



**HAL**  
open science

# Effets des mutations sur la structure et la dimérisation de l'alpha-synucléine et implications pour la maladie de Parkinson : réponses apportées par la dynamique moléculaire à gros grains

Adrien Guzzo

► **To cite this version:**

Adrien Guzzo. Effets des mutations sur la structure et la dimérisation de l'alpha-synucléine et implications pour la maladie de Parkinson : réponses apportées par la dynamique moléculaire à gros grains. Physique [physics]. Université Bourgogne Franche-Comté, 2022. Français. NNT : 2022UBFCK024 . tel-03884160

**HAL Id: tel-03884160**

**<https://theses.hal.science/tel-03884160>**

Submitted on 5 Dec 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**THÈSE DE DOCTORAT DE L'ÉTABLISSEMENT UNIVERSITÉ DE  
BOURGOGNE FRANCHE-COMTÉ**

**PRÉPARÉE AU LABORATOIRE INTERDISCIPLINAIRE CARNOT DE  
BOURGOGNE**

École doctorale n°553  
Carnot Pasteur

Doctorat de physique

Par

Mr. Adrien Guzzo

**Effets des mutations sur la structure et la dimérisation de  
l'alpha-synucléine et son implication dans la maladie de  
Parkinson : réponses apportées par la dynamique  
moléculaire à gros grains.**

Thèse présentée et soutenue à Dijon, le 9 juin 2022

Composition du jury :

Mr. Mounir Tarek	Docteur, Université de Lorraine	Président du jury, Rapporteur
Mr. Fabio Sterpone	Docteur, Université Paris-Cité	Rapporteur
Mr. Eric Bourillot	Maître de conférences, Université de Bourgogne	Examineur
Mr. Marco Cecchini	Maître de conférences, Université de Strasbourg	Examineur
Mme Catherine Etchebest	Professeure, Université Paris-Cité	Examinatrice
Mr. Patrick Senet	Professeur, Université de Bourgogne	Directeur de thèse
Mr. Gia Maisuradze	Senior Research Associate, Cornell University	Co-encadrant



# Remerciements

Je tiens tout d'abord à remercier les membres de mon jury, Fabio Sterpone, Mounir Tarek, Marco Cecchini, Catherine Etchebest et Eric Bourillot pour leur intérêt porté à mon travail et d'avoir accepté de le juger.

Un grand merci à Patrick Senet de m'avoir fait confiance pour cette thèse. J'ai grandement apprécié travailler avec vous pendant ces 3 années. Vos précieux conseils, votre bienveillance et votre sympathie ont rendu cette expérience très agréable et très enrichissante pour moi. Je remercie également les membres de l'équipe que j'ai pu côtoyer pendant ma thèse notamment Patrice Delarue qui a toujours eu la solution à tous mes problèmes avec une efficacité impressionnante. C'était toujours un grand plaisir de discuter avec vous et sans vous je ne crois pas que j'aurais réussi à faire autant de choses. Merci à Ana Rojas, pour ce super séjour que j'ai passé à New York malgré ce froid glaciale. Merci pour le temps que vous m'avez consacré, vos conseils, votre gentillesse et votre constante bonne humeur. C'était un réel plaisir de partager ces quelques jours avec vous. Merci à Gia Maisuradze. Votre passion sans faille dans ce travail était un véritable moteur pour me pousser à aller plus loin. Vos conseils et votre expertise mais également votre bonne humeur était toujours très agréable lors de nos longues visios pour discuter des résultats. Merci également pour ce très bon repas que nous avons partagé ensemble à New York! Merci également à tous les doctorants et les stagiaires qui ont pu travailler avec moi dans l'équipe. Merci à tous mes camarades doctorants du labo mais aussi ceux avec qui j'ai gardé contact, je leur souhaite à tous de réussir leur thèse pour ceux qui ne l'ont pas encore passé et de trouver un super travail pour ceux qui l'ont passé! Merci à Corentin mon compère de master. On a essayé de se soutenir mutuellement malgré la distance surtout dans la très dure partie de la rédaction. Je te souhaite de bien réussir cette dernière étape de rédaction et de continuer dans la voie qui te plaît.

Merci à ma Mathilde, de me supporter tous les jours alors que je n'ai pas toujours été très agréable (surtout sur la fin...). Heureusement que tu étais là pour me soutenir et me pousser à finir. Je sais que ça a été long (trop long même) mais j'espère quand même t'avoir rendue fière. J'ai hâte de découvrir cette nouvelle vie qui nous attend désormais. Merci à mes amis, Valentin (et ??) , Loïc (et Fanny), Alexis (et ??), Jérôme (et ??), Victor (et Julie) et Antoine (et ??). Cette thèse vous a semblé interminable mais bon grâce à vous j'ai pu parfois me sortir un peu la tête de tout ça et je vous en remercie grandement.

Enfin ma famille ... je ne serais probablement pas aussi bon que ma soeur Mathilde mais je vais essayer... Merci à mes parents qui m'ont aussi transmis ce gène de la science. Moi qui me demandait étant petit si il existait d'autres métiers que scientifique finalement je me suis bien fait avoir aussi! Merci à mon père qui m'a toujours soutenu dans tous les domaines. Merci de m'avoir toujours conseillé et surtout secoué parfois... Merci à ma mère qui m'a aussi beaucoup aidé et soutenu. Merci pour la relecture de la thèse qui n'était vraiment pas un cadeau. Merci à ma grande soeur Mathilde, et oui du coup tu y es bien dans mes remerciements de thèse! Merci pour ton aide, tes conseils et la relecture qui ne devait pas être évidente. Merci à ma moyenne soeur Maude qui a toujours les mots justes et doux pour me secouer et me remotiver. Merci à mes beaux-frères Rahim et Samuel, grâce à vous si je ne parviens pas à poursuivre dans la science je pourrais toujours me reconvertir en déménageur. Je remercie ma grand mère pour ses bons petits plats et sa constante aide. Je remercie également mon grand-père qui était là au début et que je n'oublie pas.

# Table des matières

<b>1 Introduction</b>	<b>7</b>
1.1 Les maladies neurodégénératives	7
1.2 L' $\alpha$ -synucléine	7
Lien avec la maladie de Parkinson	7
Les différentes formes structurelles prises par l' $\alpha$ -syn	8
Séquence d' $\alpha$ -synucléine	12
1.3 Portion atypique de la séquence de la protéine	13
1.4 Impact du motif répété sur la multimérisation	14
1.5 Les simulations d' $\alpha$ -synucléine	15
1.6 Récapitulatif des propriétés de l' $\alpha$ -synucléine et problématiques	16
<b>2 Méthodes</b>	<b>19</b>
2.1 Simulations des protéines	19
Programme de simulation de dynamique moléculaire : UNited-RESidue	19
La méthode du Replica Exchange	22
Les données de simulations	23
2.2 Détections des structures secondaires	25
Détection des hélices et des feuillets $\beta$	26
Détection de structures particulières	27
2.3 Analyses des simulations	30
Densité de probabilité des couples hélices/feuilleter $\beta$	30
Analyses rattachées à la notion de contact	31
<b>3 Résultats : Etude globale des monomères d'<math>\alpha</math>-synucléine</b>	<b>35</b>
"Missense Mutations Modify the Conformational Ensemble of the $\alpha$ -Synuclein Monomer Which Exhibits a Two-Phase Characteristic" <a href="#">[1]</a>	35
<b>4 Résultats : Etude globale des dimères d'<math>\alpha</math>-synucléine</b>	<b>53</b>
"Wild-type $\alpha$ -synuclein and variants occur in different disordered dimers and pre-fibrillar conformations in early stage of aggregation"	53
<b>5 Résultats : Etudes du segment 38-53 de l'<math>\alpha</math>-synucléine</b>	<b>89</b>
5.1 Comportement des angles diédraux des résidus du segment	89

5.2	Mise en évidence des trois types de structures principales	91
	La classe hairpin : les différentes conformations possibles	92
	La classe tresse	97
	La classe aléatoire	98
5.3	Etude approfondie de la tresse	99
	Résidus favorables à la formation d'une tresse	99
	Simulation du segment 38-53 seule avec UNRES	100
	Les différentes portions du segment 38-53 : impact de chaque portion sur la formation potentielle d'une tresse	101
	Stabilité de la tresse avec un champ de force tous atomes	103
	Formation de tresse chez $A\beta$	104
5.4	Généralisation de l'étude des boucles dans les monomères d'alpha-synucléine	106
5.5	Analyses des trois types de structures dans les dimères	108
5.6	Conclusion : Études du segment 38-53 de l' $\alpha$ -synucléine	111
<b>6</b>	<b>Resultats : Mutations répétées sur le motif KTKEGV</b>	<b>113</b>
6.1	Les mutants	113
6.2	Etude des monomères	115
	Probabilité de formation de structures secondaires le long de la séquence	115
	Densité de probabilité des conformations en fonction du nombre de résidus en feuillet $\beta$ et en hélice	117
	Classification global : Hairpin, Tresse, Aléatoire	119
	Lien avec la forme globale de la protéine	120
6.3	Etude des dimères	122
	Régions d'agrégation pour la formation de dimères	122
	Conformations particulières	124
6.4	Conclusion : Mutations répétées sur le motif KTKEGV	131
<b>7</b>	<b>Discussion</b>	<b>133</b>
7.1	Hétérogénéité des structures de l' $\alpha$ -synucléine	133
7.2	Les points d'accroche des DTFs	135
7.3	Elongation des CTFs dans les DTFs	138
7.4	Identification de résidus clés dans l'agrégation de l' $\alpha$ -syn	140
7.5	Les différences monomères-dimères	140
7.6	Nouvelle structure secondaire : la tresse	141
7.7	Origine de la toxicité	142
<b>8</b>	<b>Conclusion et perspectives</b>	<b>145</b>

# Lexique

Dans ce manuscrit plusieurs abréviations seront utilisées pour évoquer des termes spécifiques ou des groupements d'éléments. La liste exhaustive est donnée ci dessous pour faciliter la lecture.

AA : Acides aminés

WT : Protéine sauvage de l' $\alpha$ -synucléine

CTF : Contact type fibrille. Contact entre deux résidus d'index identique entre deux monomères pour former un dimère (en anglais Native Fibril contact, Nfc)

DTF : Dimère type fibrille. Dimère qui forme au moins 5 CTFs sur des résidus d'index consécutifs (en anglais dimers with fibril native contacts, Dfnc)

FTF : Feuillet- $\beta$  de type fibrille. Feuillet- $\beta$  entre deux résidus d'index identique entre deux monomères pour former un dimère

PDB : *Protein Data Bank*. Un code PDB est attribué à chaque structure présente au sein de la banque

IDP : *Intrinsically disordered protein*. Protéine intrinséquement désordonnée

DSSP : *Define Secondary Structure of Proteins*. Algorithme de définition des structures secondaires dans les protéines

STRIDE : *Structural identification*. Algorithme de définition des structures secondaires dans les protéines

PULCHRA : *Protein Chain Reconstruction Algorithm*

CUTABI : *CUrvature and Torsion based of Alpha-helix and Beta-sheet Identification*

PSEA : *Protein Secondary Element Assignment*



CM : *Contact map*. Carte de contact

FTIR : *Fourier-transform infrared spectroscopy*. Spectroscopie infrarouge à transformée de Fourier

FRET : *Förster resonance energy transfer*. Transfert d'énergie par résonance de fluorescence

AFM : *Atomic force microscopy*. Microscopie à force atomique

RMN : Résonance magnétique nucléaire.

PRE : *Paramagnetic relaxation enhancement*. Résonance paramagnétique électronique

SANS : *Small-angle neutron scattering*. Diffusion des neutrons aux petits angles

SAXS : *Small-angle X-ray scattering*. Diffusion des rayons X aux petits angles

NAC : *Non Amyloid Component*

# Chapitre 1

## Introduction

### 1.1 Les maladies neurodégénératives

Le 21<sup>ème</sup> siècle est confronté à l'augmentation de maladies du cerveau émergentes liées au vieillissement de la population : les maladies neurodégénératives. Des maladies chroniques progressives touchent le système nerveux central des patients et entraînent un grand nombre de complications telles que l'invalidité, ou la dépendance allant jusqu'à l'hospitalisation. Parmi ce groupe de maladies, on retrouve la maladie d'Alzheimer qui touche plus d'un million de personnes en France, la maladie de Parkinson avec environ 160 000 personnes traitées et des maladies du motoneurone causées par la sclérose latérale amyotrophique (SLA).

Les traitements actuels contre ces maladies ont une efficacité variant d'un patient à l'autre et ont un impact uniquement sur les symptômes. Il n'existe pas à ce jour de traitement de prévention. Il existe toutefois des recommandations pouvant limiter l'apparition précoce de ces maladies en ayant un mode de vie sain rythmé par une activité physique régulière et une alimentation variée. Face à ces fléaux, de nombreuses équipes de recherche partout dans le monde travaillent dans le but d'apporter des réponses sur les origines de ces maladies. La piste principale jouant un rôle central dans la maladie de Parkinson est la protéine appelée  $\alpha$ -synucléine ( $\alpha$ -syn). De nombreuses autres maladies neurodégénératives, appelées synucléopathies, sont également liées à cette protéine [2, 3, 4, 5, 6].

### 1.2 L' $\alpha$ -synucléine

#### Lien avec la maladie de Parkinson

Le phénomène d'agrégation de la protéine  $\alpha$ -syn est au centre de la problématique de la maladie de Parkinson. On retrouve ce genre de problématique également pour la maladie d'Alzheimer avec l'agrégation de la protéine  $A\beta$  [7, 8, 9]. L' $\alpha$ -syn est une petite protéine de 140 acides aminés classée dans la catégorie des protéines intrinsèquement désordonnées (*intrinsically disordered protein* (IDP)) [10, 11, 12]. On retrouve cette protéine très abondamment dans le cerveau [13] mais ses fonctions principales sont encore mal comprises à ce jour.

Les maladies liées à l' $\alpha$ -syn se caractérisent par une très forte concentration de cette protéine sous forme de filaments dans des inclusions sphériques, appelées corps de Lewy, situées dans le cytoplasme des neurones [14, 15, 16, 17].

La maladie est sporadique pour une grande majorité des cas, cependant des facteurs génétiques peuvent intervenir et faciliter la formation de la maladie de deux façons :

- La protéine sauvage (*wild-type* (WT)) est surexprimée en raison de la triplication du gène SNCA.
- Une mutation pathogène ponctuelle intervient sur le gène. On en recense plusieurs H50Q, G51D, A53E, A30P, A53T et E46K [18, 19, 20, 21, 22, 23, 24].

Le terme "toxicité" est complexe à aborder lorsqu'on traite de la protéine  $\alpha$ -syn. Le WT, comme les mutants, sont considérés comme potentiellement toxiques puisqu'ils peuvent engendrer la maladie. Cependant, on peut considérer certains mutants comme plus toxiques que d'autres par différents facteurs. Le principal facteur est la facilité de formation d'inclusions dans les cellules engendrée par la protéine. La perméabilisation des membranes entraînée par certaines structure de l' $\alpha$ -syn peut également être un autre facteur. Les mutations E46K et A53T, étudiés dans cette thèse, entraînent des formes plus précoces de la maladie de Parkinson. Le mutant E46K augmente considérablement la quantité d'inclusions dans les cellules et altère le schéma d'agrégation [25]. Le mutant A53T ne présente pas plus d'inclusions que le WT [25], cependant, dans une étude sur des vers, la vitesse et la fréquence des flexions du corps a été fortement réduite en présence du mutant A53T et le taux de paralysie a fortement augmenté [26]. Enfin le mutant A30P, également étudié dans cette thèse, réduit la quantité de corps d'inclusions dans les cellules et facilite la formation d'oligomères plutôt que d'inclusions [25]. Une étude expérimentale par microscopie à force atomique (AFM) montre qu'un mélange de mutants A53T forme plus rapidement des fibrilles qu'un mélange de protéines sauvages et qu'un mélange de mutants A30P ralentit la formation des fibrilles [27].

## Les différentes formes structurelles prises par l' $\alpha$ -syn

L' $\alpha$ -syn est qualifiée de protéine caméléon par le très grand nombre de conformations différentes qu'elle est capable d'adopter [28, 29]. Dans des conditions physiologiques, on va majoritairement la retrouver sous forme désordonnée. Elle est capable de former des structures secondaires, notamment des hélices- $\alpha$  sous forme d'une structure tétramérique [30, 31] en équilibre avec une forme monomérique, riche également en hélices- $\alpha$ , en interaction avec une membrane cellulaire. [32].

*In vitro*, dans des conditions physiologiques, l' $\alpha$ -syn peut s'agréger en formant des structures appelées "fibrilles amyloïdes" constituées de plusieurs milliers de monomères. De nombreuses études montrent également le fort impact de structures transitoires sur la toxicité de la protéine comme par exemple les oligomères constitués de plusieurs dizaines de monomères ou

encore les dimères. Une multitude de méthodes expérimentales ont permis d'étudier ces différentes structures. Nous donnons ici quelques résultats majeurs pour chaque famille de structure.

### - Les fibrilles et les proto-filaments

Les proto-filaments sont constitués d'une succession de couches de monomères d' $\alpha$ -syn formant chacune des feuillets  $\beta$  avec ses voisins de manière parfaitement alignée c'est à dire que les mêmes résidus de monomères adjacents se font faces. L'association de plusieurs proto-filaments est appelée "fibrille", par abus de langage on nomme également un proto-filament isolé comme étant une fibrille. Plusieurs types de conformations possibles ont été observés dans des expériences *in vitro* dont le but est de reproduire le comportement d'agrégation d' $\alpha$ -syn *in vivo* [33, 34, 35]. Deux exemples de ces proto-filaments sont donnés à la figure 1.1. Le nombre de couches de monomères a été réduit à deux uniquement pour la visibilité de la représentation 3D. La région de formation des feuillets  $\beta$  se situe entre les résidus 35-97 pour la structure de code PDB 2n0a [33] et 37-92 pour la structure de code PDB 6ssx [34]. Les deux types de structures présentent la formation de différentes boucles permettant un repliement de la structure sur elle même.

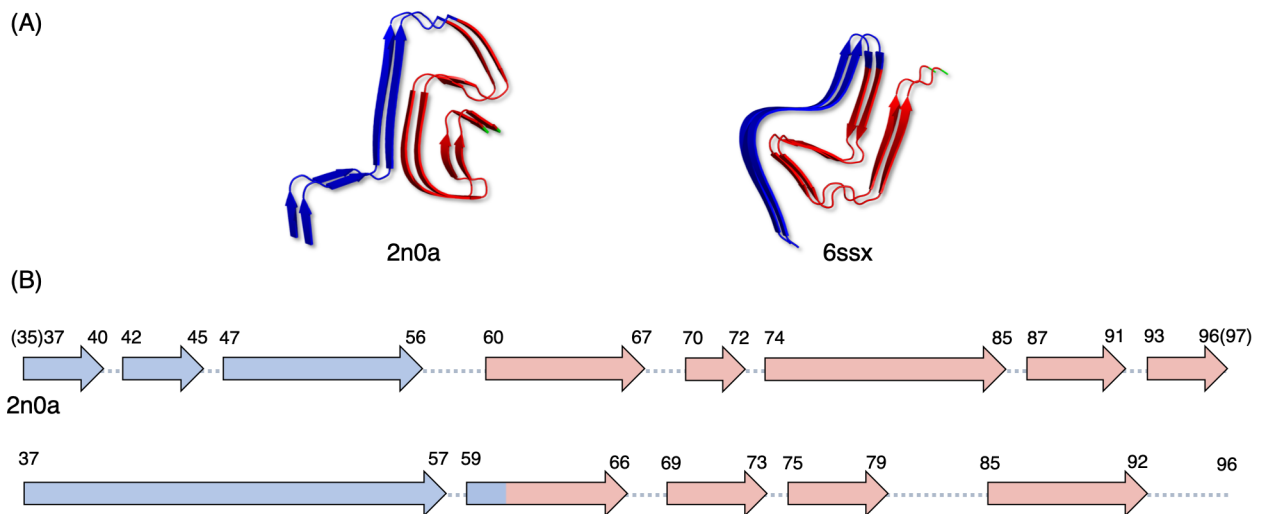


FIGURE 1.1 – panneau (A) : Représentations 3D de deux types de proto-filaments découverts expérimentalement. Uniquement deux monomères sont représentés ici du résidu 37 au 96. Le Nterm est coloré en bleu, le NAC (*Non Amyloid Component*) en rouge et le Cterm en vert (seulement un résidu est visible). panneau (B) : Mise en évidence des portions formant des feuillets  $\beta$  inter-chainés dans la séquence des deux conformations de fibrilles

Les fibrilles sont composées de milliers de monomères d' $\alpha$ -syn [9], ils sont donc plus faciles à analyser expérimentalement grâce à leur taille. Une étude directement réalisée lors de l'autopsie d'un patient atteint de la maladie de Parkinson a permis de déterminer par la méthode de diffraction des rayons X la présence de multitude de feuillets  $\beta$  [36]. En utilisant la méthode de cryo-microscopie électronique sur des patients atteints d'atrophie multisystémique, deux types de proto-filaments associés à des molécules non-protéiques ont été découverts. Ce résultat a

permis de mettre en évidence la différence des filaments présents dans ce type de maladie et dans la maladie de Parkinson montrant ainsi la grande diversité de conformères possibles pour cette protéine [37]. D'autres types de fibrilles sous forme de rubans ont également été découverts [38]. La technique de résonance magnétique nucléaire (RMN) a permis d'étudier l'agrégation de l' $\alpha$ -syn sur une vésicule unilamellaire. Le monomère initialement sous sa forme riche en hélices est capable de se réarranger pour prendre une forme en L semblable à la forme retrouvée dans les fibrilles matures [39]. En solution, des intermédiaires riches en hélices peuvent se former durant la transition observée lors de la croissance d' $\alpha$ -syn des structures désordonnées en fibrilles riches en feuillets  $\beta$  [40]. L'impact des structures prises par l' $\alpha$ -syn sur la perméabilisation membranaire est un critère de toxicité souvent étudié expérimentalement. Il est suggéré que les proto-filaments riches en feuillets  $\beta$  liés à des vésicules synthétiques perméabilisent transitoirement ces vésicules et les détruisent [41].

Les mutations ont un effet direct sur la formation et l'arrangement des fibrilles *in vitro* [35]. A53T et E46K vont accélérer le processus de formation des fibrilles tandis que A30P aura tendance à le ralentir comparé au WT [42, 43]. La formation de petits oligomères se différencie également entre les mutants et le WT [42]. A30P et A53T ont une plus grande tendance à former des agrégats non fibrillaires par rapport au WT [44].

### - Les oligomères

Le mécanisme de formation des fibrilles est encore peu connu à ce jour. Le processus de formation est relativement long et pourrait faire intervenir des structures transitoires oligomériques formées de plusieurs dizaines de monomères. Ces petits oligomères dont la structure est hétérogène semblent se consumer à mesure que les fibrilles se forment [8] et à l'inverse apparaître lorsque des fibrilles se désagrègent [38].

Un très grand nombre de techniques expérimentales ont permis d'étudier ces petits oligomères. De nombreuses caractéristiques différentes ont été identifiées dans ces travaux. La liste de l'ensemble des résultats est donnée ci dessous.

La méthode de diffusion des neutrons aux petits angles (SANS) a montré la présence d'oligomères sphériques métastables qui ne comportent pas de structures secondaires particulières, avec un diamètre de 10 nm [45]. Des oligomères de taille similaire composés de 30 monomères, de forme ellipsoïdale ont été mis en évidence par la méthode de diffusion des rayons X aux petits angles (SAXS). Ils sont composés d'un noyau compact et d'une enveloppe souple. Ces oligomères semblent empêcher la nucléation primaire et les étapes d'élongation des fibrilles [46]. Deux groupes d'oligomères composés de 18 et 29 monomères ont été identifiés sous forme de pore par microscopie cryo-électronique. Un grand nombre de feuillets  $\beta$  anti-parallèles ont été observés sur plus de 35% de la protéine [47].

L'analyse par spectroscopie infrarouge valide la forte proportion de feuillets  $\beta$  anti-parallèles dans des oligomères contraints de se réarranger pour former des fibrilles possédant des feuillets  $\beta$  parallèles [48]. Une autre étude par AFM démontre l'hétérogénéité des oligomères capables de

prendre plusieurs types de formes sur des surfaces solides comme une forme sphérique, allongée ou annulaire [27].

Les oligomères sont considérés comme plus toxiques que les fibrilles [27, 25]. Des observations d'oligomères annulaires sous forme de tube constitués de feuillets  $\beta$  perturbent l'intégrité de la membrane [41, 49]. La perméabilisation des membranes est facilitée par l'exposition des chaînes latérales hydrophobes des oligomères au solvant et entraîne un dysfonctionnement des cellules. Les oligomères sont également responsables d'une production excessive d'espèces oxygénées qui entraîne la mort cellulaire [50]. Les fibrilles préformées de l' $\alpha$ -syn associées à des protéines transmembranaires facilitent l'endocytose d'oligomères toxiques [51].

### - Les dimères

Parmi les oligomères, l' $\alpha$ -syn forme également des structures très simples : les dimères. Ces petites entités ne doivent pas être négligées car elles peuvent jouer un rôle important dans la toxicité de l' $\alpha$ -syn. L'accumulation de dimères favorise la formation de fibrilles. Ce processus semble être accéléré par les mutations A30P et A53T [52].

La mise en évidence de dimères dans la phase de latence d'agrégation de l' $\alpha$ -syn a été réalisée par le biais de plusieurs méthodes. La microscopie de fluorescence par réflexion totale interne a mis en évidence deux types de dimères qui diffèrent par leur durée de vie. Pour le WT, les moins stables ont une durée de vie d'environ 200-300 ms tandis que les plus stables peuvent aller jusqu'à 3 s. A30P augmente significativement la durée de vie des deux types de dimères avec 700 ms et 5 s, respectivement [53].

Les études par AFM d' $\alpha$ -syn sur des surfaces solides ont confirmé la présence des dimères les plus stables découverts par la méthode précédente [54]. A faible pH, des différences entre les dimères du WT et des mutants ont été observées. Différents segments d'interactions entre les monomères sont favorisés par A53T et E46K comparé au WT. A30P favorise au contraire un seul segment d'interaction [55]. Le monomère adopte une structure compacte sphérique mais connaît des structures intermédiaires plus étendues. Les dimères sont composés majoritairement de deux monomères sphériques mais aussi d'un monomère sphérique et d'un monomère allongé en quantité moindre du moins sur des surfaces [56].

La méthode de transfert d'énergie entre molécules fluorescentes (FRET) a montré que la population des dimères ayant une longue durée de vie est plus grande pour les mutants que pour le WT. La concentration des dimères pendant la phase de latence d'agrégation est identique pour le WT et les mutants. Les auteurs suggèrent donc que ce n'est pas la quantité de dimères qui serait le paramètre le plus pertinent mais les propriétés structurales de ces dimères pendant la phase d'auto-assemblage dans le processus de neuro-dégénérescence [42].

L'ensemble de ces données montre qu'il existe des différences significatives entre le WT et ses mutants à l'échelle des dimères et plus particulièrement pour A30P.

