such as temperature, pressure, pH, solvent ionic strength, oxidation potential, and protein concentration can also influence protein misfolding, nucleation, and fiber elongation, and must be taken into account. These factors should be and often are included in amyloid models.

Finally, only *in vitro* studies have been used in the development of growth models. Numerous questions, such as the influence of surfaces on amyloid aggregation (through heterogeneous nucleation catalyzed by a surface), remain important as fibrillization may arise near cellular membranes *in vivo*. Such aspects have yet to be included in these models.

I.7 Conclusion

Synthesizing all of the observations, it seems that amyloid proteins find a large spectrum of methods to survive and to accelerate the nucleation and elongation process. These methods, from the presence of secondary nucleation processes to the formation of intermediary species during the formation of fibrils, are guided by the biochemical nature of the amino acid sequence of proteins and seem different from one protein to

another. Amyloid proteins possess a high degree of plasticity and mechanistic freedom and can grow under many different cellular conditions [263]. This property of amyloid proteins makes nucleation more complex compared with a classical process.

The models presented here convey a global picture for amyloid fibril growth, but few aspects are treated at a molecular scale because of the technical challenge of observing atomic details of insoluble amyloid fibers. The computing power now available allows us to obtain more atomic details by means of atomistic simulations. For example, we can run simulations with enough particles, and over timescales long enough to observe a nucleation growth as well as fiber growth. As a result, numerical tools are necessary to complement experimental studies, in order to develop more accurate and complete models, incorporating molecular details of the nucleation and growth of amyloid proteins.