- Le tétramère

L' $\alpha$ -syn ne posséderait pas que des structures agrégées toxiques. En effet, une forme tétramérique riche en hélices a été observée dans différentes études expérimentales et jouerait un rôle protecteur dans la toxicité de l' $\alpha$ -syn [57, 58, 59]. De la dynamique moléculaire associée à des données RMN ont montré que des structures en trimères et en tétramères riches en hélices ou en feuillets  $\beta$  sont possibles. Ces structures, bien qu'elles possèdent une certaine stabilité, ne représenteraient qu'une minorité des structures de l' $\alpha$ -syn en solution avec moins de 10% des conformations [60]. Des simulations de dynamique moléculaire montrent que le tétramère serait stabilisé par des interactions hydrophobes et des ponts salins entre les monomères [59]. L'incapacité à former un ensemble de multimères comprenant des oligomères de différentes tailles et des tétramères est suggérée comme étant une cause de la toxicité de l' $\alpha$ -syn [58, 30].

Séquence d' $\alpha$ -synucléine

La protéine  $\alpha$ -syn est présente abondamment dans le cerveau et plus particulièrement dans les neurones, elle se compose de 140 acides-aminés (AA) et se divise en trois régions :

- Le N-terminal (Nterm) (1-60) contient un motif répété KTKEGV, de manière imparfaite plusieurs fois dans la séquence (résidus en rouge [figure 1.2](#)). Cette région est également chargée positivement avec 11 lysines (chargées +), 6 acides glutamiques (chargés -) et un acide aspartique (chargé -) amenant la charge globale du motif à +4.

- Le NAC (61-95) pour *Non Amyloid Component* est la région centrale très hydrophobe de la protéine. Elle est constituée d'un grand nombre de valines et d'alanines représentant pratiquement la moitié des résidus dans cette région (15/34).

- Le C-terminal (Cterm) (96-140) est une région très flexible de la protéine majoritairement déstructurée. Elle est composée de 15 résidus chargés négativement (5 acides aspartiques et 10 acides glutamiques) et de 3 lysines chargées positivement amenant une charge globale à -12.

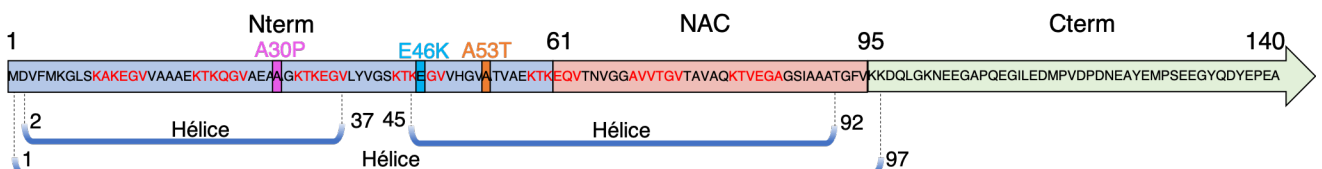


FIGURE 1.2 – Schéma récapitulatif de la séquence de l' $\alpha$ -syn et de ses caractéristiques. Les trois régions principales sont données en bleu, rouge et vert pour le Nterm, le NAC et le Cterm respectivement. Le motif imparfait KTKEGV est mis en évidence en rouge dans la séquence. La position des mutations ponctuelles A30P, E46K et A53T sont données avec le code couleur utilisé tout au long de ce manuscrit. La position des hélices formées lorsque la protéine est liée à une membrane sont données en bleu sous l'axe.

La séquence atypique de chacune de ces régions leur permet d'avoir de nombreuses propriétés différentes. En effet, le motif KTKEGV présent dans le Nterm est responsable de la possible liaison de la protéine avec une membrane. Celle-ci se fait sous deux formes différentes, une première révélée par RMN à haute résolution présente deux hélices 2-37 et 45-92 séparées par une boucle [61] lorsque la protéine est liée à un micelle, et une seconde déterminée par résonance paramagnétique électronique (PRE) en interaction avec une membrane lipidique formant une hélice étendue 1-97 [62]. Le Nterm joue un rôle dans l'agrégation de la protéine en solution lorsqu'il est exposé au solvant [43]. Le NAC contient un segment très hydrophobe 71-82 jugé nécessaire à l'agrégation de la protéine [63]. Un lien a d'ailleurs été fait entre la formation de fibrilles dans le NAC et une autre maladie neurodégénérative : la maladie d'Alzheimer [64]. Le Cterm possède plusieurs fonctions de part sa grande flexibilité avec notamment la régulation de la formation de fibrilles *in vitro* [65]. Le Cterm et le Nterm s'opposent en terme de charge, il est suggéré que leur interaction électrostatique joue un rôle dans les processus secondaires de nucléation des fibrilles [66]. De multiples contacts transitoires à longue distance entre le Cterm et les deux autres régions ont été observés par PRE [67]. Il a été suggéré que ces contacts sont protecteurs et empêchent l'agrégation. Cependant, ces contacts à longue distance entre le Cterm et le Nterm ne représentent qu'environ 14% de l'ensemble conformationnel. Dans ces conditions la région 68-78 du NAC est exposée au solvant ce qui permettrait son agrégation malgré ces contacts à longue distance [28].

### 1.3 Portion atypique de la séquence de la protéine

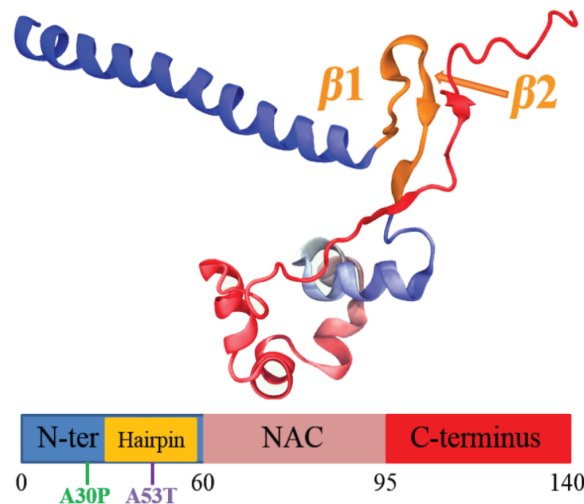


FIGURE 1.3 – Représentation 3D de la hairpin découverte dans la région 38-53 (en orange) dans la publication Yu et al. [68]

Des simulations de dynamique moléculaire ont révélé la présence d'une hairpin (épingle à cheveux en français) sur le segment de la séquence 38-53 de l' $\alpha$ -syn [68]. La représentation 3D de cette structure extraite de la publication [68] est donnée figure 1.3. Des simulations ont été réalisées pour le WT et les mutants A30P et A53T dont les mutations sont proches de la région



de la hairpin [?]. En comparant les différentes protéines, les auteurs se sont aperçus que la hairpin se replie bien plus vite pour les mutants que pour le WT. La hairpin formée interagit dans la majorité des cas avec la portion 118-130 du Cterm sous forme de feuillets  $\beta$ . Les auteurs suggèrent que la hairpin pourrait être une région favorable pour réaliser un empilement de feuillets  $\beta$  semblables à ceux retrouvés dans les fibrilles. Le repliement du Cterm pourrait être une protection contre cette agrégation.

## 1.4 Impact du motif répété sur la multimérisation

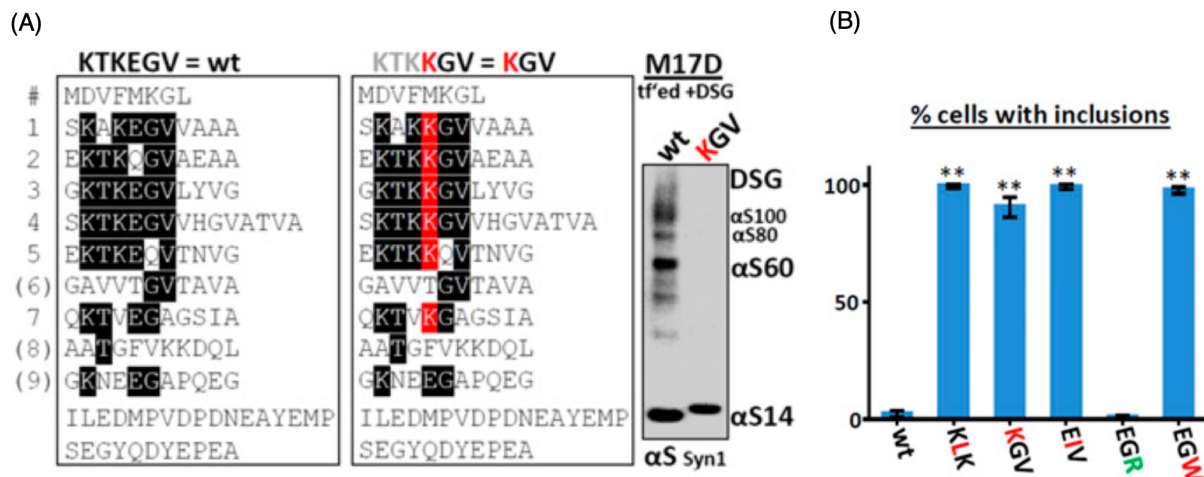


FIGURE 1.4 – Panneau (A) : Séquence du WT et d’un mutant comportant des mutations répétées sur le résidu en 4<sup>ème</sup> position du motif. La figure de droite montre les différents états multimériques de l’ $\alpha$ -syn avec la forme monomérique légendée  $\alpha$ S14 (bande du bas) et les différentes formes multimériques présentes sur le haut de la membrane de *western blot* seulement pour le WT.

Panneau (B) : Pourcentage des cellules contenant des inclusions d’ $\alpha$ -syn pour les protéines étudiées

Des analyses *in vitro* et *in vivo* du rôle du motif répété KTKEGV sur la toxicité de l’ $\alpha$ -syn ont été réalisées [58]. Deux résultats importants sont extraits de cette publication. Le premier résultat est l’absence d’une région indispensable à la multimérisation. Plusieurs portions de la protéine ont été supprimées par groupe de 10 résidus. Malgré ces suppressions, la protéine sauvage d’ $\alpha$ -syn a continué à se multimériser. Les auteurs suggèrent que le motif KTKGEV pourrait jouer un rôle compensatoire si une partie de la protéine est manquante. Plusieurs mutants ont été synthétisés pour comprendre l’effet de ce motif tout au long de la séquence. Pour chaque mutant, les mutations sont faites sur la même position au sein du motif. Des mutations franches telles que la modification d’un acide glutamique (chargé négativement) en lysine (chargé positivement) ont été réalisées pour que l’impact soit maximal. Certains mutants ont perdu leur capacité à former des tétramères (figure 1.4 panneau A) et possèdent une grande facilité à produire des inclusions de type Lewy (toxiques) dans les cellules (figure 1.4 panneau B). A l’inverse, les protéines favorisant la formation du tétramère semblent former

bien moins d'inclusions. Les auteurs suggèrent donc que la présence des tétramères jouerait un rôle protecteur réduisant la toxicité de la protéine.

## 1.5 Les simulations d' $\alpha$ -synucléine

L'état de l'art présenté précédemment sur les différentes recherches réalisées sur la protéine  $\alpha$ -syn depuis une vingtaine d'années, nous montre l'intérêt des avancées technologiques pour mieux comprendre les protéines. De nombreuses nouvelles méthodes expérimentales ont permis de réaliser des analyses plus poussées de la protéine et de donner des résultats majeurs avec notamment les découvertes assez récentes des fibrilles. La méthode expérimentale est cependant coûteuse puisqu'elle nécessite beaucoup de matériel et connaît plusieurs limites. Les études *in vitro* imposent l'étude de l' $\alpha$ -syn en forte concentration et rendent difficile l'étude de la protéine à son échelle monomérique. De plus, les méthodes d'imagerie ne peuvent capter qu'une petite portion des nombreuses conformations que peut prendre l' $\alpha$ -syn et laisse une grande part d'inconnue sur autres les conformations possibles.

Aucune technique expérimentale n'a permis à ce jour de décrire l'aspect atomistique des dimères. Il n'existe pas encore d'idées précises des effets des mutations sur la structure des dimères. Une autre approche, complémentaire, consiste à utiliser les simulations numériques. Les avancées technologiques informatiques et la connaissance des interactions physiques à l'échelle atomique ont permis notamment de construire des programmes de dynamique moléculaire capables de reproduire le comportement des protéines., avec une précision croissante. Il existe plusieurs types de dynamique moléculaire.

La dynamique moléculaire tous atomes permet de simuler l'entièreté de la protéine à l'échelle atomique. Cette méthode, gourmande en temps de calcul, ne permet cependant pas actuellement de réaliser des simulations sur des temps supérieurs à plusieurs microsecondes avec un centre de calcul standard. Cela limite donc l'espace des conformations exploré. Dans le cas de l' $\alpha$ -syn, c'est un problème majeur car la protéine a un espace de conformation très grand de part son caractère intrinsèquement désordonné. Des temps longs de simulations sont donc nécessaires pour explorer un maximum de conformations possibles. De plus, la grande flexibilité de l' $\alpha$ -syn lui donne la possibilité d'explorer des formes très étendues, ce qui nécessite pour en rendre compte, une grande boîte de simulation d'au moins 40-50 nm impliquant un nombre considérable de molécules représentant le solvant. L'intérêt de la seconde méthode de dynamique moléculaire appelée "à gros-grains" est de pouvoir permettre de simuler la protéine en s'affranchissant de ces problèmes. Le système est simplifié en utilisant des entités regroupant plusieurs atomes. Ces entités moyennent le comportement des atomes qui les constituent. Le solvant est implicite ce qui permet de simplifier l'environnement. On pourra grâce à cette méthode réaliser des temps de simulations longs sur des monomères et même simuler des multimères.

L' $\alpha$ -syn a été étudiée à l'aide de la dynamique moléculaire, en couplant souvent les résultats avec des données expérimentales. Une étude tous atomes montre que la variation de pH fait diminuer l'interaction Cterm-NAC [69]. Des données de réticulation à courte distance cou-

plées à de la dynamique moléculaire discrète ont permis de montrer des structures globulaires compactes de l' $\alpha$ -syn comprenant des contacts inter-résidus composés de structures secondaires transitoires notamment dans le NAC et Cterm contenant des feuillets  $\beta$  et des hairpins [70]. L'interaction membrane- $\alpha$ -syn a été aussi étudiée par dynamique moléculaire et montre notamment la formation par les oligomères de structures sous forme de pores dans une membrane [71]. Plusieurs études mettent en évidence une forte proportion d'hélices dans le Nterm [72] [68] [73] tandis que d'autres études n'en montrent que peu [74] [75]. La présence de feuillets  $\beta$  sous forme d'épingle à cheveux appelé "hairpin" dans la région 38-53 est également discutée par des simulations la représentant largement [68] et d'autres montrant plutôt la formation d'hélices dans cette région [72]. Les désaccords induits par les différentes méthodes de calculs montrent la grande difficulté d'étudier cette protéine à l'aide de ces outils. Des études des dimères ont déjà été réalisées en utilisant la dynamique moléculaire ultra-simplifiée [56] ou des petits fragments d' $\alpha$ -syn [76] mais ces méthodes ne prennent pas en compte la dynamique du polymère ou biaisent les conditions initiales des simulations en utilisant des structures natives de la protéine, comme par exemple celle associée à une membrane de phospholipide [77].

## 1.6 Récapitulatif des propriétés de l' $\alpha$ -synucléine et problématiques

L'agrégation de l' $\alpha$ -syn est un processus complexe, thermodynamique, qui nécessite d'étudier l'équilibre entre les phases agrégées et libres en solution. C'est un problème multi-échelle. Dans cette thèse, nous examinons les monomères isolés et les premières étapes de l'association de deux monomères, ainsi que les différences entre le WT et les mutants en espérant identifier des régions particulières pertinentes pour l'association. Les conclusions de ces simulations seront qualitativement comparées à l'expérience et des hypothèses seront émises tout en gardant en tête les limitations de notre étude : i) système minimal d'une ou deux protéines, ii) un champ de force à gros grains et solvant implicite, iii) temps effectif de simulation de l'ordre de la milliseconde beaucoup plus petit que la phase de latence de l'agrégation (heures).

Le schéma récapitulatif donné [figure 1.5](#) permet de visualiser les différentes formes connues prises par l' $\alpha$ -syn . En partant de la forme désordonnée centrale sur le schéma la protéine est capable d'évoluer vers i) une forme riche en hélice liée à une membrane lipidique, ii) un tétramère, iii) une structure monomérique favorable à la formation de multimères évoluant en fibrilles pour finalement former des inclusions dans le cerveau.

Les questions auxquelles nous cherchons à répondre dans cette thèse sont matérialisées sur le schéma ([figure 1.5](#)) :

Q1 : Quelles sont les différentes structures prises par le monomère désordonné ?

Nous chercherons à répondre à cette question en identifiant les différentes structures secondaires (hélices, feuillets  $\beta$  et boucles) présentes chez les monomères de l' $\alpha$ -syn afin d'avoir une image globale des caractéristiques de la structure.

Q2 : Quelles structures monomériques de l' $\alpha$ -syn seraient favorables à l'agrégation ?

Parmi les structures identifiées dans la Q1, nous chercherons à détecter les structures qui semblent favorables ou défavorables à la formation de dimères.

Q3 : Quelles sont les régions favorables à la formation de dimère type fibrille chez l' $\alpha$ -syn ?

Nous chercherons à identifier les différents types de dimères comportant les caractéristiques retrouvées dans les fibrilles amyloïdes pour comprendre les étapes préliminaires d'agrégation et ainsi identifier les régions clés de l' $\alpha$ -syn .

Q4 : Existe-t-il d'autres types de dimères chez l' $\alpha$ -syn ?

Les études expérimentales ont montré la présence de différents types de dimères, nous nous attendons donc à obtenir d'autres types de dimères que ceux se rapprochant d'une fibrille.

Ces quatre questions sont les quatre axes de recherche principaux de cette thèse. Ces axes seront étudiés à la fois pour le WT mais également pour les différents mutants afin de réaliser des comparatifs et ainsi de contribuer à comprendre au mieux les causes structurelles initiales responsables des synucléopathies.

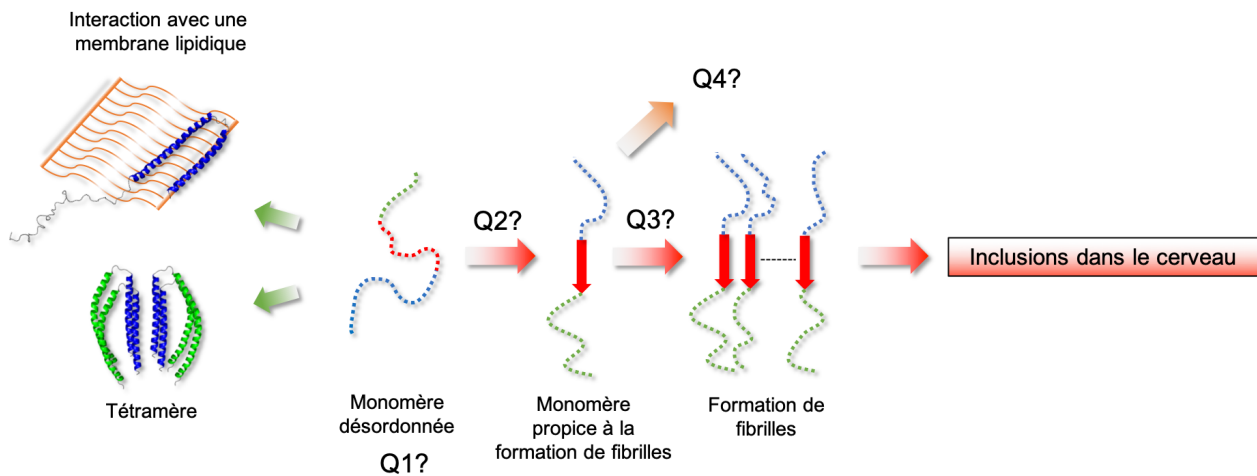


FIGURE 1.5 – Schéma représentatif des différentes conformations connues prises par l' $\alpha$ -syn .



# Chapitre 2

## Méthodes

Dans cette partie, nous expliciterons d’abord la méthode de calcul des simulations de dynamique moléculaire et les différents paramètres utilisés puis nous ferons un récapitulatif des trajectoires simulées pour les différents types de protéines. Nous détaillerons ensuite les différentes techniques d’analyse et de construction des figures utilisées à la fois dans les publications et dans les autres résultats présentés dans cette thèse.

### 2.1 Simulations des protéines

#### Programme de simulation de dynamique moléculaire : UNited-RESidue

La majorité des trajectoires de l’ $\alpha$ -syn étudiées dans cette thèse a été réalisée avec le programme UNited-RESidue (UNRES). Ce modèle à gros grains a été développé par le groupe d’Harold Scheraga à l’université de Cornell (Etats-Unis) et par Adam Liwo à l’université de Gdansk (Pologne) [78]. Ce programme simplifie la représentation d’une protéine en ne prenant que deux entités par acide aminé (figure 2.1). Une première entité notée  $P_i$  (rond noir) est située à mi-distance entre deux  $C^\alpha$  ( $1/2dC_i$ ). Une seconde entité notée  $SC_i$  (rond grisé) est reliée au  $C^\alpha$  par une liaison virtuelle  $dX_i$  positionnée au centre de masse de la chaîne latérale et donne également son orientation (angles  $\alpha_i, \beta_i$ ). Le système est donc très simplement décrit par les deux vecteurs  $dC_i$  et  $dX_i$  ou par l’ensemble des angles  $\alpha_i, \beta_i$  ainsi que les angles de liaison  $\theta_i$  et diédraux  $\gamma_i$  construits sur les liaisons virtuelles entre les  $C^\alpha$ . A partir de ces valeurs, UNRES va générer en fichier de sortie les coordonnées de 2 entités pour chaque acide aminé :  $C^\alpha$  et  $C^\beta$ .

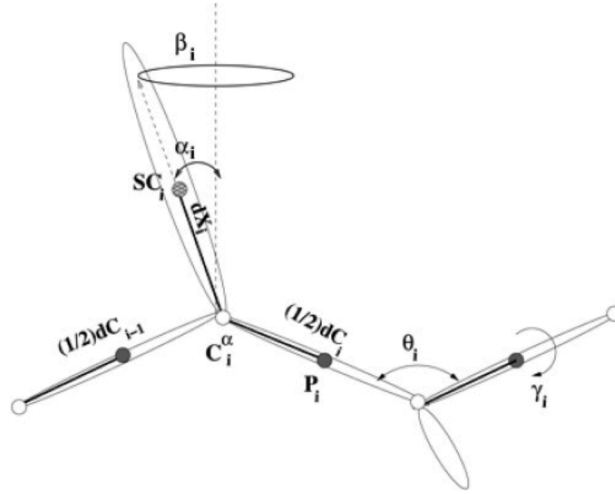


FIGURE 2.1 – Schéma du modèle simplifié d’une chaîne polypeptidique défini par UNRES. Les entités  $P_i$  (ronds noirs) et  $SC_i$  (ronds grisés) forment les deux types d’éléments utilisés par UNRES pour la simulation de la dynamique de la chaîne (voir texte). Les ronds blancs définissent les positions réelles des  $C^\alpha$ .

La réduction de la représentation de la chaîne polypeptidique, couplée à l’utilisation d’un solvant implicite, permet de réduire considérablement les temps de calculs de près de 1000 à 4000 fois comparé à de la simulation tous atomes et donne la possibilité avec UNRES de réaliser des simulations longues d’une protéine de l’ordre de la milliseconde (temps effectif).

Le champ de force a été dérivé comme un potentiel de force moyen d’une chaîne polypeptidique dans un solvant dans lequel tous les degrés de liberté ont été moyennés à l’exception des coordonnées  $P_i$  et  $SC_i$ . La fonction d’énergie contient plusieurs termes : les interactions locales ou des interactions entre les différentes régions de la protéine. Elle est obtenue en décomposant le potentiel de force moyen en plusieurs facteurs regroupés par type d’interaction à la fois entre les différents gros grains mais aussi de manière intrinsèque à chaque gros grain :

$$\begin{aligned}
 U_{chaîne-unique} = & \sum_j \sum_{i < j} U_{SC_i SC_j} + \sum_{ss} U_{Cys_{iss} Cys_{jss}} + w_{SCp} \sum_j \sum_{ij} U_{SC_i p_j} + \\
 & w_{el} \sum_j \sum_{i < j-1} U_{p_i p_j} + w_{tor} \sum_i U_{tor}(\gamma_i) + w_{tord} \sum_i U_{tord}(\gamma_i, \gamma_{i+1}) + \\
 & w_b \sum_i U_b(\theta_i) + w_{rot} \sum_i U_{rot}(\alpha_i, \beta_i) + \sum_{m=3}^6 w_{coor}^{(m)} U_{coor}^{(m)} + w_{vib} \sum_i U_{vib}(d_i)
 \end{aligned}$$

Les différentes sommes parcourent les AA  $i$  et  $j$ . Les termes  $U_{SC_i SC_j}$  concernent l’hydrophobicité ainsi que l’interaction implicite de la chaîne latérale avec le solvant [79, 80].  $U_{Cys_{iss} Cys_{jss}}$  est lié à l’énergie des liaisons disulfides (non présentes dans  $\alpha$ -syn).  $U_{SC_i p_j}$  correspond à un potentiel de volume exclu pour l’interaction entre les deux entités définies par UNRES. La moyenne des énergies d’interactions électrostatiques entre les entités  $p_i$  est donnée par le terme  $U_{p_i p_j}$ . Les potentiels de torsions et de doubles torsions  $U_{tor}$  et  $U_{tord}$  impliquent la rotation d’une ou de deux liaisons virtuelles données. Les termes  $U_b$  et  $U_{rot}$  concernent la courbure des liai-

sons virtuelles et la rotation de la chaîne latérale respectivement. Les corrélations (à différents ordres  $m$ ) entre l'électrostaticité des entités  $p_i$  et les interactions locales au sein de la chaîne principale sont données par  $U_{\text{cor.}(m)}$ . Enfin le terme  $U_{\text{vib}}$  est un simple potentiel harmonique dépendant de la taille des liaisons virtuelles ( $d_i$ ). La valeur  $w$  indique le poids de chaque terme dans l'équation.

Afin d'appliquer UNRES sur des protéines possédant plusieurs chaînes, un second potentiel inter-chaînes a été construit [81] :

$$\begin{aligned}
 U_{\text{inter-chaînes}} = & \sum_j \sum_{i < j} U_{SC_i^k SC_j^l} + \sum_{ss} U_{Cys_{iss}^k Cys_{jss}^l} + w_{SCp} \sum_j \sum_{ij} U_{SC_i^k p_j^l} \\
 & + w_{SCp} \sum_j \sum_{ij} U_{p_i^k SC_j^l} + w_{el} \sum_i \sum_j U_{p_i^k p_j^l} + \sum_{m=3}^6 w_{\text{cor,nonadj}}^{(m)} U_{\text{cor,nonadj}}^{(m)}
 \end{aligned}$$

Ce potentiel contient des termes semblables au potentiel pour une chaîne unique mais cette fois appliqués aux deux chaînes identifiées par  $k$  et  $l$ . Tous les termes ramenés à des éléments intrinsèques à la chaîne ne sont pas présents dans l'équation ( $U_{\text{tor}}, U_{\text{tord}}, U_b, U_{\text{rot}}$  et  $U_{\text{vid}}$ ).

Pour simuler l'effet de concentration présent lors d'une étude expérimentale comprenant plusieurs chaînes, un dernier potentiel, appelé potentiel de confinement  $U_{\text{conf}}$  est ajouté. Le système d'étude comprenant plusieurs sites est ainsi confiné dans une sphère. Si les sites du système sortent des frontières de la sphère,  $U_{\text{conf}}$  va produire une attraction les ramenant vers le centre de la sphère. Le but de ce confinement est de favoriser l'interaction entre les différentes chaînes à concentration finie et d'éviter qu'elles ne s'éloignent indéfiniment par diffusion.

Le potentiel global d'interaction à chaîne multiple regroupe donc les deux potentiels détaillés précédemment ainsi que le potentiel de confinement :

$$U_{\text{oligomere}} = \sum_k U_{\text{chaîne-unique}}^k + \sum_k \sum_{l > k} U_{\text{inter-chaînes}}^{k,l} + U_{\text{conf}}$$

Le champ de force UNRES a été appliqué avec succès pour simuler le repliement des protéines [80, 82, 83] et la dynamique conformationnelle de protéines chaperons [84].

Dans le cas de la protéine d'étude de cette thèse, l' $\alpha$ -syn, ce type de programme est très intéressant puisque cette protéine est capable de visiter un nombre conséquent de conformations dont certaines possèdent un rayon de giration très large. Avec un solvant explicite (comme pour un grand nombre de simulations tous atomes), les tailles de boîtes de simulation à utiliser seraient considérables et par conséquent elles ne permettraient que de faire de très courtes simulations et une étude des multimères extrêmement limitée.



## La méthode du Replica Exchange

Les protéines peuvent explorer plusieurs états. La protéine peut rester bloquée dans un état si la barrière énergétique entre un état et un autre état est trop grande. La probabilité de franchir cette barrière énergétique dépend de la température et à température ambiante est parfois proche de zéro. Par conséquent, une simulation de dynamique moléculaire à température constante ne permettra pas de définir tout l'espace conformationnel de la protéine sur le temps de simulation aux températures physiologiques (courbe bleue à basse température sur la [figure 2.2](#)).

Dans notre cas, l' $\alpha$ -syn est connue pour pouvoir réaliser un grand nombre de conformations différentes. Afin d'en visiter un maximum, il est nécessaire d'utiliser une méthode de simulation de dynamique moléculaire appelée "Replica Exchange". Plusieurs répliques à différentes températures, y compris les températures physiologiques, vont être simulées. Sur la [figure 2.2](#), la courbe orange et la courbe bleue représentent deux répliques à deux températures différentes : une basse température (courbe bleue) et une haute température (courbe orange). Les régions accessibles pour la protéine étudiée au cours de la simulation sont données par la coloration des différents puits. On remarque une barrière énergétique quasi infranchissable pour la réplique à basse température. A haute température cela devient possible car le profil des puits d'énergie est beaucoup plus plat. La méthode de Replica Exchange va permettre de réaliser des échanges entre les simulations à différentes températures afin d'aider la protéine à franchir ces barrières énergétiques. La méthode d'échange se base sur le critère de Metropolis [\[85\]](#). A intervalles réguliers, l'échange entre deux conformations pourra se faire suivant une probabilité d'acceptance :

$$P_{acc} = \begin{cases} 1 & \text{si } \Delta_{ij} \leq 0 \\ \exp(-\Delta_{ij}) & \text{si } \Delta_{ij} \geq 0 \end{cases}$$

avec

$$\Delta_{ij} = (\beta_j - \beta_i)[E(x_j) - E(x_i)]$$

L'élément  $\beta_k$  est rattaché à la température inverse de la réplique k,  $E$  à l'énergie potentielle de l'état considéré et  $x_k$  la configuration de la réplique k. La probabilité d'acceptance dépend donc de deux paramètres : l'énergie potentielle de la réplique et la température de celle-ci.

- Si  $T_j > T_i$  et  $E(x_j) \leq E(x_i)$  alors l'échange est accepté.
- Sinon la probabilité d'échange va décroître à mesure que la différence d'énergie potentielle augmente.

Ces équations montrent l'importance d'avoir des températures suffisamment proches pour que les distributions d'énergies potentielles se recoupent et puissent permettre de maximiser les chances d'échange et d'avoir ainsi une simulation optimale de l'espace conformationnel de la protéine.

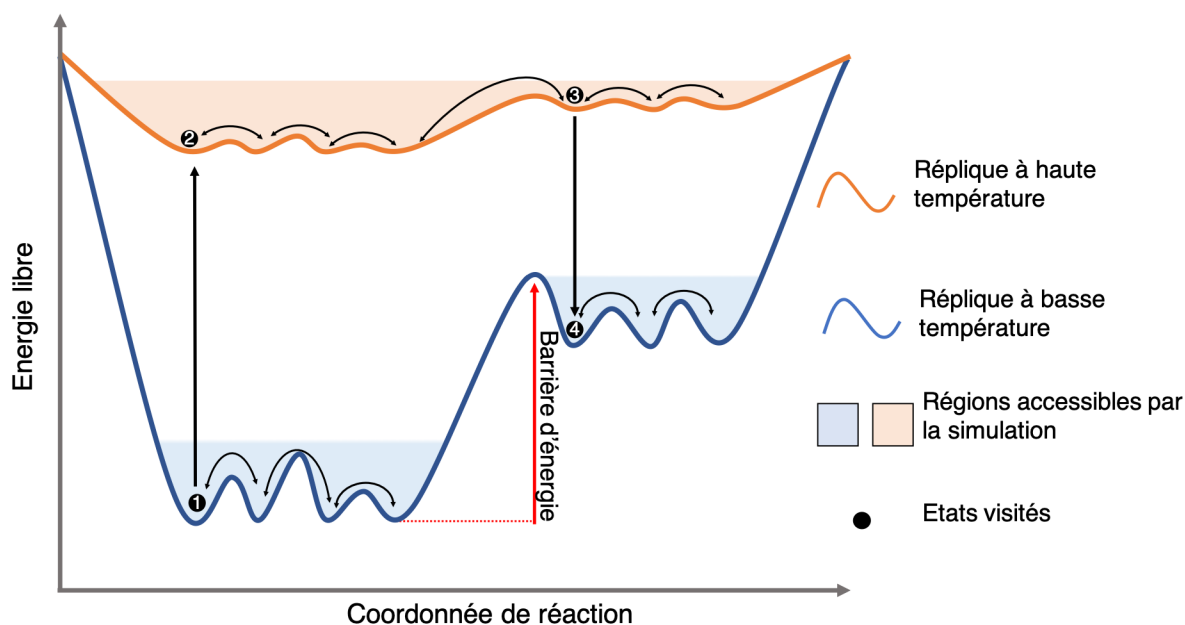


FIGURE 2.2 – Représentation schématique de la méthode du *Replica Exchange*. Les profils d'énergie libre le long d'une coordonnée de réaction sont représentés pour une réplique à basse température (courbe bleue) et une réplique à haute température (courbe orange). Les puits d'énergie libre colorés sont accessibles par la réplique correspondante et les flèches noires courbées montrent la possibilité de déplacement entre ces puits. Les flèches noires rectilignes donnent les possibilités d'échange idéaux entre les deux répliques en partant de l'état 1 suivant le critère de Métropolis.

La [figure 2.2](#) présente le cas de figure idéal. En partant de l'état 1, si les conditions de températures et d'énergies potentielles entre les deux répliques sont respectées, l'échange peut se faire et ainsi la conformation à l'état 1 va passer à haute température. Un second échange pourra être réalisé à l'état 3 et ainsi atteindre l'état 4, dans ce cas idéal la barrière d'énergie a été franchie et on visite donc de nouveaux états à la température la plus basse.

## Les données de simulations

Les trajectoires ont été calculées avec le programme UNRES (voir [section 2.1](#) page [19](#)) en utilisant la méthode du Replica Exchange (voir [section 2.1](#) page [22](#)) et le thermostat Berendsen pour 8 protéines :

- la protéine sauvage d' $\alpha$ -syn (WT).
- trois mutants retrouvés chez des patients atteints de la maladie de Parkinson : A30P, A53T et E46K .
- quatre mutants issus de la publication Dettmer et al. [\[58\]](#).

La [figure 2.3](#) résume les différents paramètres des simulations réalisées. La température la plus basse est de 300K pour 32 trajectoires, les températures suivantes sont données suivant un gradient exponentiel avec 310K, 323K, 337K, 353K, 370K pour 8 trajectoires chacune. Le point de départ des trajectoires est deux monomères d' $\alpha$ -syn séparés de 25 Å dont les angles sont contraints de façon à ce que les deux protéines aient une forme rectiligne.

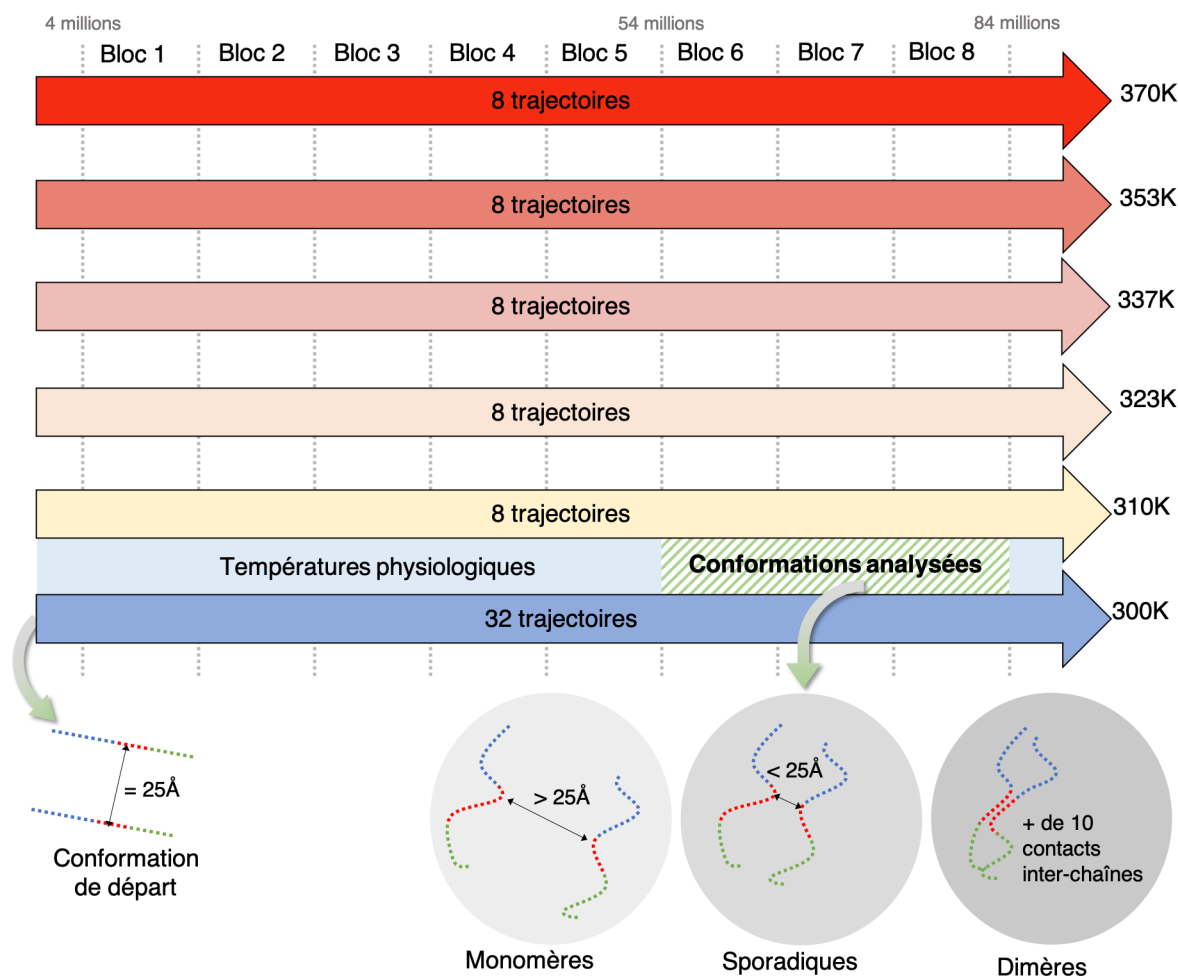


FIGURE 2.3 – Récapitulatif des différents paramètres utilisés pour la réalisation des trajectoires avec la méthode *Replica Exchange*. Les flèches de couleur représentent les différentes répliques à différentes températures. Les barres verticales en pointillés donnent le nombre de pas simulés dans ces trajectoires. Les schémas situés en dessous présentent la conformation de départ et les différentes classes de conformations définies pour  $\alpha$ -syn dans cette thèse au sein des trajectoires simulées.

L'ensemble des conformations d' $\alpha$ -syn a été divisé en trois classes distinctes : monomères, dimères et sporadiques. Une chaîne d' $\alpha$ -syn fait partie de la classe monomère si chacun de ses résidus est au moins à 25 Å des résidus de la seconde chaîne. La classe dimères concerne deux chaînes dont au moins 10 résidus sont en interactions (définies par la distance de contact) avec 10 résidus de l'autre chaîne. Enfin, la troisième classe, sporadique, contient toutes les autres conformations possibles.

Le pas d'intégration d'UNRES est de 4.9 fs correspondant à temps effectif de 4.9 ps [86]. Le temps effectif global est de 412 microsecondes par trajectoire. Les conformations des 4 premiers millions de pas sont considérées comme trop proches de la conformation de départ et n'ont donc pas été utilisées. Dans la continuité de ces 4 premiers millions, nous avons réalisé 8 blocs de 10 millions de pas chacun. A partir des classes "monomères" et "dimères" nous avons vérifié la convergence à l'aide du rayon de giration pour les monomères du WT (figure 2.4 panneau de droite) et du contact moyen inter-chaines pour les dimères (figure 2.4 panneau de gauche)

(voir [section 2.3](#) page [33](#)) . Les courbes des blocs 6, 7 et 8 sont très proches pour les deux propriétés. Nous avons donc considéré que la convergence est atteinte pour ces 3 blocs et nous n'avons donc analysé que cette partie de notre statistique correspondant à un temps effectif de 147 microsecondes par trajectoire. On utilisera les conformations à des températures comprises entre 300K et 310K proches de la température physiologique dans les 72 trajectoires (zone hachurée verte à la [figure 2.3](#)). Dans cet ensemble, on retrouve au total environ 1,08 millions de conformations soit environ 5,3 millisecondes de simulations (temps effectif) sur un ensemble global (en prenant en compte les 84 millions de pas réalisés et les 72 trajectoires réalisées) de 29.7 millisecondes de temps effectif..

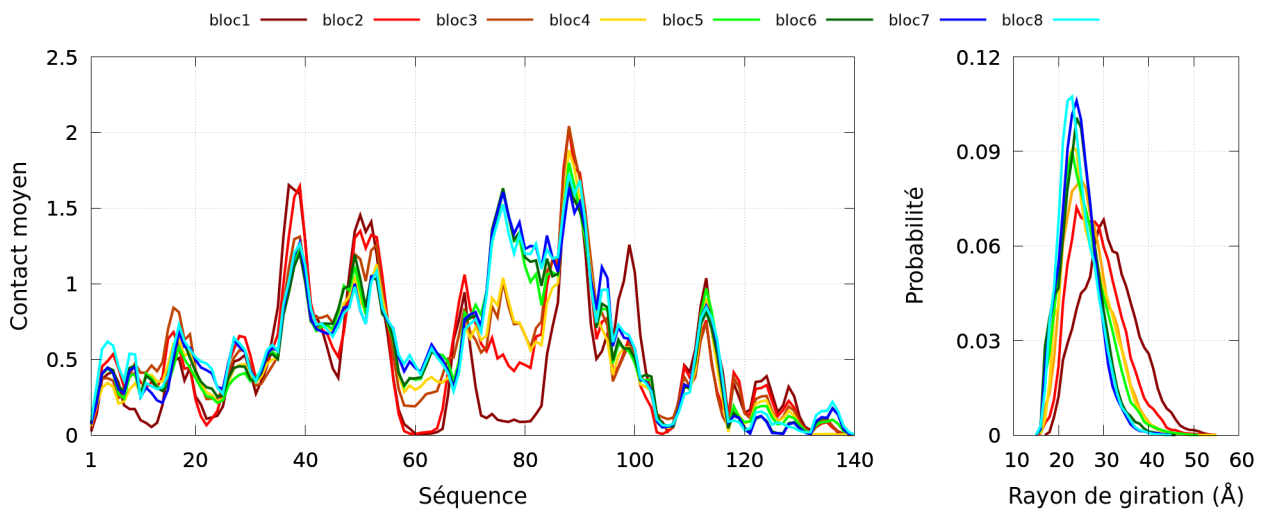


FIGURE 2.4 – Contact moyen inter-chaînes des dimères (panneau de gauche) et distribution de probabilité du rayon de giration des monomères (panneau de droite). Les deux figures sont réalisées pour le WT et présentent les courbes de chaque bloc ([figure 2.3](#)) suivant le gradient de couleur donné dans la légende au dessus de la figure.

## 2.2 Détections des structures secondaires

Les protéines sont caractérisées par une succession d'acides aminés. Cet enchainement différent d'une protéine à l'autre lui donne ses caractéristiques en formant différents types de structures secondaires telles que les hélices, les feuillets  $\beta$  ou les boucles. Au sein des hélices, il existe des sous familles : les hélices  $3_{10}$ , les hélices- $\alpha$  (majoritairement représentées dans les protéines) et hélices  $\pi$ . La formation d'un de ces types d'hélices dépend de la liaison hydrogène formée entre les groupements N-H de l' $AA_i$  et C=O de l' $AA_{i+3}$  pour les hélices  $3_{10}$ ,  $AA_{i+4}$  (pour les hélices- $\alpha$ ) ou  $AA_{i+5}$  pour les hélices  $\pi$ . Concernant les feuillets  $\beta$ , il en existe deux types : parallèles si les deux brins sont orientés de la même façon ou anti-parallèles si leurs orientations s'opposent. La boucle est une structure bien plus variable, elle est surtout remarquable lorsqu'elle est couplée avec un feuillet pour former une hairpin.

## Détection des hélices et des feuillets $\beta$

En utilisant un modèle à gros grains tel que UNRES, les fichiers de sortie donnant les coordonnées cartésiennes de la protéine ne contiennent que deux entités par AA :  $C^\alpha$  et  $C^\beta$ . Sans la position de la totalité des atomes de la chaîne principale, il est impossible d'utiliser les outils conventionnels tels que DSSP [87][88] ou STRIDE [89] pour assigner à chaque AA une structure secondaire. L'unique option pour utiliser ces programmes est de reconstruire la structure en tous atomes à partir de la structure en gros grains en utilisant par exemple MMTSB [90] ou PULCHRA [91]. Cependant le coût en terme de temps de calcul est considérable. Nous avons donc construit un programme nommé CUTABI (CURvature and Torsion based on Alpha-Helix and Beta-sheet Identification). En utilisant uniquement les positions des  $C^\alpha$ , le programme est capable de détecter les hélices et les feuillets  $\beta$  par le biais des angles de liaison  $\theta$  et diédraux  $\gamma$ . Dans le cas des feuillets- $\beta$ , une combinaison d'angles  $\theta$  et  $\gamma$  (définis [figure 2.5](#)) associée à des distances entre les  $C^\alpha$  est nécessaire.

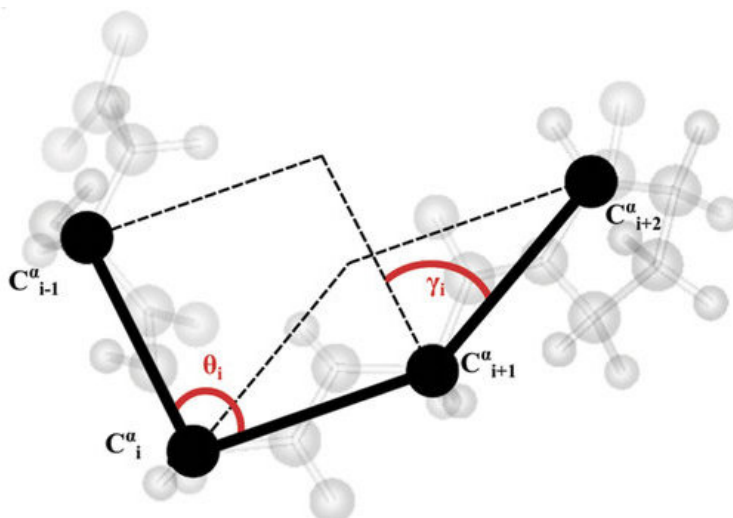


FIGURE 2.5 – Pour un résidu  $i$ ,  $\gamma_i$  est l'angle diédral formé par la liaison virtuelle reliant quatre  $C^\alpha$  successifs ( $i-1$ ,  $i$ ,  $i+1$  et  $i+2$ ) le long de la séquence d'AA et  $\theta_i$  est l'angle de liaison formé par la liaison virtuelle reliant trois  $C^\alpha$  successifs ( $i-1$ ,  $i$ ,  $i+1$ ) le long de la séquence d'Å.

Le programme a été conçu à partir de 14 652 structures expérimentales contenant moins de 40% d'identité de séquence extraites de la base de donnée ASTRAL [92][93]. Les critères de CUTABI ont été ajustés pour reproduire les résultats de DSSP. Nous sommes parvenu à obtenir un accord de 93% avec DSSP pour les 14 652 structures en étant 10 fois plus rapide. Les performances détaillées de CUTABI comparées à DSSP et à un autre programme de détection à partir de gros-grains appelé PSEA [94] sont décrites dans une publication de notre équipe donnée dans la partie [chapitre 3 \[1\]](#).

CUTABI ne peut détecter que des hélices constituées de minimum 4 AA. Les hélices  $3_{10}$  pouvant être de 3 AA ne peuvent donc pas être reconnues par le programme. Les hélices ne se forment que pour des interactions à courte distance avec la chaîne c'est pourquoi une succession simple d'angles est suffisante pour définir une hélice. En effet si les angles  $\theta_k$ ,  $\theta_{k+1}$ ,  $\theta_{k+2}$  et  $\theta_{k+3}$  sont compris entre  $80^\circ$  et  $105^\circ$  et que les angles  $\gamma_{k+1}$  et  $\gamma_{k+2}$  sont compris entre  $30^\circ$  et  $80^\circ$  alors

les résidus  $k$ ,  $k+1$ ,  $k+2$  et  $k+3$  sont considérés comme appartenant à une hélice ( [figure 2.6](#) panneau de gauche).

CUTABI détecte des feuilletts  $\beta$  dont les brins font minimum 2 AA formant donc un feuillet de minimum 4 AA. Les ponts  $\beta$  n'incluant que 1 seul AA ne peuvent donc pas être considérés dans notre définition. Pour que les AA  $k$ ,  $k+1$ ,  $i$  et  $i+1$  appartiennent à un feuillet  $\beta$  il faut que  $\theta_k$ ,  $\theta_{k+1}$ ,  $\theta_i$  et  $\theta_{i+1}$  soient compris entre  $100^\circ$  et  $155^\circ$  et que  $\gamma_k$  et  $\gamma_i$  soient supérieur à  $80^\circ$  ou inférieur à  $-80^\circ$ . Des conditions de distance de moins de  $5.5 \text{ \AA}$  entre  $k$  et  $i$  et entre  $k+1$  et  $i+1$  ainsi qu'une distance de moins de  $6.8 \text{ \AA}$  entre  $k+1$  et  $i+2$  (flèches rouges et bleues respectivement sur la [figure 2.6](#) panneau de droite) s'ajoutent aux conditions angulaires pour détecter le feuillet. Des conditions similaires s'appliquent pour détecter un feuillet  $\beta$  anti-parallèle.

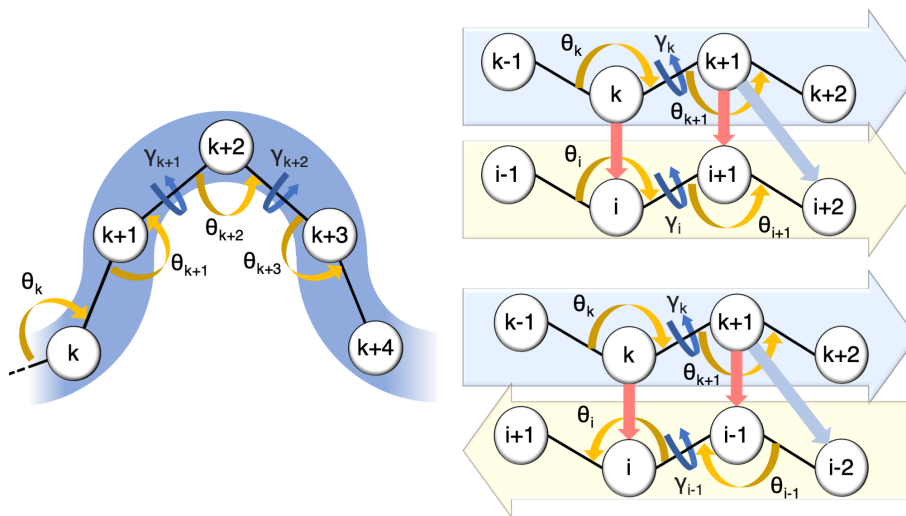


FIGURE 2.6 – Paramètres définissant les résidus comme appartenant à une hélice ou à un feuillet  $\beta$ . Les sphères blanches définissent la position des  $C^\alpha$  des AA. Les flèches oranges et bleues montrent respectivement les angles  $\theta$  et  $\gamma$  impliqués. Pour les feuilletts  $\beta$ , les flèches rouges et bleues donnent les distances entre les AA nécessaires dans la définition.

## Détection de structures particulières

Les feuilletts  $\beta$  sont capables de former des structures particulières appelées "hairpins" en anglais pour "épingles à cheveux". L'apparition d'une hairpin est une caractéristique structurale intéressante chez les protéines. Elle a notamment été détectée par simulations chez  $\alpha$ -syn sur le segment 38-53 de la séquence de la protéine [\[68\]](#). Il s'est donc avéré utile de pouvoir classer les conformations contenant ces structures dans notre ensemble de données afin de comparer avec les résultats obtenus dans d'autres types de simulation de dynamique moléculaire. Au cours de notre analyse, nous nous sommes aperçus qu'un second type de structure proche de la hairpin pouvait intervenir chez  $\alpha$ -syn. Cette structure s'apparente à deux brins d'une hairpin qui se sont entremêlés en faisant une rotation sur l'axe de la hairpin. Elle a été nommée dans cette thèse "tresse". Une représentation schématique de cette structure est donnée [figure 2.7](#).

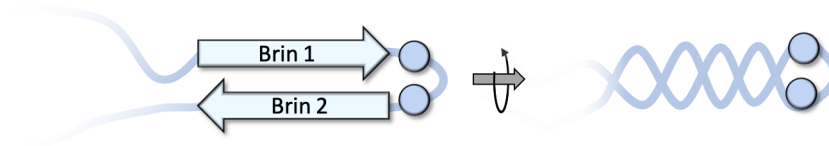


FIGURE 2.7 – Représentation schématique d'une hairpin (à gauche) et d'une tresse (à droite)

### - Détection des hairpins

Pour détecter les hairpins en utilisant CUTABI, il nous faut passer par la détection des boucles au sein d'une hairpin. La "boucle" décrite par DSSP est complexe à détecter par CUTABI si elle ne se situe pas au sein d'une hairpin. Dans le cas où elle se situe au sein d'une hairpin elle peut être composée de 2 à 5 résidus. Les travaux publiés sur la hairpin d' $\alpha$ -syn dans le segment 38-53 définissent la boucle de cette hairpin comme étant composée des résidus 45 et 46 [68]. Pour comparer nos résultats directement avec ceux déjà publiés nous avons donc étendu CUTABI à la détection des hairpins contenant une boucle de 2 AA.

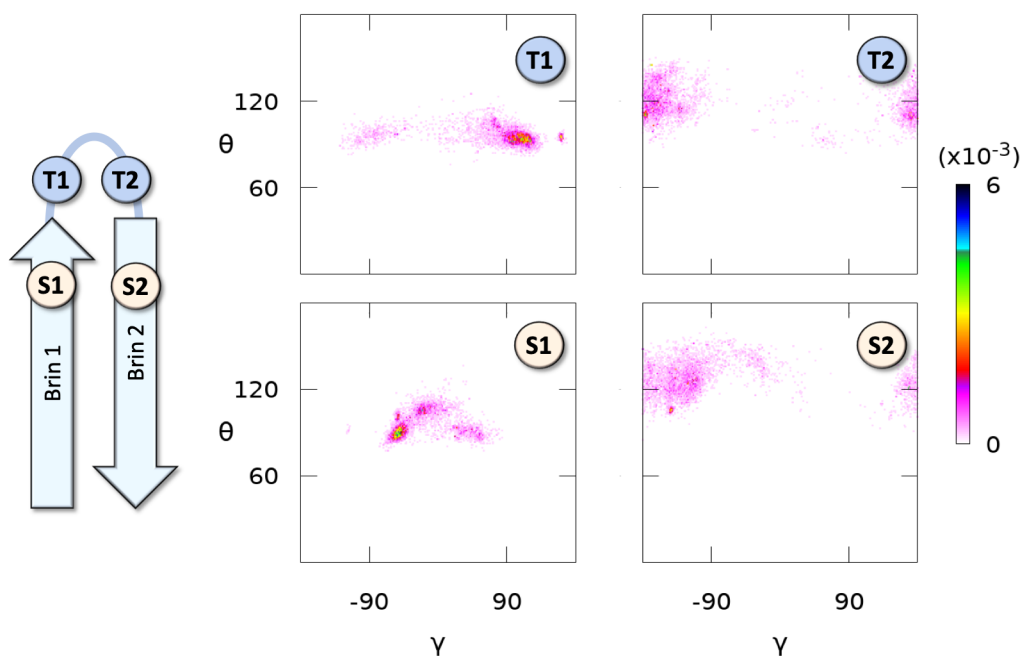


FIGURE 2.8 – Paysages d'énergie libre des couples d'angle  $\theta$  et  $\gamma$  retrouvés pour les résidus situés dans une hairpin d'après DSSP dans les structures de la base ASTRAL contenant au moins une hairpin. Les cartes sont associées à chaque résidu dont la position est donnée dans le schéma à gauche des cartes. Le gradient de couleur indique la probabilité d'apparition d'un couple.

Le set de données utilisé pour paramétrer CUTABI pour la détection d'hairpin est constitué des structures ASTRAL déjà utilisées pour la détection des hélices et des feuillet  $\beta$ . Nous avons cherché dans ces structures expérimentales toutes les régions pour lesquelles DSSP détecte un enchaînement de 2 AA en feuillets  $\beta$  puis 2 AA dans une boucle puis de nouveaux 2 AA en

feuillet  $\beta$  soit BBTTBB (selon les codes DSSP : B=Beta-sheet (feuillet  $\beta$ ), T=Turn (boucle)). Si la protéine possède au moins une structure de la sorte, elle sera utilisée pour l'analyse. On dénombre ainsi un total de 4735 structures possédant ce critère au moins une fois. A partir de ces structures on va chercher à paramétrer CUTABI de sorte à ce qu'il retrouve les mêmes enchainements BBTTBB que DSSP. Les paysages d'énergie libre dans l'espace des angles  $(\theta, \gamma)$  des AA en feuillets  $\beta$  adjacents à la boucle, notés S1 et S2 et des AA dans la boucle, notés T1 et T2, sont donnés à la [figure 2.8](#). Seuls les AA S1 et T1 possèdent un couple d'angles assez spécifique contrairement aux AA T2 et S2 qui forment des angles propices à une conformation de feuillets  $\beta$ . Le paysage d'énergie libre de S1 montre trois tâches dont une plus probable formées pour des angles  $\gamma$  négatifs. Après différents tests, il s'avère que la valeur de l'angle  $\gamma$  est assez flexible au sein d'une boucle, nous avons donc choisi d'utiliser le spectre complet allant de  $-64^\circ$  à  $64^\circ$  pour  $\gamma$  et  $80^\circ$  à  $120^\circ$  pour l'angle  $\theta$  recouvrant ainsi la quasi totalité des points. L'AA T1 a un angle  $\theta$  entre  $80^\circ$  et  $110^\circ$  et un angle  $\gamma$  pour une grande partie positif. On remarque ainsi trois zones, tout d'abord une première très probable et assez large au delà de  $90^\circ$  qui s'étend en réduisant sa probabilité pour des angles  $\gamma$  inférieurs à  $90^\circ$ . Symétriquement à cette tâche on retrouve une autre tâche pour des  $\gamma$  négatifs mais stastiquement bien moins probable. Une troisième tâche très localisée apparaît ensuite proche d'un  $\gamma$  de  $180^\circ$ . On utilisera les  $\gamma$  entre  $50^\circ$  et  $180^\circ$  pour englober les deux tâches majeures et on laissera de côté la tâche des  $\gamma$  négatifs car sa prise en compte pourrait avoir tendance à nous faire surestimer la détection des boucles. Les gammes d'angles utilisées sont assez larges et ces simples conditions ne sont pas suffisantes pour estimer la position d'une hairpin dans une séquence. Nous avons donc ajouté, en complément, la vérification selon les critères de CUTABI que les AA S1-1 et S2+1 sont bien localisés dans un feuillet  $\beta$ .

Dans la suite de ce manuscrit, lorsque le terme hairpin sera employé il fera allusion à une hairpin à deux résidus.

### - Détection des tresses

La tresse est une structure encore jamais étudiée à notre connaissance dans la littérature. Elle se caractérise par des angles  $\theta/\gamma$  atypiques. L'angle  $\theta$  est très localisé avec une gamme comprise entre  $110^\circ$  et  $120^\circ$ . L'angle  $\gamma$  occupe une région déjà connue comme étant une région propice à la formation d'hélice entre  $30^\circ$  et  $60^\circ$ . Le couple de ces deux angles se situe donc grâce à l'angle  $\theta$  dans une région dissociée de celle des hélices et permet donc une détection facile de ce type de structure. La tresse se constitue de deux branches anti-parallèles comme un feuillet  $\beta$ , la particularité est que ces deux branches vont s'entremêler pour former un tressage.

Considérer qu'un résidu est dans une tresse car ses angles se trouvent dans la gamme définie n'est pas une condition suffisante. Dans le cas de l'analyse du segment 38-53, une étude complémentaire (non présentée dans cette thèse) a permis d'établir un phénomène de corrélation entre les angles des AA dans une tresse. Dans le segment de séquence 38-53, les résidus 40,41,48 et 49 se trouvent au centre des brins de la tresse et forme la boucle centrale ([figure 2.9](#)). Si au moins 3 de ces 4 résidus possèdent des angles dans la gamme d'angles de la tresse, la portion



de séquence 38-53 aura une très grande probabilité de former une tresse. On utilisera ce critère pour la détecter.

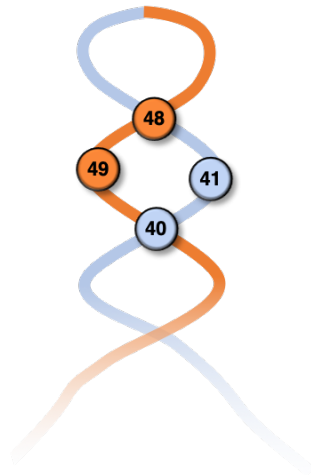


FIGURE 2.9 – Schéma représentant une tresse du résidu 38 au 53 dans l' $\alpha$ -syn . La partie montante (partant du Nterm) est donnée en dégradé bleu, la partie descendante (allant vers le Cterm) est donnée en dégradé orange. Les positions des  $C^\alpha$  40, 41, 48 et 49 utilisés pour détecter la tresse sont données par des sphères.

## 2.3 Analyses des simulations

### Densité de probabilité des couples hélices/feuillet $\beta$

L'utilisation de CUTABI nous permet avec un temps de calcul assez court de calculer la probabilité de formation d'hélices et de feuillets  $\beta$  chez  $\alpha$ -syn pour toutes nos conformations. On peut ainsi étudier de manière assez simple la probabilité de formation des structures secondaires par AA.

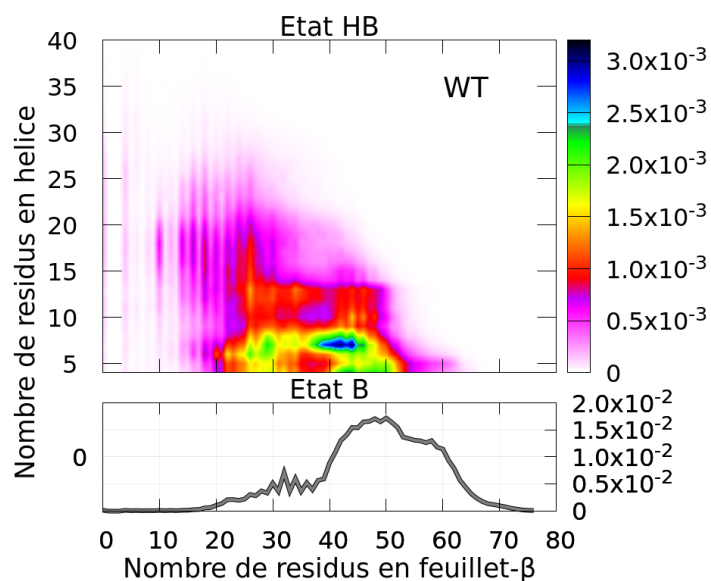


FIGURE 2.10 – Exemple de carte de couleur typique représentant le nombre de résidus en hélice par rapport au nombre de résidus en feuillet  $\beta$  pour les conformations des monomères du WT.

Pour réaliser des classes des conformations des monomères, la détection des feuillets  $\beta$  et des hélices s'est avérée être un outil très utile. En effet, grâce à CUTABI, il nous est possible de compter le nombre d'AA formant des hélices et le nombre d'AA formant des feuillets  $\beta$  pour chaque conformation. Grâce à ce couple de valeurs, on peut créer une carte en deux dimensions proposant en abscisse le nombre d'AA en feuillets  $\beta$  et en ordonnée le nombre d'AA en hélices (figure 2.10). Chaque conformation formera un point sur cette carte situé au couple de valeurs correspondant. Pour une étude statistique de l'ensemble des conformations, une troisième dimension correspondant à la couleur de ce point dépendra de la proportion de conformations capables de former ce couple de valeurs. Par soucis de visibilité dans le cas des conformations ne comportant pas d'AA en hélice (état B figure 2.10), nous avons réalisé une courbe à une dimension avec une échelle de probabilité plus adaptée donnée sur l'axe y de droite.

## Analyses rattachées à la notion de contact

### - Définition

L'analyse d'une protéine aussi complexe que l' $\alpha$ -syn nécessite de trouver, au sein de la très grande quantité de données due à la statistique très importante, des paramètres permettant de faire ressortir des informations pertinentes sur la protéine.

Un de ces paramètres est très simple : le contact. Si deux résidus se trouvent à une distance suffisamment proche, mesurée entre les positions de leur  $C^\alpha$ , ils seront considérés en contact.

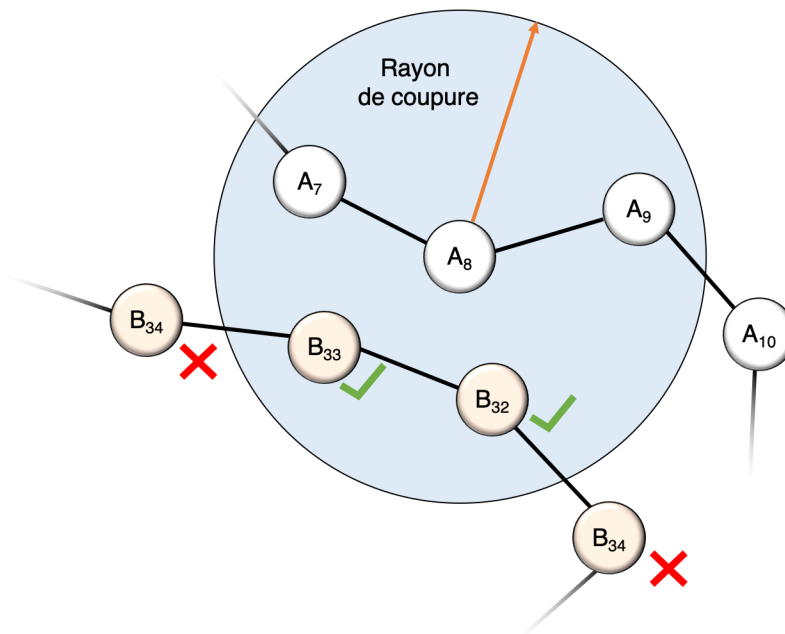


FIGURE 2.11 – Schéma explicatif de la notion de contact (ici dans le cas de contact inter-chaines). Les sphères blanches représentent les  $C^\alpha$  de la chaîne A, les sphères orangées représentent les  $C^\alpha$  de la chaîne B. Le cercle bleu est le rayon de coupure appliqué au  $C^\alpha$  analysé  $A_8$ . Les  $C^\alpha$  présents dans cette sphère,  $B_{32}$  et  $B_{33}$ , sont considérés en contact avec  $A_8$ .

On utilise deux façons de compter les contacts dépendant du rayon de coupure utilisé :

- Contact limité : on limite le nombre de contacts pour chaque  $C^\alpha$  à 1 en prenant en compte le  $C^\alpha$  le plus proche. Cette méthode est utilisée pour définir les différentes classes "monomères", "dimères" et "sporadiques" explicitées page [23](#).

- Contact global : si on utilise un rayon de coupure assez large, il est possible que plusieurs  $C^\alpha$  se retrouvent dans la sphère de contact du  $C^\alpha$  analysé, lui donnant ainsi la possibilité de faire plusieurs contacts. Ce cas de figure est illustré à la [figure 2.11](#). Le  $C^\alpha$  analysé est le  $A_8$ . Il possède 2  $C^\alpha$  de la chaîne B dans son rayon de coupure, soit  $B_{32}$  et  $B_{33}$ . Il forme donc deux contacts. On utilise cette méthode pour la majorité des analyses comme le calcul du contact moyen inter-chaînes le long de la séquence pour les dimères et les contacts intra-chaîne en ignorant les premiers et seconds voisins liés par des liaisons virtuelles.

Deux distances de coupure ont été utilisées : 5 Å et 6 Å. La distance 5 Å a été utilisée pour la séparation des différentes conformations entre les classes "monomères", "dimères" et "sporadiques". Cette distance englobe toutes les distances inter-chaînes retrouvées entre les résidus dans les structures de fibrilles expérimentales de codes PDB 2n0a (environ 4.9 Å) et 6ssx (environ 4.7 Å). L'étude des feuillets  $\beta$  intra-chaîne de la base ASTRAL qui a permis de construire le programme CUTABI a montré une distance inférieure à 5.5 Å entre les résidus des deux brins formant un feuillet  $\beta$ . Pour l'étude de l'agrégation de la protéine nous avons donc utilisé dans la plupart de nos analyses la distance de coupure de 6 Å. Ce rayon de coupure permet également de rendre plus lisible les cartes de contacts (voir paragraphe suivant).

### - Carte de contact

Les axes des ordonnées et des abscisses des cartes de contacts (CM) représentent la séquence de la protéine. Si un  $C^\alpha$  forme un contact avec un autre  $C^\alpha$  un point se formera sur la carte. Lorsqu'on analyse plusieurs conformations en même temps, une échelle de couleur vient colorer le point en fonction de sa probabilité d'apparition. Une interpolation des points est réalisée pour chaque carte de sorte à éviter la formation discrète de points rendant la carte de contact moins lisible.

Dans le cas des cartes de contact intra-chaîne (pour les monomères), pour un AA de position  $i$  les plus proches voisins  $i-2$ ,  $i-1$ ,  $i+1$  et  $i+2$  ne sont pas pris en compte car trop proches. La diagonale de la carte ne possède donc pas de points, cela permet une meilleure lecture des contacts proches de la diagonale au-delà des seconds voisins le long de la chaîne. Dans le cas des cartes de contact inter-chaînes (pour les dimères) tous les  $C^\alpha$  sont inclus.

On remarque plusieurs types de traces sur ces cartes :

- Les traces rectilignes perpendiculaires à la diagonale correspondent à des feuillets  $\beta$  anti-parallèles : pour une hairpin la trace coupe la diagonale de manière notée (1a) sur le panneau A sur la [figure 2.12](#) sinon il s'agira d'un feuillet  $\beta$  anti-parallèle à longue distance noté (1b) ou

d'un simple feuillet  $\beta$  anti-parallèle entre les deux chaînes noté (1b) sur le panneau (B).

- Les traces parallèles à la diagonale notées (2) sur la [figure 2.12](#) correspondent à des feuillets  $\beta$  parallèles.

Cas particuliers :

- Pour les cartes de contact intra-chaîne (panneau (A) [figure 2.12](#)) : les traces rectilignes collées à la diagonale notées (3) correspondent à la formation d'hélice.

- Pour les cartes de contact inter-chaînes (panneau (B) [figure 2.12](#)) : les traces rectilignes se superposant à la diagonale sont assimilables à des contacts de type fibrille (CTF), c'est à dire reliant les mêmes AA de chaque chaîne, notées (4).

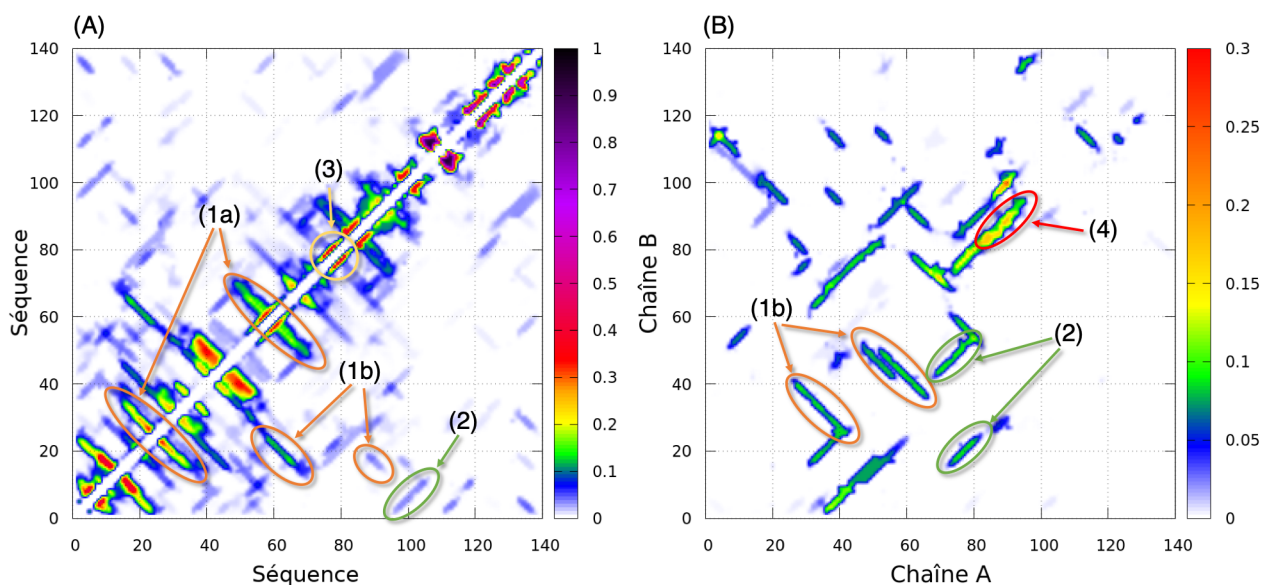


FIGURE 2.12 – Exemples de carte de contacts intra-chaîne (panneau (A)) et inter-chaînes (panneau (B)) de la protéine sauvage  $\alpha$ -syn . Les différents types de structures secondaires sont numérotés et décrites dans le texte.

### - Contact moyen le long de la séquence

Le calcul du contact moyen est utilisé pour caractériser les interactions entre deux monomères au sein de nos simulations. Le rayon de coupure utilisé est de 6 Å. Dans ce cas de figure, lors de l'analyse de chaque AA on utilise la notion de contact global. Ce nombre de contact moyen est la somme des contacts de l'acide aminé dans les deux chaînes (Exemple : Contact  $AA_8 = \text{Contacts } AA_8 \text{ chaîne A} + \text{Contacts } AA_8 \text{ chaîne B}$ ) dans l'ensemble des conformations analysées, divisée par le nombre de conformations analysées et par 2 (pour le nombre de chaînes). La courbe de la [figure 2.13](#) présente la séquence en abscisse avec la valeur de contact moyen en ordonnée pour chaque résidu. L'intérêt de cette analyse est d'avoir une image de la capacité d'un AA à attirer l'autre chaîne pour former des contacts qui sont potentiellement des contacts dans des feuillets  $\beta$ .

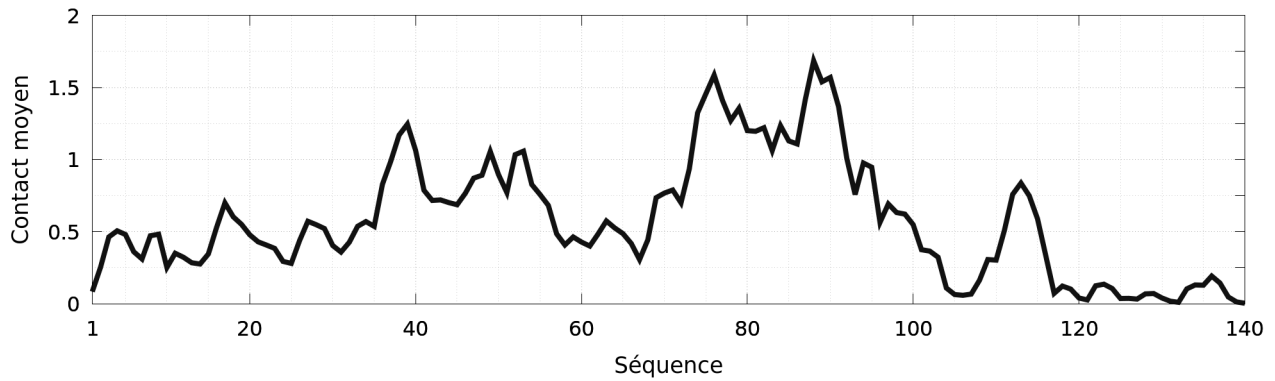


FIGURE 2.13 – Exemple typique de la courbe de contact moyen le long de la séquence pour les dimères de l' $\alpha$ -syn WT.

### - Carte de couleur du nombre de contacts intra-chaine et du nombre de contact inter-chaines

La formation d'interactions entre les deux monomères est également caractérisée en calculant le nombre de contacts intra-chaine ( $n_{intra}$ ) et inter-chaines ( $n_{inter}$ ). L'utilisation de ce couple de valeurs pour un dimère permet de réaliser une carte en deux dimensions dont l'abscisse est le nombre de contacts inter-chaines et l'ordonnée le nombre de contacts intra-chaine. Un exemple de ce type de carte est donné (figure 2.14). Les contacts sont comptabilisés en sommant tous les contacts réalisés par chaque AA suivant la notion de contact globale. Chaque point de la carte représente un couple de valeur  $n_{inter}/n_{intra}$ . La couleur de ces points est définie par l'échelle de couleur donnée en  $-\ln\left(\frac{P(n_{inter},n_{intra})}{P_{max}}\right)$  calculées à partir de la densité de probabilité  $P$  du couple  $n_{inter}/n_{intra}$  entre les résidus d'un ensemble de dimères et  $P_{max}$ , la valeur maximale de la probabilité de la carte.

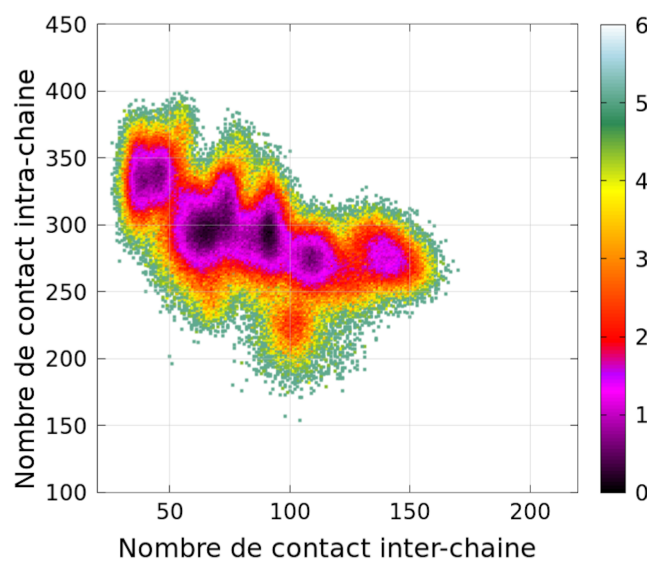


FIGURE 2.14 – Exemple typique de carte de couleur du nombre de contacts intra-chaine en fonction du nombre de contact inter-chaines pour les dimères du WT.

# Chapitre 3

## Résultats : Etude globale des monomères d' $\alpha$ -synucléine

La première publication de cette thèse "Missense mutations modify the conformational ensemble of the  $\alpha$ -synuclein monomer which exhibits a two-phase characteristic" publiée le 29 novembre 2021 par notre équipe présente l'étude des monomères d' $\alpha$ -syn du WT et des mutants A30P, A53T et E46K simulés par le programme de dynamique moléculaire UNRES.

La partie Méthode de la publication explicite le développement du programme CUTABI (C $\alpha$ Urvature and Torsion base of Alpha-helix and Beta-sheet Identification), que nous avons développé, permettant la détection des hélices et des feuillets  $\beta$  en utilisant uniquement une combinaison d'angles et de distances des  $C^\alpha$  de la protéine. Basé sur un ensemble de 14 652 structures expérimentales, les paramètres de CUTABI ont été ajustés sur les résultats de l'algorithme tous atomes DDSP avec un accord de 93%.

L'analyse des monomères à l'aide de CUTABI nous a permis de réaliser une carte à deux dimensions de la densité de probabilité du nombre de résidus en hélices en fonction du nombre de résidus en feuillets  $\beta$  montrant deux groupes principaux de conformations ou états. Un premier état a été identifié où les monomères ne sont constitués que de feuillets  $\beta$ , représentant environ 40% des conformations du WT, A30P et E46K et seulement 25% des conformations de A53T. Le second état contient des monomères constitués d'un mélange d'hélices et de feuillets  $\beta$ . Nous avons mis en avant le fait que les caractéristiques globales de la protéine, telle que la distribution de probabilité du rayon de giration, est insuffisante pour caractériser les différences entre le WT et les mutants mais que l'analyse des hélices et des feuillets  $\beta$  permet d'établir des comportements variants entre les différentes protéines.



# Missense Mutations Modify the Conformational Ensemble of the $\alpha$ -Synuclein Monomer Which Exhibits a Two-Phase Characteristic

Adrien Guzzo<sup>1</sup>, Patrice Delarue<sup>1</sup>, Ana Rojas<sup>2</sup>, Adrien Nicolai<sup>1</sup>, Gia G. Maisuradze<sup>3</sup> and Patrick Senet<sup>1,3\*</sup>

<sup>1</sup>Laboratoire Interdisciplinaire Carnot de Bourgogne, UMR 6303 CNRS-Université de Bourgogne Franche-Comté, Dijon, France, <sup>2</sup>Schrödinger, Inc., New York, NY, United States, <sup>3</sup>Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY, United States

## OPEN ACCESS

### Edited by:

Piero Andrea Temussi,  
University of Naples Federico II, Italy

### Reviewed by:

Leonid Breydo,  
West Pharmaceutical Services,  
United States  
Fabio Sterpone,  
UPR9080 Laboratoire de Biochimie  
Théorique (LBT), France

### \*Correspondence:

Patrick Senet  
psenet@u-bourgogne.fr

### Specialty section:

This article was submitted to  
Structural Biology,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 29 September 2021

**Accepted:** 25 October 2021

**Published:** 29 November 2021

### Citation:

Guzzo A, Delarue P, Rojas A, Nicolai A,  
Maisuradze GG and Senet P (2021)  
Missense Mutations Modify the  
Conformational Ensemble of the  
 $\alpha$ -Synuclein Monomer Which Exhibits a  
Two-Phase Characteristic.  
Front. Mol. Biosci. 8:786123.  
doi: 10.3389/fmolb.2021.786123

$\alpha$ -Synuclein is an intrinsically disordered protein occurring in different conformations and prone to aggregate in  $\beta$ -sheet structures, which are the hallmark of the Parkinson disease. Missense mutations are associated with familial forms of this neuropathy. How these single amino-acid substitutions modify the conformations of wild-type  $\alpha$ -synuclein is unclear. Here, using coarse-grained molecular dynamics simulations, we sampled the conformational space of the wild type and mutants (A30P, A53P, and E46K) of  $\alpha$ -synuclein monomers for an effective time scale of 29.7 ms. To characterize the structures, we developed an algorithm, CUTABI (CUrvature and Torsion based of Alpha-helix and Beta-sheet Identification), to identify residues in the  $\alpha$ -helix and  $\beta$ -sheet from  $C^\alpha$ -coordinates. CUTABI was built from the results of the analysis of 14,652 selected protein structures using the Dictionary of Secondary Structure of Proteins (DSSP) algorithm. DSSP results are reproduced with 93% of success for 10 times lower computational cost. A two-dimensional probability density map of  $\alpha$ -synuclein as a function of the number of residues in the  $\alpha$ -helix and  $\beta$ -sheet is computed for wild-type and mutated proteins from molecular dynamics trajectories. The density of conformational states reveals a two-phase characteristic with a homogeneous phase (state B,  $\beta$ -sheets) and a heterogeneous phase (state HB, mixture of  $\alpha$ -helices and  $\beta$ -sheets). The B state represents 40% of the conformations for the wild-type, A30P, and E46K and only 25% for A53T. The density of conformational states of the B state for A53T and A30P mutants differs from the wild-type one. In addition, the mutant A53T has a larger propensity to form helices than the others. These findings indicate that the equilibrium between the different conformations of the  $\alpha$ -synuclein monomer is modified by the missense mutations in a subtle way. The  $\alpha$ -helix and  $\beta$ -sheet contents are promising order parameters for intrinsically disordered proteins, whereas other structural properties such as average gyration radius,  $R_g$ , or probability distribution of  $R_g$  cannot discriminate significantly the conformational ensembles of the wild type and mutants. When separated in states B and HB, the distributions of  $R_g$  are more significantly different, indicating that global structural parameters alone are insufficient to characterize the conformational ensembles of the  $\alpha$ -synuclein monomer.

**Keywords:**  $\alpha$ -synuclein, amyloid, Parkinson disease, molecular dynamics, dictionary of secondary structure of proteins, CUrvature and Torsion based of Alpha-helix and Beta-sheet Identification, PSEA

## 1 INTRODUCTION

$\alpha$ -helices and  $\beta$ -sheets are the two main secondary structures stabilized by backbone hydrogen bonds in proteins. The propensity of each residue to form an  $\alpha$ -helix or a  $\beta$ -sheet depends on the amino-acid sequence and on the fold (Chou and Fasman, 1974; Smith et al., 1994; Pace and Scholtz, 1998; Bhattacharjee and Biswas, 2010). The formation of long  $\beta$ -sheets differs significantly from  $\alpha$ -helices as the former necessarily involves long-distance contacts in the sequence and depends on the mean curvature of the fold (Koh et al., 2006; Bhattacharjee and Biswas, 2010). The content of these two secondary structure elements is a criterion for classification of protein native structures and characterization of protein folding kinetics and pathways (Konagurthu et al., 2020). In intrinsically disordered proteins (IDPs) (Wright and Dyson, 1999; Uversky et al., 2000; Uversky, 2019; Deiana et al., 2019),  $\alpha$ -helices and  $\beta$ -sheets are metastable secondary structures. As for folded proteins, a natural extension of protein classification of IDP can be based on the content of these two secondary structural elements in a statistical sense. As an IDP has no native state, the relative content of  $\alpha$ -helices ( $\alpha$ ) and of  $\beta$ -sheets ( $\beta$ ) possibly defines a fingerprint of each conformational state in a two-dimensional ( $\alpha$ ,  $\beta$ ) propensity map (Ullman et al., 2011). Here, we aim to build such an ( $\alpha$ ,  $\beta$ ) effective free-energy map for  $\alpha$ -synuclein ( $\alpha$ -syn), wild-type (WT), and A30P, A53T, and E46K mutants from first principles by using coarse-grained molecular dynamics (MD) simulations with the UNRES (UNited RESidue) force field (Liwo et al., 2001; Maisuradze et al., 2010).

$\alpha$ -syn is a 140 amino-acid protein abundant in the brain (Jakes et al., 1994; Mollenhauer et al., 2008). It plays a central role in the onset of the Parkinson disease (PD) and other neurological disorders, named synucleopathies (Soto, 2003; Chiti and Dobson, 2006; Stefanis, 2012; Chiti and Dobson, 2017; Tanudjojo et al., 2021). In these diseases,  $\alpha$ -syn is found in high concentration, as filaments, in spherical inclusions (Lewy bodies) located in the neuron cytoplasm (Spillantini et al., 1997; Breydo et al., 2012; Lashuel, 2020; Trinkaus et al., 2021). Familial cases of PD are induced either by an overexpression of WT  $\alpha$ -syn due to  $\alpha$ -syn gene triplication or by pathogenic mutations in  $\alpha$ -syn gene corresponding to single amino-acid substitution, namely, H50Q, G51D, A53E, A30P, A53T, and E46K (Polymeropoulos et al., 1997; Krüger et al., 1998; Zarranz et al., 2004; Fuchs et al., 2008; Appel-Cresswell et al., 2013; Pasanen et al., 2014; Petrucci et al., 2016). These inherited forms of PD have phenotypes that are similar to the sporadic PD, but patients with the A53T mutant exhibit an early onset of PD (Petrucci et al., 2016). The role of  $\alpha$ -syn in PD development and the mechanisms of its aggregation and of the aggregate propagation are still debatable (Henrich et al., 2020; Sang et al., 2021) and involve  $\alpha$ -syn-lipid interactions (Galvagnion et al., 2015; Suzuki et al., 2018; Antonschmidt et al., 2021), a liquid-liquid phase transition (Ray et al., 2020), and a secondary nucleation and fibril fragmentation (Cremades et al., 2012;

Cascella et al., 2021; Kumari et al., 2021). Moreover,  $\alpha$ -syn aggregation might disregulate the mitochondrial function, and increasing the level of free radicals and alterations in this pathway may be involved in the pathogenesis of PD (Hsu et al., 2000; Devi et al., 2008).

From a structural point of view,  $\alpha$ -syn is a chameleon protein (Ullman et al., 2011; Mor et al., 2016): it is disordered (IDP) in solution under physiological conditions, in equilibrium with a minor  $\alpha$ -helical tetrameric form in the cytoplasm (Bartels et al., 2011; Wang et al., 2011) and  $\alpha$ -helical when bounded to a cell membrane (Fusco et al., 2018). When incubated under physiological conditions *in vitro*,  $\alpha$ -syn aggregates into fibrils with polymorphic cross- $\beta$ -sheet conformations, in which a core of  $\beta$ -strands is aligned perpendicular to the fibril axis forming extended regular  $\beta$ -sheets with different arrangements (Tuttle et al., 2016; Guerrero-Ferreira et al., 2019; Guerrero-Ferreira et al., 2020). In addition to cylindrical fibrils, ribbon aggregates have been also observed (Bousset et al., 2013). The different polymorphs characterized *in vitro* are believed to mimic the  $\alpha$ -syn filament structures in synucleopathies.

Three main regions were identified in the primary sequence of  $\alpha$ -syn regarding their role in  $\alpha$ -syn conformational dynamics and aggregation. The N-terminal region (residues 1–60) contains a number of imperfect repeats, with the consensus motif KTKEGV, strongly similar to that found in the amphipathic helices and responsible for membrane binding (Perrin et al., 2000). High-resolution NMR structures revealed a broken helix featuring two curved  $\alpha$ -helices of  $\alpha$ -syn (residues 2–37 and 45–92) bound to micelles (Ulmer et al., 2005), whereas electron paramagnetic resonance characterizations feature an extended helix for  $\alpha$ -syn bound to a lipid membrane (first 97 residues) (Cheng et al., 2013). In solution, exposure of this N-terminal region to solvents is correlated to aggregation propensity (Stephens et al., 2020). The central hydrophobic region of  $\alpha$ -syn (residues 61–95), called NAC (non-amyloid component), contains a hydrophobic stretch of 12 residues 71VTGVTAVAQKTV82 necessary for the aggregation (Giasson et al., 2001). Fibrils of the NAC region are also found in plaques of Alzheimer's disease (Uéda et al., 1993). The flexible acidic C-terminal region regulates fibril formation *in vitro* seeding experiments (Murray et al., 2003) and plays a role in the secondary nucleation process of amyloids via electrostatic interactions with the lysine-rich N-tail (Kumari et al., 2021). Transient, long-range interactions between the negatively charged C-terminus (residues 120–140) with the positively charged N-terminus and NAC (residues 30–100) were observed by paramagnetic relaxation enhancement (PRE) for WT  $\alpha$ -syn (Dedmon et al., 2005). It is hypothesized that these transient interactions are responsible for the larger compactness of  $\alpha$ -syn compared to a disordered chain (Dedmon et al., 2005). Structures with long-range contacts between the N-terminal and C-terminal represent about 14% of the conformational ensemble, and in a significant fraction of these structures, residues 68–78



(NAC) are exposed to a solvent, in contrast to the hypothesis that such long-range transient interactions prevent aggregation (Ullman et al., 2011).

Mutations affect the arrangement and growth of the fibrils *in vitro* (Guerrero-Ferreira et al., 2020). Compared to WT, A53T and E46K mutations aggregate faster, whereas A30P aggregates more slowly than WT (Tosatto et al., 2015; Stephens et al., 2020). In addition, WT, A53T, E46K, and A30P differ in the formation of the different early oligomeric moieties (Tosatto et al., 2015). The mutants A30P and A53T show a greater propensity to form non-fibrillar aggregates than WT (Li et al., 2002), and A53T promotes seeded aggregation in human neurons (Tanudjojo et al., 2021). The rate of lipid-induced aggregation and secondary nucleation have been found to differ by multiple orders of magnitude depending on which missense mutation is involved (Flagmeier et al., 2016). NMR studies have shown an increase flexibility of  $\alpha$ -syn in nanosecond–microsecond time scales and a reduction of contacts between C- and N-terminals in mutants (Bertoncini et al., 2005). All single amino-acid substitutions have thus both a kinetic and structural effect on the formation of oligomeric structures. The purpose of the present theoretical study is the characterization of the differences between the conformational ensembles of monomeric WT and mutants, which may contribute to our understanding of the early steps of the aggregation process in solution.

The majority of monomeric WT  $\alpha$ -syn conformations have no secondary structures (Ullman et al., 2011). The average helical and  $\beta$ -strand contents of WT  $\alpha$ -syn deduced from restrained MD of  $\alpha$ -syn fragments and NMR data are about 3 and 11%, respectively (Ullman et al., 2011). The maximum fraction of residues with helical and  $\beta$ -sheet secondary structures was found to be 20% (28 residues) and 28% (39 residues), respectively (Ullman et al., 2011). Little is known on how the missense mutations modify the propensity of  $\alpha$ -syn to form secondary structures. As the  $\alpha$ -helical/ $\beta$ -sheet equilibrium is central to the folding polymorphism of  $\alpha$ -syn, we decided to characterize the propensities of these two secondary structures in WT, A30P, A53T, and E46K monomers using MD.

Because the huge conformational space of  $\alpha$ -syn is out of range for all-atom MD simulations in the explicit solvent, we applied coarse-grained UNRES MD (Maisuradze et al., 2010) to sample the structures of WT and the most studied A30P, A53T, and E46K monomers. The effective time scale of UNRES is 3 orders of magnitude larger than the all-atom time scale (Khalili et al., 2005). To ensure the convergence of the conformational sampling of the monomeric states, we applied replica exchange MD (see the Material and Methods section). The total effective time scale of the present simulations is 29.7 ms (72 replicas of 412  $\mu$ s each) for each protein studied. One of the gold standards to quantify the secondary structure elements of a protein from its structure is the Dictionary of Secondary Structure of Proteins (DSSP) (Kabsch and Sander, 1983; Touw et al., 2015) algorithm based on a simplified model of hydrogen bonds. Application of DSSP to coarse-grained structures simulated by UNRES requires to build a compatible all-atom structure from the  $C^\alpha$  coordinates of the UNRES model using reconstruction programs (Feig et al., 2004; Rotkiewicz and Skolnick, 2008). To avoid the high computational

cost of all-atom reconstructions from coarse-grained coordinates, we developed here an algorithm which assigns an  $\alpha$ -helix or a  $\beta$ -sheet secondary structure to each residue based on the  $C^\alpha$ - $C^\alpha$  distances and on the coarse-grained angles formed by  $C^\alpha$ - $C^\alpha$  pseudobonds, which correspond to the local curvature and torsion of the protein main chain (Grassein et al., 2020). The accuracy of the present algorithm, named CUTABI (CURvature and Torsion based of Alpha helix and Beta-sheet Identification), to quantify the  $\beta$ -sheet content of proteins is improved compared to an existing algorithm based on  $C^\alpha$  coordinates [P-SEA (Labesse et al., 1997)] and is comparable to the accuracy of DSSP (see the Material and Methods section). For each structure of the conformational ensemble of WT and mutants monomers, the number of residues in  $\alpha$ -helix or in  $\beta$ -sheet was computed with CUTABI. The probability density of this two-dimensional descriptor was computed by using the conformations at 300 and 310 K. Analysis of these maps and of the conformational ensembles of the WT protein and mutants revealed subtle effects of the single amino-acid substitutions which are possibly related to the differences observed in oligomerization between WT and mutants (see the Results and Discussion section).

## 2 MATERIALS AND METHODS

### 2.1 Coarse-Grained Molecular Dynamics Simulations

Detailed descriptions of the UNRES force field and its parameterization are available in the reference (Liwo et al., 2019) and at <http://www.unres.pl>. Therefore, it will be only briefly outlined here. In the UNRES force field, a polypeptide chain is represented as a sequence of  $C^\alpha$  atoms with united peptide groups located halfway of the virtual  $C^\alpha$ - $C^\alpha$  bonds and united side chains (SCs) attached to the  $C^\alpha$  atoms. The force field has been derived as the potential of mean force (PMF) of a system of polypeptide chain(s) in the solvent, where all degrees of freedom except the coordinates of the  $C^\alpha$  atoms and SC centers have been averaged out. The effective energy function contains local and site-site interactions as well as multibody terms, which have been obtained by decomposing the PMF into factors corresponding to clusters of interactions within and between coarse-grained sites (Liwo et al., 2001). The SC-SC interaction potentials implicitly include the contribution from solvation (Liwo et al., 2001; Maisuradze et al., 2010). The force field was calibrated to reproduce the structure and thermodynamics of small model proteins and applied with success to simulate protein folding (Maisuradze et al., 2010; Zhou et al., 2014; Sieradzan et al., 2021) and large-scale conformational dynamics (Gołaś et al., 2012).

All structures of  $\alpha$ -syn (WT and mutants) were extracted from replica exchange MD trajectories generated with the UNRES force field. A total of 72 trajectories were computed for each protein: 32 trajectories at 300 K and 8 trajectories at each of the following temperatures, 310 K, 323 K, 337 K, 353 K, and 370 K. Each trajectory was started with two fully unfolded monomers separated by a distance of 25 Å. The Cartesian coordinates of  $C^\alpha$  and SC beads were saved every 1,000 integration steps. The

integration time step in UNRES is 4.9 fs, corresponding to an effective actual time step of about 4.9 ps. Due to the implicit integration of fast motions, the time scale of UNRES compared to the experimental time scale is indeed accelerated by a factor of 1,000 (Khalili et al., 2005). For each trajectory, the first 4 million steps were discarded. After this relaxation period, 8 consecutive runs of 10 million steps were used for the analysis of each trajectory. Convergence of the conformational ensemble simulated for each protein was monitored by the probability densities of intra-chain and inter-chain contacts and of the radius of gyration. For each trajectory, the convergence was achieved for the last 3 runs of 10 million steps, i.e., for a statistics of 30 million steps (an effective time scale of 147 microseconds) out of 84 million steps (an effective time scale of 412 microseconds). Only structures at 300 and 310 K are reported here as they are close to the physiological temperature. The structures were saved every 1,000 integration steps, leading to about 1 million of converged conformations analyzed for each protein combining the results at 300 and 310 K.

Since the simulations are performed on two monomers, both isolated non-interacting monomer conformations and aggregated monomers were observed in the converged MD trajectories. The analysis and description of the dimeric conformations are out of the scope of the present paper, which is dedicated to isolated monomers, and they will be described elsewhere. The monomeric structures studied in the present work were extracted from the converged replica exchange MD trajectories if they obey the following condition: no residues at a distance smaller than 20 Å from the other monomer. The fraction of monomers out of all the conformers simulated at 300 and 310 K is 55% for WT, 65% for A30P, 50% for A53T, and 48% for E46K. The monomeric state is finally described here by about 1 million of structures of the converged production part of the MD trajectories at 300 and 310 K, representing a sampling on an effective time scale of 4.9 ms.

## 2.2 Curvature and Torsion Based of $\alpha$ -Helix and $\beta$ -Sheet Identification: An Algorithm for Secondary Structure Determination Based on $C^\alpha$ Coordinates

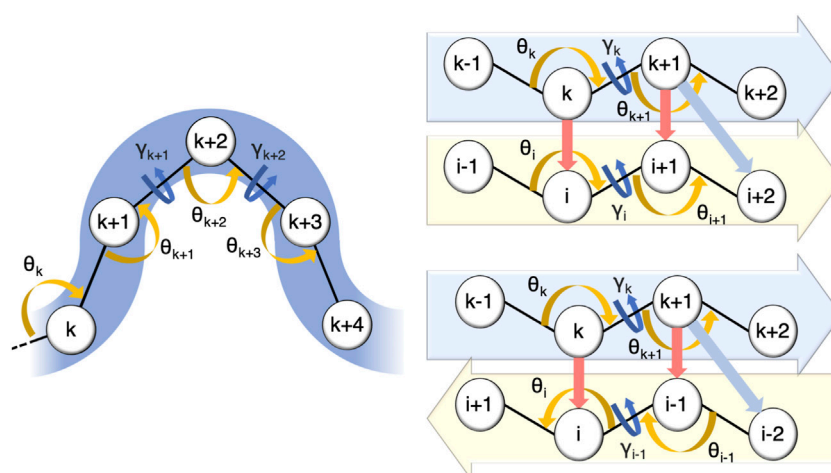
Secondary structure elements (SSEs) are important descriptors of the native state of proteins (Konagurthu et al., 2020) and of the conformational ensemble of IDP, as shown in the next section for  $\alpha$ -syn. The main local structures stabilized by backbone hydrogen bonds are helices ( $\alpha$ -,  $3_{10}$ -, and  $\pi$ -helices) and  $\beta$ -sheets. Deviations from the canonical definitions of these secondary structures based on Ramachandran angles are common in the Protein Data Bank. Therefore, a practical and widely used method to assign an SSE to a residue is to apply the DSSP algorithm which is based on the calculation of a simplified energy function describing backbone hydrogen bonds (Kabsch and Sander, 1983). Application of DSSP to a protein structure requires the knowledge of the Cartesian coordinates of all backbone atoms. For structures simulated by coarse-grained force fields, like UNRES, only a subset of these coordinates is known, those of the  $C^\alpha$  atoms. Assignment of the SSE based on  $C^\alpha$

coordinates can be performed using DSSP but at the extra computational cost of the reconstruction of an all-atom representation of the coarse-grained structure using software like MMTSB (Feig et al., 2004) or PULCHRA (Rotkiewicz and Skolnick, 2008). Alternatively, the SSE can be defined from the  $C^\alpha$  coordinates only [P-SEA algorithm (Labesse et al., 1997)]. Here, we have chosen the second faster option and have developed the algorithm CUTABI by analyzing 14,652 experimental structures with less than 40% of sequence identity extracted from the ASTRAL database (Fox et al., 2014; Chandonia et al., 2019). The parameters of CUTABI were adjusted to reproduce the DSSP assignment of helices and  $\beta$ -sheets for the ASTRAL database. The agreement between CUTABI and DSSP is 93% for the structures in the database with 10 times (DSSP + PULCHRA) to 30 times (DSSP + MMTSB) less computational cost.

### 2.2.1 Parameters of Curvature and Torsion Based of $\alpha$ -Helix and $\beta$ -Sheet Identification

Assuming a constant virtual bond length between  $C^\alpha$  atoms of successive residues, a chain of  $N$  amino acids is fully characterized by  $N - 3$  torsion angles  $\gamma_n$ , built from the positions of  $C_{n-1}^\alpha$ ,  $C_n^\alpha$ ,  $C_{n+1}^\alpha$ , and  $C_{n+2}^\alpha$  with  $n = 2$  to  $N - 2$ ; and  $N - 2$  bond angles  $\theta_n$ , built from  $C_{n-1}^\alpha$ ,  $C_n^\alpha$ , and  $C_{n+1}^\alpha$ , with  $n = 2$  to  $N - 1$ . These angles have clear geometrical meanings: they are respectively the discrete version of the local curvature ( $\theta_n$ ) and the local torsion ( $\gamma_n$ ) of the chain formed by the successive  $C^\alpha - C^\alpha$  virtual bonds (Grassein et al., 2020). From a mathematical point of view, the local curvature and torsion fully describe the structure of a string and form a complete set of local order parameters for protein folding (Grassein et al., 2020). For proteins, the curvature has a limited range with  $\theta$  varying between  $80^\circ$  and  $160^\circ$ , whereas  $\gamma$  can take nearly any value between  $-180^\circ$  and  $+180^\circ$ . The SSEs correspond statistically to specific areas in the coarse-grained ( $\gamma$ ,  $\theta$ ) maps. **Supplementary Figure S1** shows the distribution of ( $\gamma$ ,  $\theta$ ) angles computed by the DSSP algorithm for the 14,652 experimental structures extracted from the ASTRAL database for helix (H + G + I) and  $\beta$ -sheet (E) (in brackets the one-letter SSE codes in DSSP). The areas corresponding to helix and  $\beta$ -sheet identified by DSSP in the ( $\gamma$ ,  $\theta$ ) map are nicely reproduced by the CUTABI algorithm, as shown in **Supplementary Figure S1**, using an identification of the SSE based on the coordinates of the  $C^\alpha$  atoms as follows.

In CUTABI, the minimal size of a helix is set to 4 residues. Helices with less than 3 residues, such as short  $3_{10}$  helices, are thus not counted. Because helices involve short-range interactions along the amino-acid sequence, the local curvature  $\theta$  and torsion  $\gamma$  parameters are sufficient to characterize this SSE. To decide if a set of four residues  $k$ ,  $k + 1$ ,  $k + 2$ , and  $k + 3$  pertains to a helix, a combination of the coarse-grained angles formed by residues ( $C^\alpha$  atoms)  $k - 1$ ,  $k$ ,  $k + 1$ ,  $k + 2$ ,  $k + 3$ , and  $k + 4$  is considered, as shown in **Figure 1**. Residues  $k$ ,  $k + 1$ ,  $k + 2$ , and  $k + 3$  pertain to a helix if the angles  $\theta_k$ ,  $\theta_{k+1}$ ,  $\theta_{k+2}$ , and  $\theta_{k+3}$  belong to an interval between  $80^\circ$  and  $105^\circ$  and the angles  $\gamma_{k+1}$  and  $\gamma_{k+2}$  belong to an interval between  $30^\circ$  and  $80^\circ$ . A window of 4 residues is slid along the sequence to evaluate the residues pertaining to a helix in each  $\alpha$ -syn conformation.



**FIGURE 1** | Parameters defining residues pertaining to a helix and to a parallel or to an anti-parallel  $\beta$ -sheet in CUTABI (see text). The spheres represent the positions of  $C^\alpha$  atoms of the residues. Orange and dark-blue arrows show the  $\theta$  and  $\gamma$  angles involved, respectively. For  $\beta$ -sheets, red and light-blue arrows point out the distances between residues involved in their definitions.

In CUTABI, the minimal size of a  $\beta$ -strand is set to 2 residues, i.e., a  $\beta$ -sheet cannot be smaller than 4 residues. The  $\beta$ -strands of 1 residue forming  $\beta$ -bridges (B code in DSSP) are thus not considered. To define parallel and anti-parallel  $\beta$ -sheets (Figure 1), both local curvature and torsion parameters as well as distances between the two  $\beta$ -strands forming the sheet are necessary. If the four residues  $k$ ,  $k + 1$ ,  $i$ , and  $i + 1$  pertain to a parallel  $\beta$ -sheet (Figure 1), then the angles  $\theta_k$ ,  $\theta_{k+1}$ ,  $\theta_i$ , and  $\theta_{i+1}$  must be between  $100^\circ$  and  $155^\circ$  and the angles  $\gamma_k$  and  $\gamma_i$  must be smaller than  $-80^\circ$  or larger than  $80^\circ$ . In addition, the distances between  $k$  and  $i$  and  $k + 1$  and  $i + 1$  (Figure 1, red arrows) must be smaller than  $5.5 \text{ \AA}$ , and the distance between  $k + 1$  and  $i + 2$  (Figure 1, light blue arrows) must be smaller than  $6.8 \text{ \AA}$ . Similar conditions must be met for the anti-parallel  $\beta$ -sheet, as shown in Figure 1.

### 2.2.2 Performance of Curvature and Torsion Based of $\alpha$ -Helix and $\beta$ -Sheet Identification Compared to Dictionary of Secondary Structure of Proteins and P-SEA

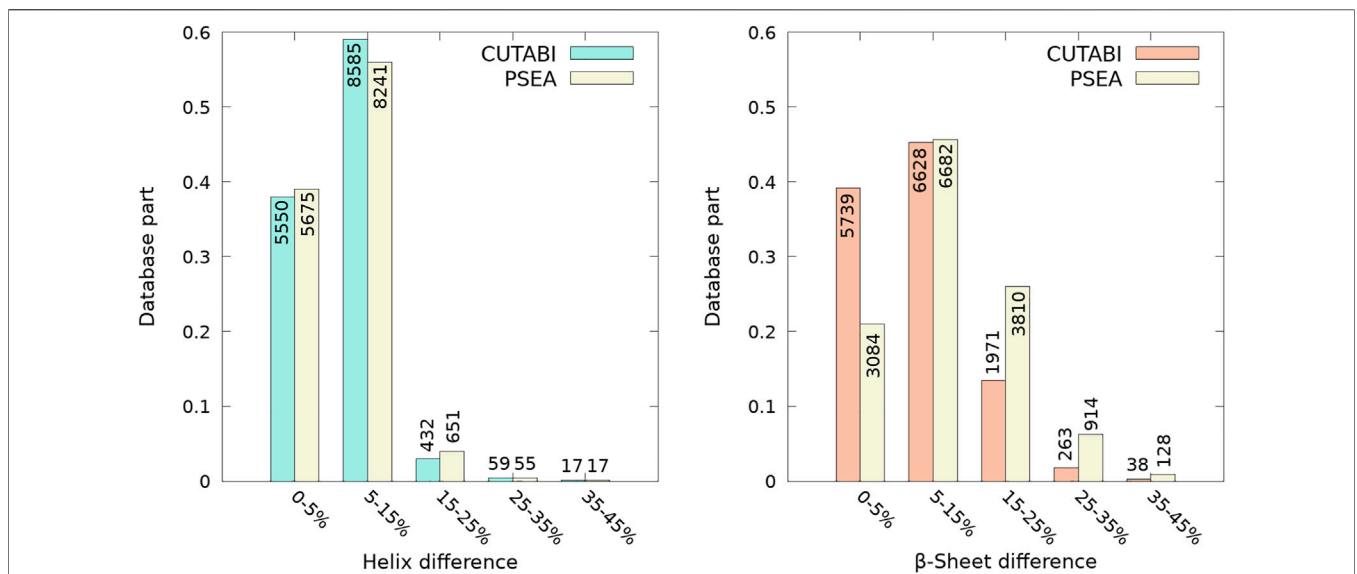
The SSEs computed with the parameters defined in Figure 1 were compared to the SSEs calculated with DSSP (based on all-atom coordinates) (Kabsch and Sander, 1983) and P-SEA (based on  $C^\alpha$  coordinates only) (Labesse et al., 1997). Figure 2 shows the percentage of difference, i.e., the number of residues having an SSE different in the coarse-grained algorithms P-SEA and CUTABI compared to DSSP, divided by the sequence length. For helices, as shown in Figure 2 (left panel), 96% of structures ( $5,550 + 8,585 = 14,235$  structures) do not have more than 15% of residues with an SSE different in CUTABI and in DSSP. The results of P-SEA are similar. Considering DSSP as a gold standard, the CUTABI accuracy is improved compared with P-SEA on the evaluation of residues in the  $\beta$ -sheet. As shown in Figure 2 (right panel), 5,739 structures examined (40% of the database) have less than 5% of residues pertaining to an SSE

different in CUTABI compared to DSSP. This number is only 3,084 (20% of the database) in P-SEA. In total, assignments of SSE by CUTABI and DSSP agree for 84% of the structures studied ( $5,739 + 6,628 = 12,367$  structures) with a difference compared to DSSP of maximum 15% of the sequence length. The average of the percentage of agreement over all the structures of the database (14,652) for all SSEs between CUTABI and DSSP is 93% for a computational time at least 10 times smaller.

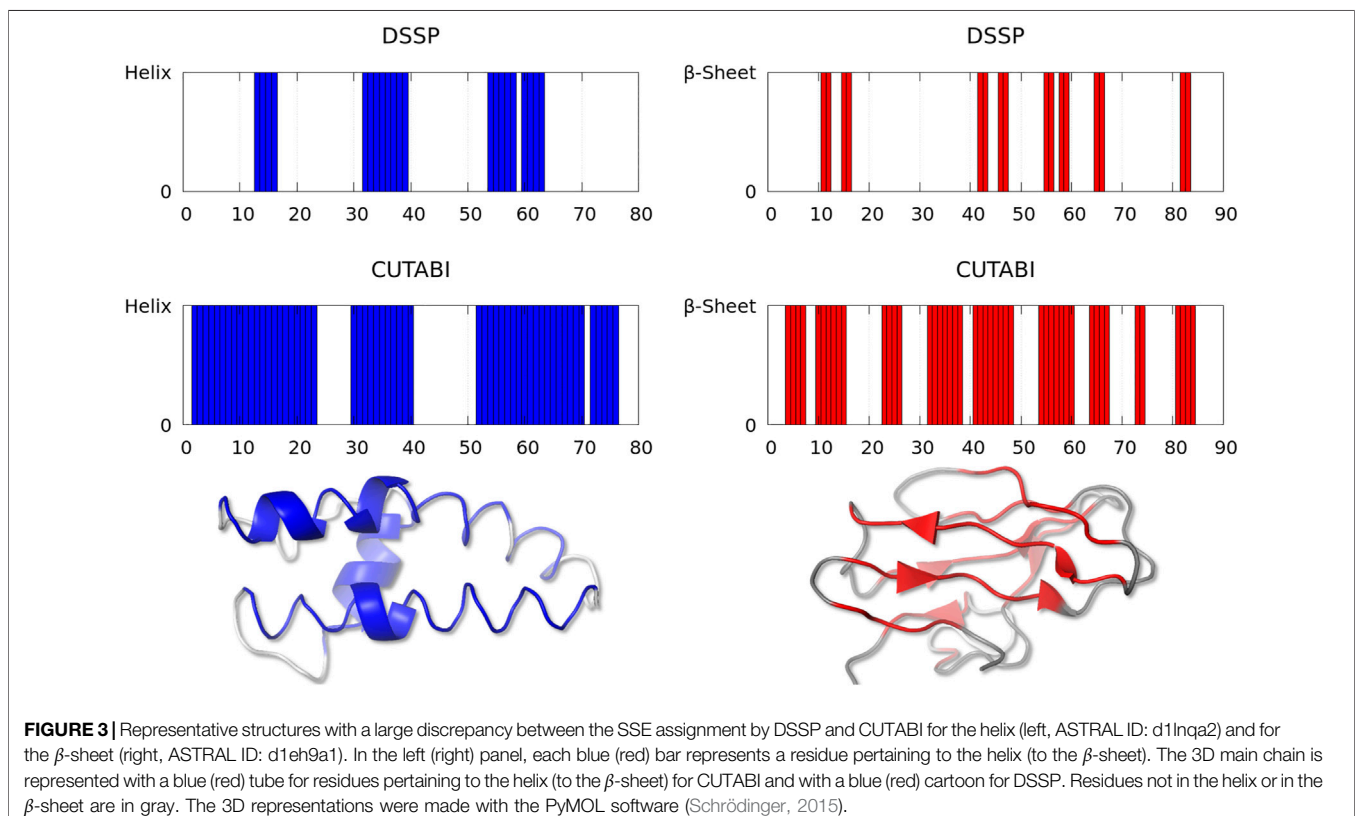
As shown in Figure 2, one observes a large difference of secondary structure assignments between CUTABI and DSSP for a very small number of proteins. Most structures in this category were measured by X-ray diffraction but with a low resolution ( $>3 \text{ \AA}$ ). In addition, as hydrogen atom positions are not detected in the X-ray (except for ultra-high resolution), the application of DSSP to these structures may be less accurate as DSSP is based on the calculation of hydrogen bond energy. To illustrate the precise origin of this finding, we examined the structures and selected two representative structures with 35% of difference in SSE: one for the helix (Figure 3, left blue panel) and another for the  $\beta$ -sheet (Figure 3, right red panel).

Figure 3 (left panel) represents the fragment (from residue 19 to 98) of a calcium-gated potassium channel (PDB ID: 1lnq and ASTRAL ID: d1lnqa2). In this example, CUTABI detects a much larger number of residues in the helix than DSSP. However, the 3D representation shows that although these extra residues missed by DSSP are not in a canonical helix, the overall shape of the main chain is indeed helical.

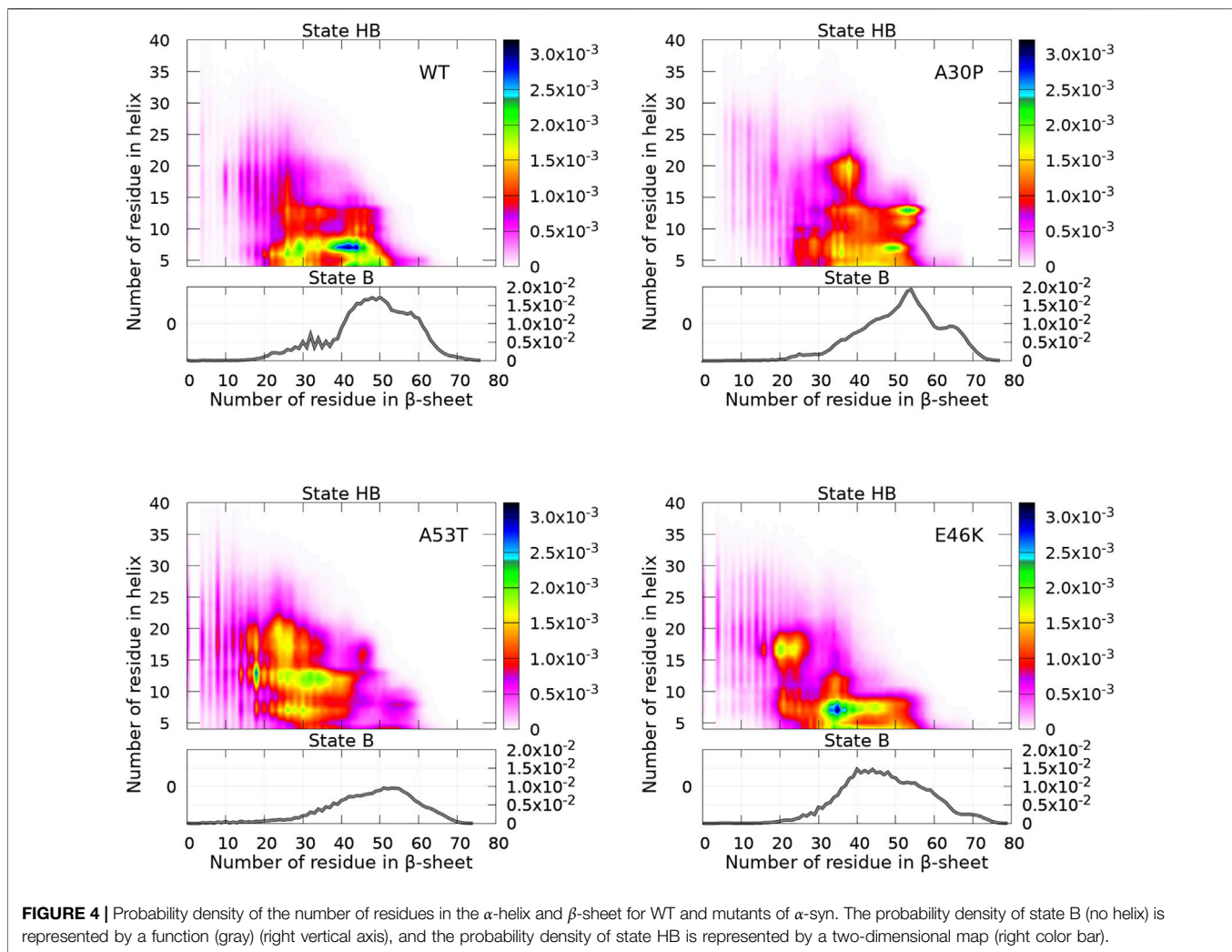
Figure 3 (right panel) represents the fragment (from residue 1 to 90) of a glycosyltrehalose trehalohydrolase (PDB ID: 1eh9 and ASTRAL ID: d1eh9a1). The number of residues in the  $\beta$ -sheet is much larger in CUTABI than in DSSP. As in the case of the helix, the 3D representation indicates that the additional residues in the  $\beta$ -sheet in CUTABI are part of a main chain segment with the overall shape of the  $\beta$ -sheet, although not canonical, due probably to the low experimental resolution.



**FIGURE 2** | Comparison between the identification of residues pertaining to the helix and to  $\beta$ -sheet between the all-atom DSSP algorithm and coarse-grained CUTABI and P-SEA algorithms. Each bar of the histograms represents the number of structures of the ASTRAL database as a function of a range of percentages of the difference between CUTABI (green and orange) and DSSP and between P-SEA (yellow) and DSSP. For each protein, the percentage is calculated as the number of residues having a different secondary structure in the coarse-grained algorithms and DSSP, divided by the protein sequence length.



**FIGURE 3** | Representative structures with a large discrepancy between the SSE assignment by DSSP and CUTABI for the helix (left, ASTRAL ID: d1lnqa2) and for the  $\beta$ -sheet (right, ASTRAL ID: d1eh9a1). In the left (right) panel, each blue (red) bar represents a residue pertaining to the helix (to the  $\beta$ -sheet) for CUTABI and with a blue (red) tube for DSSP. Residues not in the helix or in the  $\beta$ -sheet are in gray. The 3D representations were made with the PyMOL software (Schrödinger, 2015).



### 3 RESULTS

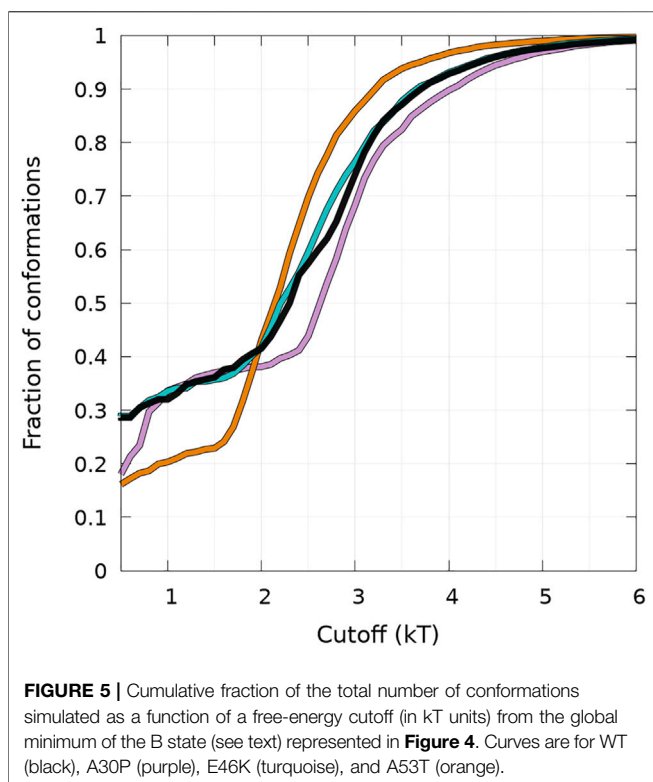
#### 3.1 $\alpha$ -Helix and $\beta$ -Sheet Propensities of $\alpha$ -syn Divide the Conformations Into Two Distinct States

The algorithm CUTABI was applied to the  $\alpha$ -syn conformations to compute the number of residues in the  $\alpha$ -helix ( $\alpha$ ) and  $\beta$ -sheet ( $\beta$ ) of each structure for WT and mutants. Each conformation has ( $\alpha$ ,  $\beta$ ) coordinates. The resulting probability densities in the ( $\alpha$ ,  $\beta$ ) space are represented in **Figure 4**. In these maps, only the residues from the N-terminal and NAC regions were considered for the calculations [as can be seen in the next section (**Figure 8**), the C-terminal region does not contribute to SSE differences between WT and mutants].

A major observation is that the conformations are divided into two distinct states for the N-terminal + NAC region: an ensemble of conformations with no residue in the helix (state B) and the rest of conformations (state HB). The highest probability of observing a conformation in state B is an order of magnitude

larger than that of state HB, as can be seen by comparing (the scale of) two-dimensional and one-dimensional plots in **Figure 4**. In addition, **Figure 4** reveals clear differences between WT and mutants.

First, we discuss the global differences between states B of WT and mutants (one-dimensional functions in **Figure 4**). The maximum of the distributions is at 50, 54, 53, and 40 for WT, A30P, A53T, and E46K, respectively. The distribution is sharper for A30P, which has the largest number of conformations with the largest number of residues (between 60 and 70) in the  $\beta$ -sheet. Clearly, A53T has the lowest number of conformations in state B. This is even better seen in **Figure 5**, showing the fraction of conformations within a free-energy difference cutoff from the global minimum of state B for each protein. With  $P_{\max}$  being the maximum of probability at (0,  $\beta$ ) (in the B state) and  $P$  being the probability at ( $\alpha$ ,  $\beta$ ) ( $\alpha \geq 0$ , in the B or HB states), the free-energy cutoff is computed as  $-\ln(\frac{P_{\max}}{P})$  in  $kT$  units, where  $k$  is the Boltzmann constant and  $T$  is the temperature. Within 1  $kT$ , there is 32% of the conformations found for WT, E46K, and A30P and only 20% for A53T, as shown in **Figure 5**.



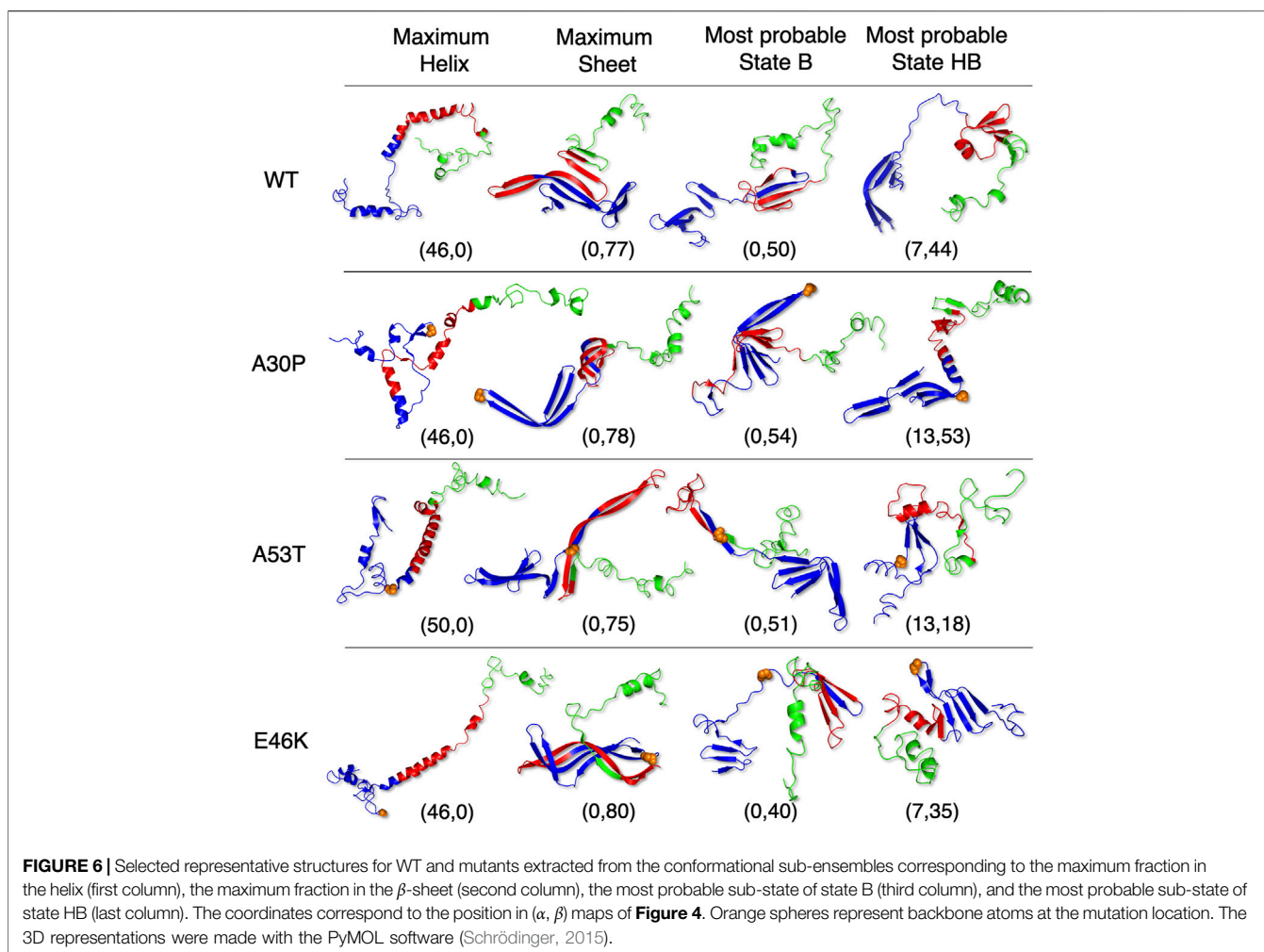
The results illustrated in **Figure 5** emphasize the two-state behavior. For A30P, there is even a third state below 1 kT. By definition, the derivative of the curves represented in **Figure 5** represents the Density Of conformations or micro-States (DOS). The nearly linear behavior of the curves for WT and E46K up to 2 kT means a rather constant DOS corresponding to state B (one-dimensional probability densities in **Figure 4**). The change of slope at 2 kT for these proteins points up the separation between the states, i.e., the onset of state HB, i.e., a state with a mixture of  $\alpha$ -helices and  $\beta$ -sheets. State B represents thus 40% of the conformations for WT and E46K. Between 2 kT and about 3 kT, the curves for WT and E46K in **Figure 5** are linear with a slope larger than that in the B state, which corresponds to a larger DOS in state HB. Beyond 3 kT, the DOS decreases and reaches a plateau for all proteins. For A53T, state B contains less conformations and the onset of the state HB occurs at about 1.5 kT. State B represents only about 25% of the conformations for A53T. The DOS of the HB state for A53T is higher than that in its B state, and it is also higher than the DOS of the HB state for WT and E46K. The case of A30P is special. State B has the highest DOS up to 1 kT, and it becomes similar to the DOS for WT and E46K up to 2.5 kT. The third state below 1 kT is clearly visible as a shoulder with a large proportion of conformations with 60–70 residues in the  $\beta$ -sheet in the one-dimensional probability distribution, shown in **Figure 4**. Finally, state B in A30P also concerns 40% of the structures. The DOS in the HB state for A30P is similar to the one of A53T.

**Figure 5** shows that state HB encompasses about 60% of the conformations for WT, A30P, and E46K (all conformations beyond 2 kT for WT and E46K and beyond 2.5 kT for A30P)

and 75% of the conformations for A53T (all conformations beyond 1.5 kT). Although the DOS is relatively constant, local maxima occur in the HB two-dimensional maps (**Figure 4**). The most probable ( $\alpha$ ,  $\beta$ ) conformations occur at (7,44), (13,53), (13,18), and (7,35) for WT, A30P, A53T, and E46K. There is a significant difference between A53T and WT, A30P, and E46K. For each protein, selected structures for the maximum of probability of the B state and for the local maxima of the probability of the HB state are represented in **Figure 6**. They illustrate the expected diversity of conformations of an IDP. It is, however, important to emphasize that each ( $\alpha$ ,  $\beta$ ) pair represents an ensemble of conformations. It is impossible to represent the entire diversity of these sub-ensembles. For example, for the WT protein, the maximum of the B state at (0,50) corresponds to 19,901 structures, and the maximum of the HB state at (7,44) corresponds to 3,569 structures. In **Figure 6**, the structures with a maximum of residues in the helix or in the  $\beta$ -sheet are shown. The maximum number of residues in the helix is 50 for A53T compared to 46 for the other proteins and compared to the 82 (Ulmer et al., 2005) and 97 residues (Cheng et al., 2013) in helices when  $\alpha$ -syn is bound to membranes.

The average radius of gyration  $R_g$  is a common global structural parameter in polymer science, and it is interesting to relate this property to states B and HB. For a random coil represented by a self-avoiding walk in a good solvent (i.e., for which interactions between monomers and solvent molecules are energetically favorable),  $R_g = 0.367bN^\nu$ , where  $b$  is the length of the so-called statistical segment and  $\nu$  is a fractal exponent. In three dimensions, we take  $b = 7.6 \text{ \AA}$  (twice the distance between two  $C^\alpha$ ),  $R_g = 54 \text{ \AA}$  for the size of  $\alpha$ -syn ( $N = 140$ ) (Victor et al., 1994). The average radius of gyration  $R_g$  computed from the UNRES trajectories is 24.7, 25.1, 26.0, and 25.2  $\text{\AA}$  for WT, A30P, A53T, and E46K, respectively. Only A53T has a significantly larger  $R_g$  than WT. Other authors reported an average radius of gyration of  $R_g \approx 23 \text{ \AA}$  for WT and a distribution of  $R_g$  narrower than the distribution of a random coil of a similar sequence length (Dedmon et al., 2005; Allison et al., 2009).

The distributions of  $R_g$  of the conformational ensemble simulated with UNRES for WT and mutants are very similar to each other, as shown in **Figure 7**. The peak of the probability density is lower for A30P. When the global distribution is divided into states B and HB, one reaches the same conclusion for state HB (**Figure 7**, middle panel), but one observes more significant differences between the proteins for state B (**Figure 7**, bottom panel). For state B, the peak of A53T is the highest and a sub-population appears clearly on the left side of the distribution for A30P and E46K. This sub-population is hardly visible in the global probability distribution (**Figure 7**, top panel) as a shoulder. Examination of the structures corresponding to this sub-state of the B state reveals that the structures of A30P and E46K have a large proportion of contacts between regions 1–20 and 96–140. The average number of contacts of the structures with  $17.9 \text{ \AA} < R_g < 18.1 \text{ \AA}$  is 5.2, 14.9, 0.1, and 9.0 for WT, A30P, A53T, and E46K, respectively. The large number of contacts between the two extremities of the protein for A30P and E46K mutants compared to WT and A53T explains the peak at 18  $\text{\AA}$ .

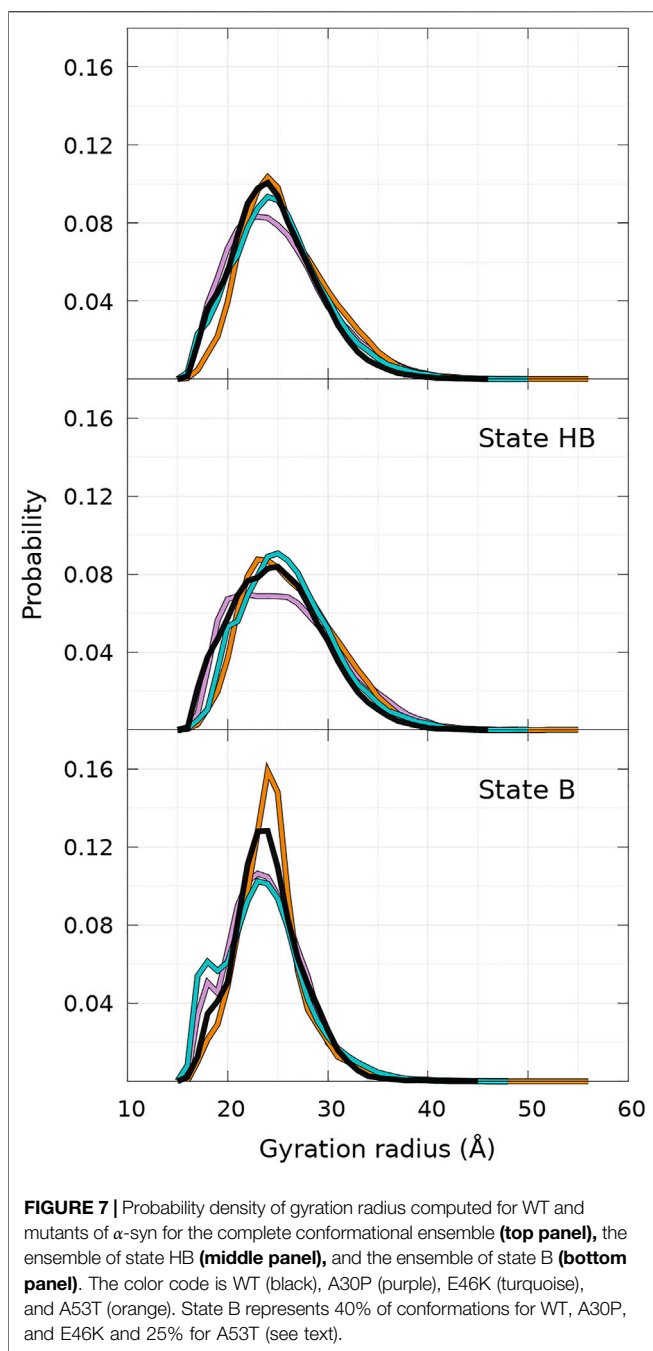


### 3.2 Secondary Structure Element Propensities as a Function of the Position in the Amino-Acid Sequence: Differences and Similarities Between the Wild Type and Mutants

As shown in **Figure 8**, helices are found in four main regions: two in the C-terminal region (residues 119–125 and 127–130), one in the NAC region (residues 75–82), and one overlapping the N-terminal and NAC regions (residues 53–65) for both WT and mutants. As claimed in the previous section, there are no significant differences between the propensities of SSE for WT and mutants in the C-terminal region. The major differences between WT and mutants occur in the region 53–65, which has a peak for the helix propensity at residue K58. For all mutants, the probability to form an  $\alpha$ -helix in this region is larger than that for WT. For A53T, the probability of residue K58 to pertain to a helix is more than twice higher than one for the WT protein ( $P_{WT} = 0.25$  compared to  $P_{A53T} = 0.56$ ). Mutation 53 occurs in the N-terminal part of a helical region of WT, but its effect is not trivial as the mutation could be naively expected to decrease the helicity. Indeed, the propensity to be part of a helix is 1.45 for

alanine and 0.82 for threonine according to the empirical helix propensity scale (Chou and Fasman, 1974). The increase of helicity observed in the 53–65 region of the sequence upon single mutation is clearly not a local effect. Similarly, single amino-acid substitutions, A30P and E46K, also increase significantly the helicity in the region of 53–65.

Compared to other proteins, another significant difference is observed for A53T in the helical region located at residues 18–22, where the probability to form an  $\alpha$ -helix is significantly larger for this mutant. For example, the peak at E20 corresponds to the following probabilities:  $P_{WT} = 0.07$ ,  $P_{A30P} = 0.04$ ,  $P_{A53T} = 0.14$ , and  $P_{E46K} = 0.05$ . The significant increase of helicity in the 18–22 and 53–65 regions explains why state B of A53T is less populated, as shown in **Figures 4, 5** and discussed in the previous section. The presence of an  $\alpha$ -helix in the 18–22 region might explain why mutation A18T induces significant modification of the  $\alpha$ -syn polymerization (Kumar et al., 2018). Simulations of this mutant are scheduled in the future. For A30P, the probability to observe a helix in the NAC region is slightly lower than for other proteins: the probabilities at A78 are  $P_{WT} = 0.37$ ,  $P_{A30P} = 0.32$ ,  $P_{A53T} = 0.40$ , and  $P_{E46K} = 0.38$ .



As shown in **Figure 8**, the probabilities to find residues in the  $\beta$ -sheet are significantly high in the N-terminal and NAC regions up to about residue 100 at the same locations for WT and mutants. For WT, the maximum of the peaks observed in **Figure 8** are  $P_{F4} = 0.57$ ,  $P_{K10} = 0.61$ ,  $P_{A17} = 0.87$ ,  $P_{T22} = 0.76$ ,  $P_{E28} = 0.65$ ,  $P_{K34} = 0.61$ ,  $P_{Y39} = 0.69$ ,  $P_{K43} = 0.50$ ,  $P_{V49} = 0.4$ ,  $P_{V55} = 0.69$ ,  $P_{V63} = 0.64$ ,  $P_{V70} = 0.32$ ,  $P_{K80} = 0.41$ ,  $P_{A90} = 0.58$ , and  $P_{V95} = 0.35$ .

As shown in **Figure 8**, the  $\beta$ -sheet probabilities are very similar for WT and E46K along the sequence and differ significantly for A30P in the region 26–80 and for A53T in the region 35–65.

The mutation A30P has a huge local impact on the probability of residue 30 to pertain to a  $\beta$ -sheet,  $P_{WT,A30} = 0.49$  and  $P_{P30} = 0.03$ . This induces an unexpected increase of the probability to occur in the  $\beta$ -sheet for the neighboring residues:  $P = 0.85$  for residues 26–29 and 32–35. Long-range effects of A30P mutation on the propensities of other residues are observed by an increase of the peaks at  $P_{Y39} = 0.76$ ,  $P_{K43} = 0.57$ ,  $P_{V70} = 0.42$ ,  $P_{T75} = 0.46$ , and  $P_{V95} = 0.46$  and by a decrease of the peaks at  $P_{V55} = 0.53$  and  $P_{V63} = 0.54$ , compared to WT. The decrease in  $\beta$ -sheet propensity in this region compared to WT agrees with their larger helical propensity in this region for the A30P mutant (**Figure 8**). For A53T, the most drastic effect of the single amino-acid substitution occurs in the region 53–65, where the probability to form  $\beta$ -sheets is significantly reduced compared to WT,  $P_{V55} = 0.32$  and  $P_{V63} = 0.40$ , in agreement with their high probabilities to be in a helix (**Figure 8**). Other significant long-range effects of the amino-acid substitution are observed at peaks  $P_{Y39} = 0.55$ ,  $P_{T44} = 0.37$ , and  $P_{V49} = 0.18$ , where the mutation A53T decreases the probability to form a  $\beta$ -sheet compared to WT. Overall, two crucial regions of the amino-acid sequence are affected by the mutations: the region 26–35, where the A30P mutation increases mainly the  $\beta$ -sheet formation, and the region 53–65, where the A53T mutation mainly decreases the  $\beta$ -sheet formation.

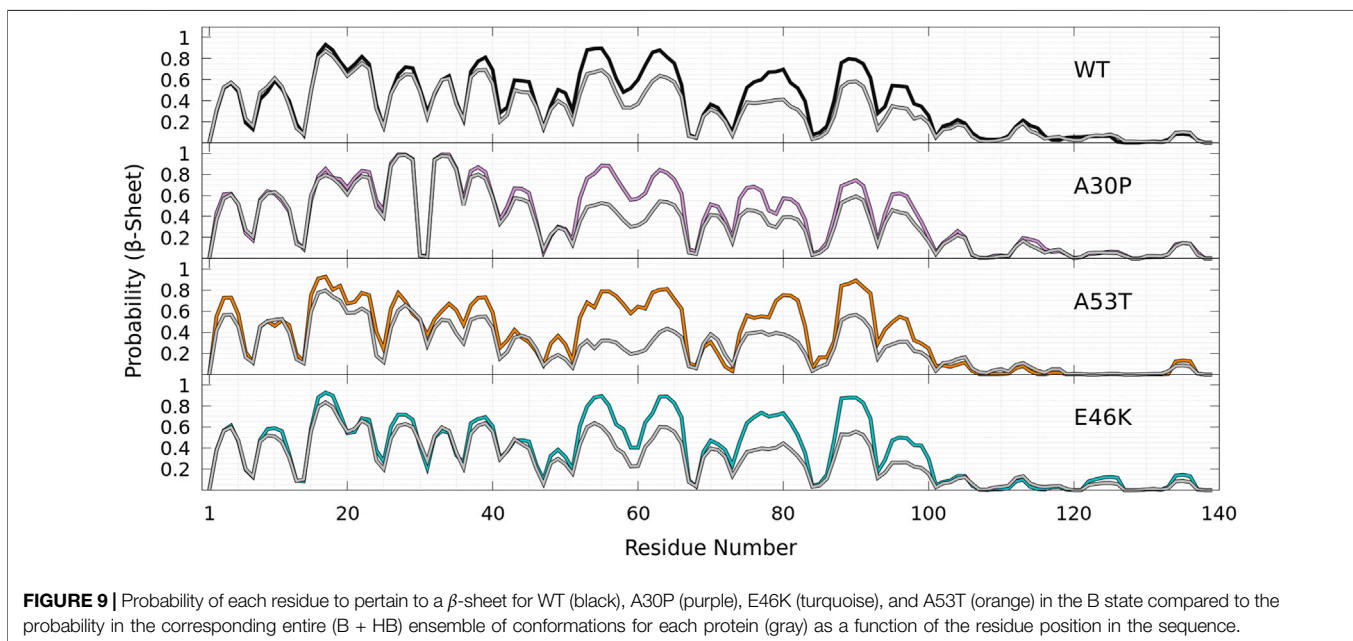
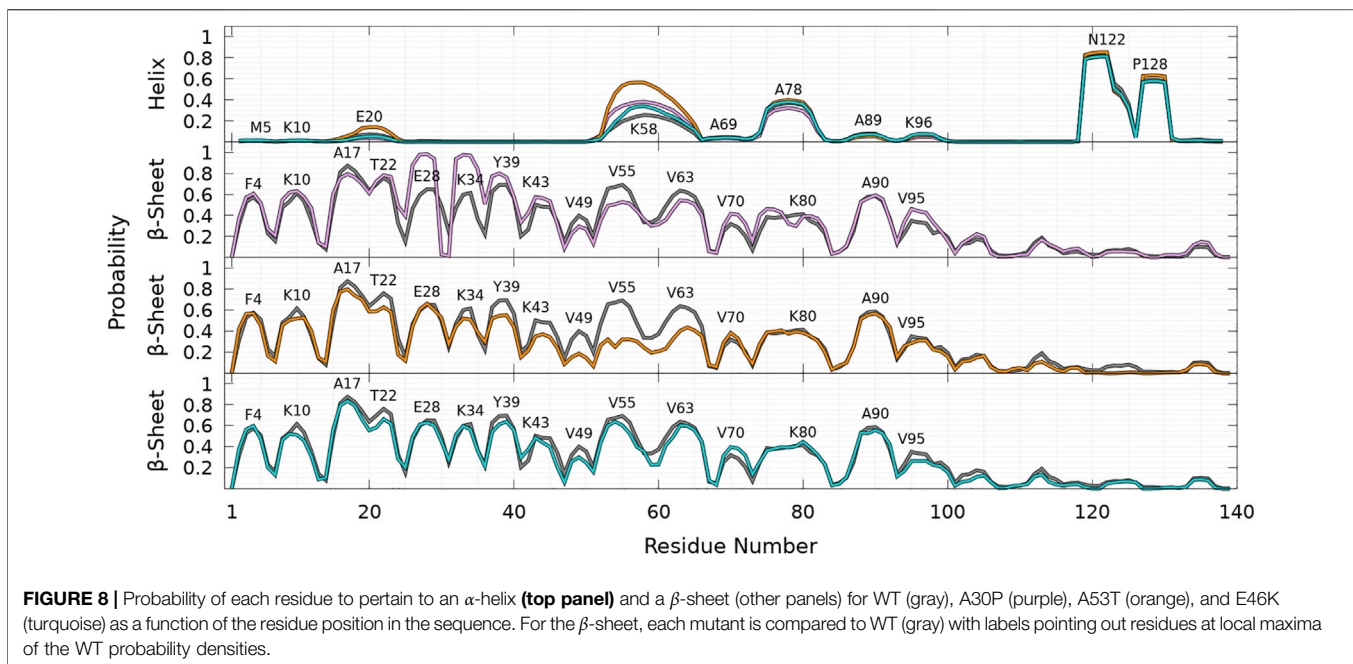
Finally, as shown in **Figure 9** for all proteins, the statistics of  $\beta$ -sheet propensity of state B differs from the global statistics (B + HB) in the region 50–100, which encompasses the two major helical regions centered at K58 and A78 (**Figure 9**, top panel). The most significant differences are observed for A53T, for which the probability to form a  $\beta$ -sheet is significantly lower in the HB state in the region 50–70, where the probability to form an helix is very high (**Figure 9**).

## 4 DISCUSSION

The present analysis is based on a first-principle (no bias or constraints applied) very large conformational sampling of the WT  $\alpha$ -syn and mutants. As in any MD simulations, the sampling is never complete and each force-field has some bias. Next, we compare the predictions of UNRES for WT and mutants to available experimental data and previous theoretical studies.

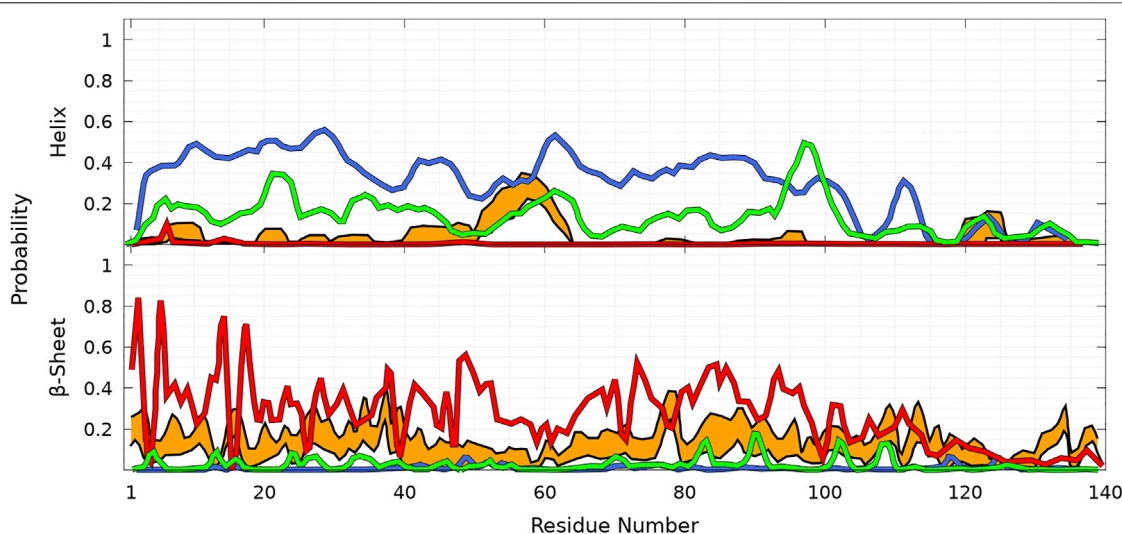
To illustrate the difficulties of producing a conformational ensemble of  $\alpha$ -syn, we compared (see **Figure 10**) the helix and  $\beta$ -sheet probabilities computed in four different works using MD simulations for WT  $\alpha$ -syn (Ullman et al., 2011; Yu et al., 2015; Ramis et al., 2019; Coskuner and Wise-Scira, 2013). Compared to the UNRES simulations and Ref. Ullman et al., 2011, the calculations of Ref. Yu et al., 2015 largely overestimate the helical properties of  $\alpha$ -syn, whereas the simulations of Ref. Ramis et al., 2019 largely underestimate them. The simulations of Ref. Coskuner and Wise-Scira, 2013 also predict a larger helical propensity than the one found in the present work (**Figure 8**) and in Ref. Ullman et al., 2011 (**Figure 10**). It is worth noting that the helical region nearby residue 60 is found in the UNRES simulations (**Figure 8**) and in Refs. Ullman et al., 2011; Coskuner and Wise-Scira, 2013 (**Figure 10**).





Comparison with our predictions shown in **Figure 8** indicates better agreement with Ref. Ullman et al., 2011 for the helical part, except that UNRES predicts a larger probability to form an  $\alpha$ -helix at residue A78. In the present work, MD simulations are performed for the entire protein. In contrast, the method of Ref. Ullman et al., 2011 consists of a combination of MD simulations of peptides of  $\alpha$ -syn with sequences which overlap each other and are extracted from the  $\alpha$ -syn sequence. The conformation of the protein is rebuilt by a combination of these fragments with weights using NMR data. The formation of transient helical segments was observed experimentally in the

WT  $\alpha$ -syn tetramer in Ref. Wang et al., 2011 between residues 4–43 and 50–103. Larger probabilities of helical tendency were found experimentally in regions 20–23 and 48–90 (Wang et al., 2011). These findings are rather in good agreement with the transient existence of helices in the  $\alpha$ -syn monomer, as shown in **Figure 8** (top panel), including the peak at A78. The stability of the tetrameric form is supported by previous all-atom MD simulations showing that the stability of this helical oligomer is due to a hydrophobic core formed by non-polar residues pertaining to the second  $\alpha$ -helix of each chain, along with salt bridges formed by positively charged lysine residues and



**FIGURE 10** | Probability of each residue of WT to pertain to an  $\alpha$ -helix (**top panel**) and a  $\beta$ -sheet (**bottom panel**) as a function of its position in the sequence from MD simulations according to four different works: Ullman et al., 2011 (orange), Yu et al., 2015 (blue), Ramis et al., 2019 (red), and Coskuner and Wise-Scira, 2013 (green). The thickness of the orange lines corresponds to the 95% confidence interval. Excerpted of the published figures of the works cited are reproduced with permissions. Further permissions related to the material excerpted should be directed to the corresponding publishers.

negatively charged aspartate and glutamate residues (Cote et al., 2018).

A detailed quantitative comparison between the UNRES  $\beta$ -sheet propensities (**Figure 8**) and those predicted in Refs. Ullman et al., 2011; Coskuner and Wise-Scira, 2013; Yu et al., 2015; Ramis et al., 2019 is difficult. However, we observe that there is a significant tendency to form  $\beta$ -sheet conformations over the entire sequence in Refs. Ullman et al., 2011; Ramis et al., 2019. On the contrary, simulations of Refs. Coskuner and Wise-Scira, 2013; Yu et al., 2015 predicted much localized and lower  $\beta$ -sheet probabilities.

Although the probabilities represented in **Figure 8** for intra-molecular  $\beta$ -sheets of an isolated monomer cannot be simply compared to the formation of intra-molecular contacts in fibrils, the peaks in UNRES predictions for the  $\beta$ -sheet propensities (**Figure 8**) agree with the pairs of residues forming intra-molecular contacts in protofilaments, i.e., 47–79, 48–78, 92–71, 93–70, 94–69, 95–68 (PDB ID: 2n0a), and 47–79, 48–78, 92–69, 93–68 (PDB ID 6h6b). In addition, the maxima in **Figure 8** for  $\beta$ -sheet propensity are located at or close to valine residues [ $P_{V3} = 0.52$ ,  $P_{V26} = 0.46$ ,  $P_{V40} = 0.57$ ,  $P_{V49} = 0.4$ ,  $P_{V55} = 0.69$ ,  $P_{V63} = 0.64$ ,  $P_{V70} = 0.32$ ,  $P_{V82} = 0.31$ ,  $P_{V95} = 0.35$ ], which is the most frequently found amino acid in  $\beta$ -sheets (Chou and Fasman, 1974). The peak at A90 is in the hydrophobic stretch 88IAAA91. Present simulations show a significant propensity to form a  $\beta$ -sheet in the region 71–83 necessary for aggregation (Giasson et al., 2001). Except for G73, the probability is indeed about 40% in this region (**Figure 8**, bottom panel). However, as mentioned above, in the present simulations, the probability of helical tendency at A78 is also about 40%. The results shown in **Figure 8** are thus compatible with the probability to form both the  $\beta$ -sheet and helix in the region 71–83. The helix can be stabilized in the tetrameric oligomer (Bartels et al., 2011; Wang et al., 2011; Cote et al., 2018).

Although the propensities along the sequence are difficult to compare to experimental data, more global metrics can be used. Indeed, an interesting experimental parameter is the average content in helix and  $\beta$ -sheet conformations of WT  $\alpha$ -syn in solution. Circular dichroism (CD) data reported an average of  $2 \pm 3\%$  and  $11 \pm 7\%$  for helix and  $\beta$ -sheet contents, respectively (Rekas et al., 2010). It is worth noting that the algorithms to extract  $\beta$ -sheet conformations from CD spectra are not as accurate as those for helices (Micsonai et al., 2015) and, as shown for  $\alpha$ -syn, the CD spectrum is also dependent on the buffer and concentration (Araki et al., 2016). Other experimental values extracted from CD for the SSE content of WT monomers were reported:  $< 2\%$  for the helix and  $30\%$  for  $\beta$ -sheet contents in Ref. Weinreb et al., 1996 and  $3 \pm 1\%$  for the helix and  $23 \pm 8\%$  for  $\beta$ -sheet fractions in Ref. Davidson et al., 1998. In their construction of the WT conformational ensemble based on MD simulations of  $\alpha$ -syn fragments using the CHARMM force field constrained by NMR data, Ullman et al. reported the values  $2\%$  for the helix and  $11\%$  for  $\beta$ -sheet fractions (Ullman et al., 2011) (orange curve in **Figure 10**). In the present work, the global proportions of residues in the helix and  $\beta$ -sheet computed from the complete converged conformational WT ensemble are 10 and 31%, respectively. There are about 3 times more residues in the  $\beta$ -sheet than in the helix, a ratio which is in agreement with the one estimated from CD in Ref. Rekas et al., 2010. Interestingly, the theoretical values reported by coarse-grained simulations of WT  $\alpha$ -syn with a different force field (red curve in **Figure 10**) are  $20 \pm 4\%$  and  $26.8 \pm 6.8\%$  for helix and  $\beta$ -sheet fractions, respectively (Ramis et al., 2019). In the present simulations, one notes that half of the helix fraction arises from the short helices located in the C-terminal region. The helix fraction without the C-terminal region is only 5% here. The fractions of the helix and  $\beta$ -sheet are on the same order of

magnitude for the mutants and WT: A30P (helix = 10%,  $\beta$ -sheet = 34%), A53T (helix = 13%,  $\beta$ -sheet = 26%), and E46K (helix = 10%,  $\beta$ -sheet = 29%). Again, A53T shows a larger deviation compared to WT in agreement with a less populated B state (Figures 4 and 5). Most likely, the present force field overestimates the formation of SSE but predicts the correct equilibrium between the two main SSEs.

Another global structural parameter is  $R_g$ . The average radius of gyration  $R_g$  measured by small-angle X-ray scattering for WT  $\alpha$ -syn in solution depends on the protein concentration, the presence of dimers or trimers, buffer type, pH, acetylation, and the source of proteins (Araki et al., 2016). For recombinant  $\alpha$ -syn in 10 mM ammonium acetate pH 7.4,  $R_g = 27.2 \pm 0.44 \text{ \AA}$  extrapolated at infinite dilution, which is comparable to the average computed value ( $R_g = 24.7 \text{ \AA}$ , Figure 7). Addition of HCl and/or NaCl increases  $R_g$  significantly to 33–40  $\text{\AA}$  (Li et al., 2002; Allison et al., 2009; Araki et al., 2016), but these effects cannot be tested with the present model. It is interesting to note that  $R_g$  for dimers in the present simulations are about 10  $\text{\AA}$  larger than that of monomers, as will be discussed elsewhere. Therefore, any mixture of monomers and dimers increases the effective radius of gyration of the solution.

The effects of the missense mutations on the structural properties can be summarized as follows: for the SSE tendency (Figure 8), the most significant effects of the single mutations are for A30P and A53T. Residue P30 has a huge effect on neighboring residues for the formation of the  $\beta$ -sheet. Residue T53 has a great influence on the helical region centered at K58. For the B state, structures with low  $R_g$  (around 18  $\text{\AA}$ ) are more compact for A30P and E46K due to the large average number of contacts between the residues 1–20 and 96–140. For the entire ensemble of conformations, the average number of contacts between these two regions is rather similar for WT (1.84), A30P (2.12), and E46K (1.61) but significantly less for A53T (0.75). The mutant A53T is thus expected to be more flexible and less compact as shown by its larger  $R_g$ . The analysis is slightly different if the average number of contacts is computed between the entire N-terminal (1–60) and C-terminal (96–140) regions for which the average number of contacts is 4.54 (WT), 4.22 (A30P), 2.71 (E46K), and 2.61 (A53T). One finds that both E46K and A53T are much more flexible than WT. NMR studies of mutants A30P and A53T showed a reduction of contacts between C- and N-terminals for both mutants compared to WT (Bertoncini et al., 2005). The present MD results fully agree with these data for A53T.

A major finding in the present work is that an isolated  $\alpha$ -synuclein in solution occurs in two phases, which are clearly

visible in Figures 4 and 5. Unfortunately, it is difficult to construct experimentally a two-dimensional map of SSE propensities such as in Figure 4. A possible road toward such experimental analysis could be the use of Raman single-molecule spectroscopy (Leray et al., 2016; Dai et al., 2021) as  $\alpha$ -helices and  $\beta$ -sheets have been well described by Raman fingerprints (Dai et al., 2021). About 40% of conformations are in the B state for WT, A30P, and E46K and 25% for A53T, and such a significant fraction of the conformational ensemble might be detectable by spectroscopy.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Materials, and further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

PS and GGM designed research. AG and PD performed research. AG, GGM, and PS analyzed the data. AR provided technical assistance with the UNRES package. AG did the figures and graphic representations, and PS wrote the paper. All authors discussed the results and reviewed the manuscript.

## FUNDING

The calculations were performed using HPC resources from DSI-CCuB (Centre de Calcul de l'Université de Bourgogne). This work is part of the project NANO-NEURO-MED (2019–2022) supported by the EIPHI Graduate School (contract ANR-17-EUR-0002), the Conseil Régional de Bourgogne Franche-Comté, and the European Union through the PO FEDER-FSE Bourgogne 2014/2020 programs. This work was supported by a grant from the National Institutes of Health grant no. R01GM14312.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <https://www.frontiersin.org/articles/10.3389/fmolb.2021.786123/full#supplementary-material>

## REFERENCES

- Allison, J. R., Varnai, P., Dobson, C. M., and Vendruscolo, M. (2009). Determination of the Free Energy Landscape of  $\alpha$ -Synuclein Using Spin Label Nuclear Magnetic Resonance Measurements. *J. Am. Chem. Soc.* 131, 18314–18326. doi:10.1021/ja904716h
- Antonschmidt, L., Dervişoğlu, R., Sant, V., Tekwani Movellan, K., Mey, I., Riedel, D., et al. (2021). Insights into the Molecular Mechanism of Amyloid Filament Formation: Segmental Folding of  $\alpha$ -Synuclein on Lipid Membranes. *Sci. Adv.* 7, eabg2174. doi:10.1126/sciadv.abg2174
- Appel-Cresswell, S., Vilarino-Guell, C., Encarnacion, M., Sherman, H., Yu, I., Shah, B., et al. (2013). Alpha-synuclein p.H50Q, a Novel Pathogenic Mutation for Parkinson's Disease. *Mov Disord.* 28, 811–813. doi:10.1002/mds.25421
- Araki, K., Yagi, N., Nakatani, R., Sekiguchi, H., So, M., Yagi, H., et al. (2016). A Small-Angle X-ray Scattering Study of Alpha-Synuclein from Human Red Blood Cells. *Sci. Rep.* 6, 30473. doi:10.1038/srep30473
- Bartels, T., Choi, J. G., and Selkoe, D. J. (2011).  $\alpha$ -Synuclein Occurs Physiologically as a Helically Folded Tetramer that Resists Aggregation. *Nature* 477, 107–110. doi:10.1038/nature10324
- Bertoncini, C. W., Jung, Y.-S., Fernandez, C. O., Hoyer, W., Griesinger, C., Jovin, T. M., et al. (2005). Release of Long-Range Tertiary Interactions Potentiates

- Aggregation of Natively Unstructured  $\alpha$ -Synuclein. *Proc. Natl. Acad. Sci.* 102, 1430–1435. doi:10.1073/pnas.0407146102
- Bhattacharjee, N., and Biswas, P. (2010). Position-Specific Propensities of Amino Acids in the Beta-Strand. *BMC Struct. Biol.* 10, 29. doi:10.1186/1472-6807-10-29
- Bousset, L., Pieri, L., Ruiz-Arlandis, G., Gath, J., Jensen, P. H., Habenstein, B., et al. (2013). Structural and Functional Characterization of Two Alpha-Synuclein Strains. *Nat. Commun.* 4, 2575. doi:10.1038/ncomms3575
- Breydo, L., Wu, J. W., and Uversky, V. N. (2012).  $\alpha$ -Synuclein Misfolding and Parkinson's Disease. *Biochim. Biophys. Acta (Bba) - Mol. Basis Dis.* 1822, 261–285. doi:10.1016/j.bbadis.2011.10.002
- Cascella, R., Chen, S. W., Bigi, A., Camino, J. D., Xu, C. K., Dobson, C. M., et al. (2021). The Release of Toxic Oligomers from  $\alpha$ -Synuclein Fibrils Induces Dysfunction in Neuronal Cells. *Nat. Commun.* 12, 1814. doi:10.1038/s41467-021-21937-3
- Chandonia, J.-M., Fox, N. K., and Brenner, S. E. (2019). SCOPe: Classification of Large Macromolecular Structures in the Structural Classification of Proteins-Extended Database. *Nucleic Acids Res.* 47, D475–D481. doi:10.1093/nar/gky1134
- Cheng, C.-Y., Varkey, J., Ambrosio, M. R., Langen, R., and Han, S. (2013). Hydration Dynamics as an Intrinsic Ruler for Refining Protein Structure at Lipid Membrane Interfaces. *Proc. Natl. Acad. Sci.* 110, 16838–16843. doi:10.1073/pnas.1307678110
- Chiti, F., and Dobson, C. M. (2006). Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* 75, 333–366. doi:10.1146/annurev.biochem.75.101304.123901
- Chiti, F., and Dobson, C. M. (2017). Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu. Rev. Biochem.* 86, 27–68. doi:10.1146/annurev-biochem-061516-045115
- Chou, P. Y., and Fasman, G. D. (1974). Conformational Parameters for Amino Acids in Helical,  $\beta$ -Sheet, and Random Coil Regions Calculated from Proteins. *Biochemistry* 13, 211–222. doi:10.1021/bi00699a001
- Coskuner, O., and Wise-Scira, O. (2013). Structures and Free Energy Landscapes of the A53T Mutant-type  $\alpha$ -Synuclein Protein and Impact of A53T Mutation on the Structures of the Wild-Type  $\alpha$ -Synuclein Protein with Dynamics. *ACS Chem. Neurosci.* 4, 1101–1113. doi:10.1021/cn400041j
- Cote, Y., Delarue, P., Scheraga, H. A., Senet, P., and Maisuradze, G. G. (2018). From a Highly Disordered to a Metastable State: Uncovering Insights of  $\alpha$ -Synuclein. *ACS Chem. Neurosci.* 9, 1051–1065. doi:10.1021/acscchemneuro.7b00446
- Cremades, N., Cohen, S. I. A., Deas, E., Abramov, A. Y., Chen, A. Y., Orte, A., et al. (2012). Direct Observation of the Interconversion of Normal and Toxic Forms of  $\alpha$ -Synuclein. *Cell* 149, 1048–1059. doi:10.1016/j.cell.2012.03.037
- Dai, X., Fu, W., Chi, H., Mesias, V. S. D., Zhu, H., Leung, C. W., et al. (2021). Optical Tweezers-Controlled Hotspot for Sensitive and Reproducible Surface-Enhanced Raman Spectroscopy Characterization of Native Protein Structures. *Nat. Commun.* 12, 1292. doi:10.1038/s41467-021-21543-3
- Davidson, W. S., Jonas, A., Clayton, D. F., and George, J. M. (1998). Stabilization of  $\alpha$ -Synuclein Secondary Structure upon Binding to Synthetic Membranes. *J. Biol. Chem.* 273, 9443–9449. doi:10.1074/jbc.273.16.9443
- Dedmon, M. M., Lindorff-Larsen, K., Christodoulou, J., Vendruscolo, M., and Dobson, C. M. (2005). Mapping Long-Range Interactions in  $\alpha$ -Synuclein Using Spin-Label NMR and Ensemble Molecular Dynamics Simulations. *J. Am. Chem. Soc.* 127, 476–477. doi:10.1021/ja044834j
- Deiana, A., Forcelloni, S., Porrello, A., and Giansanti, A. (2019). Intrinsically Disordered Proteins and Structured Proteins with Intrinsically Disordered Regions Have Different Functional Roles in the Cell. *Plos One* 14, e0217889. doi:10.1371/journal.pone.0217889
- Devi, L., Raghavendran, V., Prabhu, B. M., Avadhani, N. G., and Anandatheerthavarada, H. K. (2008). Mitochondrial Import and Accumulation of  $\alpha$ -Synuclein Impair Complex I in Human Dopaminergic Neuronal Cultures and Parkinson Disease Brain. *J. Biol. Chem.* 283, 9089–9100. doi:10.1074/jbc.M710012200
- Feig, M., Karanicolas, J., and Brooks, C. L. (2004). MMTSB Tool Set: Enhanced Sampling and Multiscale Modeling Methods for Applications in Structural Biology. *J. Mol. Graphics Model.* 22, 377–395. doi:10.1016/j.jmgs.2003.12.005
- Flagmeier, P., Meisl, G., Vendruscolo, M., Knowles, T. P. J., Dobson, C. M., Buell, A. K., et al. (2016). Mutations Associated with Familial Parkinson's Disease Alter the Initiation and Amplification Steps of  $\alpha$ -synuclein Aggregation. *Proc. Natl. Acad. Sci. USA* 113, 10328–10333. doi:10.1073/pnas.1604645113
- Fox, N. K., Brenner, S. E., and Chandonia, J.-M. (2014). SCOPe: Structural Classification of Proteins-Extended, Integrating SCOP and ASTRAL Data and Classification of New Structures. *Nucl. Acids Res.* 42, D304–D309. doi:10.1093/nar/gkt1240
- Fuchs, J., Tichopad, A., Golub, Y., Munz, M., Schweitzer, K. J., Wolf, B., et al. (2008). Genetic Variability in the SNCA Gene Influences  $\alpha$ -Synuclein Levels in the Blood and Brain. *FASEB J.* 22, 1327–1334. doi:10.1096/fj.07-9348com
- Fusco, G., Sanz-Hernandez, M., and De Simone, A. (2018). Order and Disorder in the Physiological Membrane Binding of  $\alpha$ -Synuclein. *Curr. Opin. Struct. Biol.* 48, 49–57. doi:10.1016/j.sbi.2017.09.004
- Galvagnion, C., Buell, A. K., Meisl, G., Michaels, T. C. T., Vendruscolo, M., Knowles, T. P. J., et al. (2015). Lipid Vesicles Trigger  $\alpha$ -Synuclein Aggregation by Stimulating Primary Nucleation. *Nat. Chem. Biol.* 11, 229–234. doi:10.1038/nchembio.1750
- Giasson, B. I., Murray, I. V. J., Trojanowski, J. Q., and Lee, V. M.-Y. (2001). A Hydrophobic Stretch of 12 Amino Acid Residues in the Middle of  $\alpha$ -Synuclein Is Essential for Filament Assembly. *J. Biol. Chem.* 276, 2380–2386. doi:10.1074/jbc.M008919200
- Golaš, E., Maisuradze, G. G., Senet, P., Oldziej, S., Czaplowski, C., Scheraga, H. A., et al. (2012). Simulation of the Opening and Closing of Hsp70 Chaperones by Coarse-Grained Molecular Dynamics. *J. Chem. Theor. Comput.* 8, 1750–1764. doi:10.1021/ct200680g
- Grassein, P., Delarue, P., Nicolai, A., Neiers, F., Scheraga, H. A., Maisuradze, G. G., et al. (2020). Curvature and Torsion of Protein Main Chain as Local Order Parameters of Protein Unfolding. *J. Phys. Chem. B* 124, 4391–4398. doi:10.1021/acs.jpcc.0c01230
- Guerrero-Ferreira, R., Taylor, N. M., Arteni, A. A., Kumari, P., Mona, D., Ringler, P., et al. (2019). Two New Polymorphic Structures of Human Full-Length Alpha-Synuclein Fibrils Solved by Cryo-Electron Microscopy. *eLife* 8, e48907. doi:10.7554/eLife.48907
- Guerrero-Ferreira, R., Kovacic, L., Ni, D., and Stahlberg, H. (2020). New Insights on the Structure of Alpha-Synuclein Fibrils Using Cryo-Electron Microscopy. *Curr. Opin. Neurobiol.* 61, 89–95. doi:10.1016/j.conb.2020.01.014
- Henrich, M. T., Geibl, F. F., Lakshminarasimhan, H., Stegmann, A., Giasson, B. I., Mao, X., et al. (2020). Determinants of Seeding and Spreading of  $\alpha$ -synuclein Pathology in the Brain. *Sci. Adv.* 6, eabc2487. doi:10.1126/sciadv.abc2487
- Hsu, L. J., Sagara, Y., Arroyo, A., Rockenstein, E., Sisk, A., Mallory, M., et al. (2000).  $\alpha$ -Synuclein Promotes Mitochondrial Deficit and Oxidative Stress. *Am. J. Pathol.* 157, 401–410. doi:10.1016/s0002-9440(10)64553-1
- Jakes, R., Spillantini, M. G., and Goedert, M. (1994). Identification of Two Distinct Synucleins from Human Brain. *FEBS Lett.* 345, 27–32. doi:10.1016/0014-5793(94)00395-5
- Kabsch, W., and Sander, C. (1983). Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* 22, 2577–2637. doi:10.1002/bip.360221211
- Khalil, M., Liwo, A., Jagielska, A., and Scheraga, H. A. (2005). Molecular Dynamics with the United-Residue Model of Polypeptide Chains. II. Langevin and Berendsen-Bath Dynamics and Tests on Model  $\alpha$ -Helical Systems. *J. Phys. Chem. B* 109, 13798–13810. doi:10.1021/jp058007w
- Koh, E., Kim, T., and Cho, H.-S. (2006). Mean Curvature as a Major Determinant of Beta-Sheet Propensity. *Bioinformatics* 22, 297–302. doi:10.1093/bioinformatics/bti775
- Konagurthu, A. S., Subramanian, R., Allison, L., Abramson, D., Stuckey, P. J., Garcia de la Banda, M., et al. (2020). Universal Architectural Concepts Underlying Protein Folding Patterns. *Front. Mol. Biosci.* 7, 612920. doi:10.3389/fmolb.2020.612920
- Krüger, R., Kuhn, W., Müller, T., Woitalla, D., Graeber, M., Kösel, S., et al. (1998). AlaSOPro Mutation in the Gene Encoding  $\alpha$ -synuclein in Parkinson's Disease. *Nat. Genet.* 18, 106–108. doi:10.1038/ng0298-106
- Kumar, S., Jangir, D. K., Kumar, R., Kumari, M., Bhavesh, N. S., and Maiti, T. K. (2018). Role of Sporadic Parkinson Disease Associated Mutations A18T and A29S in Enhanced  $\alpha$ -Synuclein Fibrillation and Cytotoxicity. *ACS Chem. Neurosci.* 9, 230–240. doi:10.1021/acscchemneuro.6b00430
- Kumari, P., Ghosh, D., Vanas, A., Fleischmann, Y., Wiegand, T., Jeschke, G., et al. (2021). Structural Insights into  $\alpha$ -synuclein Monomer-Fibril Interactions. *Proc. Natl. Acad. Sci. USA* 118, e2012171118. doi:10.1073/pnas.2012171118

- Labesse, G., Colloc'h, N., Pothier, J., and Mornon, J.-P. (1997). P-SEA: A New Efficient Assignment of Secondary Structure from C $\alpha$  Trace of Proteins. *Bioinformatics* 13, 291–295. doi:10.1093/bioinformatics/13.3.291
- Lashuel, H. A. (2020). Do Lewy Bodies Contain Alpha-Synuclein Fibrils? and Does it Matter? A Brief History and Critical Analysis of Recent Reports. *Neurobiol. Dis.* 141, 104876. doi:10.1016/j.nbd.2020.104876
- Leray, A., Brulé, T., Buret, M., Colas des Francs, G., Bouhelier, A., Dereux, A., et al. (2016). Sorting of Single Biomolecules Based on Fourier Polar Representation of Surface Enhanced Raman Spectra. *Sci. Rep.* 6, 20383. doi:10.1038/srep20383
- Li, J., Uversky, V. N., and Fink, A. L. (2002). Conformational Behavior of Human  $\alpha$ -Synuclein Is Modulated by Familial Parkinson's Disease Point Mutations A30P and A53T. *Neuro.Toxicol.* 23, 553–567. doi:10.1016/S0161-813X(02)00066-9
- Liwo, A., Czaplewski, C., Pillardy, J., and Scheraga, H. A. (2001). Cumulant-based Expressions for the Multibody Terms for the Correlation between Local and Electrostatic Interactions in the United-Residue Force Field. *J. Chem. Phys.* 115, 2323–2347. doi:10.1063/1.1383989
- Liwo, A., Sieradzian, A. K., Lipska, A. G., Czaplewski, C., Joung, I., Żmudzka, W., et al. (2019). A General Method for the Derivation of the Functional Forms of the Effective Energy Terms in Coarse-Grained Energy Functions of Polymers. III. Determination of Scale-Consistent Backbone-Local and Correlation Potentials in the UNRES Force Field and Force-Field Calibration and Validation. *J. Chem. Phys.* 150, 155104. doi:10.1063/1.5093015
- Maisuradze, G. G., Senet, P., Czaplewski, C., Liwo, A., and Scheraga, H. A. (2010). Investigation of Protein Folding by Coarse-Grained Molecular Dynamics with the UNRES Force Field. *J. Phys. Chem. A* 114, 4471–4485. doi:10.1021/jp9117776
- Miconai, A., Wien, F., Kerna, L., Lee, Y.-H., Goto, Y., Réfrégiers, M., et al. (2015). Accurate Secondary Structure Prediction and Fold Recognition for Circular Dichroism Spectroscopy. *Proc. Natl. Acad. Sci. USA* 112, E3095–E3103. doi:10.1073/pnas.1500851112
- Mollenhauer, B., Cullen, V., Kahn, I., Krastins, B., Outeiro, T. F., Pepivani, I., et al. (2008). Direct Quantification of CSF  $\alpha$ -Synuclein by ELISA and First Cross-Sectional Study in Patients with Neurodegeneration. *Exp. Neurol.* 213, 315–325. doi:10.1016/j.expneurol.2008.06.004
- Mor, D. E., Ugras, S. E., Daniels, M. J., and Ischiropoulos, H. (2016). Dynamic Structural Flexibility of  $\alpha$ -Synuclein. *Neurobiol. Dis.* 88, 66–74. doi:10.1016/j.nbd.2015.12.018
- Murray, I. V. J., Giasson, B. I., Quinn, S. M., Koppaka, V., Axelsen, P. H., Ischiropoulos, H., et al. (2003). Role of  $\alpha$ -Synuclein Carboxy-Terminus on Fibril Formation *In Vitro*. *Biochemistry* 42, 8530–8540. doi:10.1021/bi027363r
- Pace, C. N., and Scholtz, J. M. (1998). A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins. *Biophys. J.* 75, 422–427. doi:10.1016/s0006-3495(98)77529-0
- Pasanen, P., Myllykangas, L., Siitonen, M., Raunio, A., Kaakkola, S., Lyytinen, J., et al. (2014). A Novel  $\alpha$ -Synuclein Mutation A53E Associated with Atypical Multiple System Atrophy and Parkinson's Disease-type Pathology. *Neurobiol. Aging* 35, e1–2180. doi:10.1016/j.neurobiolaging.2014.03.024
- Perrin, R. J., Woods, W. S., Clayton, D. F., and George, J. M. (2000). Interaction of Human  $\alpha$ -Synuclein and Parkinson's Disease Variants with Phospholipids. *J. Biol. Chem.* 275, 34393–34398. doi:10.1074/jbc.M004851200
- Petrucci, S., Ginevrino, M., and Valente, E. M. (2016). Phenotypic Spectrum of Alpha-Synuclein Mutations: New Insights from Patients and Cellular Models. *Parkinsonism Relat. Disord.* 22, S16–S20. doi:10.1016/j.parkreldis.2015.08.015
- Polymeropoulos, M. H., Lavedan, C., Leroy, E., Ide, S. E., Dehejia, A., Dutra, A., et al. (1997). Mutation in the  $\alpha$ -Synuclein Gene Identified in Families with Parkinson's Disease. *Science* 276, 2045–2047. doi:10.1126/science.276.5321.2045
- Ramis, R., Ortega-Castro, J., Casasnovas, R., Mariño, L., Vilanova, B., Adrover, M., et al. (2019). A Coarse-Grained Molecular Dynamics Approach to the Study of the Intrinsically Disordered Protein  $\alpha$ -Synuclein. *J. Chem. Inf. Model.* 59, 1458–1471. doi:10.1021/acs.jcim.8b00921
- Ray, S., Singh, N., Kumar, R., Patel, K., Pandey, S., Datta, D., et al. (2020).  $\alpha$ -Synuclein Aggregation Nucleates through Liquid-Liquid Phase Separation. *Nat. Chem.* 12, 705–716. doi:10.1038/s41557-020-0465-9
- Rekas, A., Knott, R. B., Sokolova, A., Barnham, K. J., Perez, K. A., Masters, C. L., et al. (2010). The Structure of Dopamine Induced  $\alpha$ -Synuclein Oligomers. *Eur. Biophys. J.* 39, 1407–1419. doi:10.1007/s00249-010-0595-x
- Rotkiewicz, P., and Skolnick, J. (2008). Fast Procedure for Reconstruction of Full-Atom Protein Models from Reduced Representations. *J. Comput. Chem.* 29, 1460–1465. doi:10.1002/jcc.20906
- Sang, J. C., Hidari, E., Meisl, G., Ranasinghe, R. T., Spillantini, M. G., and Klenerman, D. (2021). Super-Resolution Imaging Reveals  $\alpha$ -Synuclein Seeded Aggregation in SH-SY5Y Cells. *Commun. Biol.* 4, 1–11. doi:10.1038/s42003-021-02126-w
- Schrödinger, L. L. C. (2015). *The PyMOL Molecular Graphics System*. version 1.8. Available at <https://pymol.org/2/support.html?>
- Sieradzian, A. K., Korneev, A., Begun, A., Kachlishvili, K., Scheraga, H. A., Molochkov, A., et al. (2021). Investigation of Phosphorylation-Induced Folding of an Intrinsically Disordered Protein by Coarse-Grained Molecular Dynamics. *J. Chem. Theor. Comput.* 17, 3203–3220. doi:10.1021/acs.jctc.1c00155
- Smith, C. K., Withka, J. M., and Regan, L. (1994). A Thermodynamic Scale for the Beta-Sheet Forming Tendencies of the Amino Acids. *Biochemistry* 33, 5510–5517. doi:10.1021/bi00184a020
- Soto, C. (2003). Unfolding the Role of Protein Misfolding in Neurodegenerative Diseases. *Nat. Rev. Neurosci.* 4, 49–60. doi:10.1038/nrn1007
- Spillantini, M. G., Schmidt, M. L., Lee, V. M.-Y., Trojanowski, J. Q., Jakes, R., and Goedert, M. (1997).  $\alpha$ -Synuclein in Lewy Bodies. *Nature* 388, 839–840. doi:10.1038/42166
- Stefanis, L. (2012).  $\alpha$ -Synuclein in Parkinson's Disease. *Cold Spring Harbor Perspect. Med.* 2, a009399. doi:10.1101/cshperspect.a009399
- Stephens, A. D., Zacharopoulou, M., Moons, R., Fusco, G., Seetaloo, N., Chiki, A., et al. (2020). Extent of N-Terminus Exposure of Monomeric Alpha-Synuclein Determines its Aggregation Propensity. *Nat. Commun.* 11, 2820. doi:10.1038/s41467-020-16564-3
- Suzuki, M., Sango, K., Wada, K., and Nagai, Y. (2018). Pathological Role of Lipid Interaction with  $\alpha$ -Synuclein in Parkinson's Disease. *Neurochem. Int.* 119, 97–106. doi:10.1016/j.neuint.2017.12.014
- Tanudjojo, B., Shaikh, S. S., Fenyi, A., Bousset, L., Agarwal, D., Marsh, J., et al. (2021). Phenotypic Manifestation of  $\alpha$ -synuclein Strains Derived from Parkinson's Disease and Multiple System Atrophy in Human Dopaminergic Neurons. *Nat. Commun.* 12, 3817. doi:10.1038/s41467-021-23682-z
- Tosatto, L., Horrocks, M. H., Dear, A. J., Knowles, T. P. J., Dalla Serra, M., Cremades, N., et al. (2015). Single-molecule FRET Studies on Alpha-Synuclein Oligomerization of Parkinson's Disease Genetically Related Mutants. *Sci. Rep.* 5, 16696. doi:10.1038/srep16696
- Touw, W. G., Baakman, C., Black, J., te Beek, T. A. H., Krieger, E., Joosten, R. P., et al. (2015). A Series of PDB-Related Databanks for Everyday Needs. *Nucleic Acids Res.* 43, D364–D368. doi:10.1093/nar/gku1028
- Trinkaus, V. A., Riera-Tur, I., Martínez-Sánchez, A., Bäuerlein, F. J. B., Guo, Q., Arzberger, T., et al. (2021). *In Situ* Architecture of Neuronal  $\alpha$ -Synuclein Inclusions. *Nat. Commun.* 12, 2110. doi:10.1038/s41467-021-22108-0
- Tuttle, M. D., Comellas, G., Nieuwkoop, A. J., Covell, D. J., Berthold, D. A., Kloepper, K. D., et al. (2016). Solid-State NMR Structure of a Pathogenic Fibril of Full-Length Human  $\alpha$ -synuclein. *Nat. Struct. Mol. Biol.* 23, 409–415. doi:10.1038/nsmb.3194
- Ueda, K., Fukushima, H., Masliah, E., Xia, Y., Iwai, A., Yoshimoto, M., et al. (1993). Molecular Cloning of cDNA Encoding an Unrecognized Component of Amyloid in Alzheimer Disease. *Proc. Natl. Acad. Sci.* 90, 11282–11286. doi:10.1073/pnas.90.23.11282
- Ullman, O., Fisher, C. K., and Stultz, C. M. (2011). Explaining the Structural Plasticity of  $\alpha$ -Synuclein. *J. Am. Chem. Soc.* 133, 19536–19546. doi:10.1021/ja208657z
- Ulmer, T. S., Bax, A., Cole, N. B., and Nussbaum, R. L. (2005). Structure and Dynamics of Micelle-Bound Human  $\alpha$ -Synuclein. *J. Biol. Chem.* 280, 9595–9603. doi:10.1074/jbc.M411805200
- Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000). Why Are “natively Unfolded” Proteins Unstructured under Physiological Conditions? *Proteins* 41, 415–427. doi:10.1002/1097-0134(20001115)41:3<415::aid-prot130>3.0.co;2-7
- Uversky, V. N. (2019). Intrinsically Disordered Proteins and Their “Mysterious” (Meta)Physics. *Front. Phys.* 7. doi:10.3389/fphy.2019.00010
- Victor, J. M., Imbert, J. B., and Lhuillier, D. (1994). The Number of Contacts in a Self-Avoiding Walk of Variable Radius of Gyration in Two and Three Dimensions. *J. Chem. Phys.* 100, 5372–5377. doi:10.1063/1.467151

- Wang, W., Perovic, I., Chittiluru, J., Kaganovich, A., Nguyen, L. T. T., Liao, J., et al. (2011). A Soluble  $\alpha$ -synuclein Construct Forms a Dynamic Tetramer. *Proc. Natl. Acad. Sci.* 108, 17797–17802. doi:10.1073/pnas.1113260108
- Weinreb, P. H., Zhen, W., Poon, A. W., Conway, K. A., and Lansbury, P. T. (1996). NACP, A Protein Implicated in Alzheimer's Disease and Learning, Is Natively Unfolded. *Biochemistry* 35, 13709–13715. doi:10.1021/bi961799n
- Wright, P. E., and Dyson, H. J. (1999). Intrinsically Unstructured Proteins: Re-assessing the Protein Structure-Function Paradigm. *J. Mol. Biol.* 293, 321–331. doi:10.1006/jmbi.1999.3110
- Yu, H., Han, W., Ma, W., and Schulten, K. (2015). Transient  $\beta$ -hairpin Formation in  $\alpha$ -synuclein Monomer Revealed by Coarse-Grained Molecular Dynamics Simulation. *J. Chem. Phys.* 143, 243142. doi:10.1063/1.4936910
- Zarranz, J. J., Alegre, J., Gómez-Esteban, J. C., Lezcano, E., Ros, R., Ampuero, I., et al. (2004). The New Mutation, E46K, of  $\alpha$ -synuclein Causes Parkinson and Lewy Body Dementia. *Ann. Neurol.* 55, 164–173. doi:10.1002/ana.10795
- Zhou, R., Maisuradze, G. G., Suñol, D., Todorovski, T., Macias, M. J., Xiao, Y., et al. (2014). Folding Kinetics of WW Domains with the United Residue Force Field

for Bridging Microscopic Motions and Experimental Measurements. *Proc. Natl. Acad. Sci. USA.* 111, 18243–18248. doi:10.1073/pnas.1420914111

**Conflict of Interest:** Author AR is a employee of Schrödinger, Inc. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Guzzo, Delarue, Rojas, Nicolai, Maisuradze and Senet. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Chapitre 4

## Résultats : Etude globale des dimères d' $\alpha$ -synucléine

Le second article issu de cette thèse et soumis pour publication dans *Frontiers in molecular biology*, "Wild-type  $\alpha$ -synuclein and variants occur in different disordered dimers and pre-fibrillar conformations in early stage of aggregation", traite de l'analyse des dimères présents dans notre ensemble de structures. Nous avons pu mettre en évidence la présence de dimères réalisant des CTF appelés dimères type fibrilles (DTF). Ces dimères ne représentent que 9 à 15% des conformations de dimères présents dans notre ensemble, le reste étant désordonné. Deux régions principales d'agrégation ont été identifiées sur les segments 36-55 et 66-95. La formation de dimères a grandement déstabilisé la formation d'hélices et de feuillets  $\beta$  au sein de chaque monomère comparé à leur comportement seul présenté dans la précédente publication. La formation de CTF se produit majoritairement dans le Nterm pour A53T et E46K tandis que le WT et A30P favorisent plutôt la fin du NAC. Ces résultats se corrélaient fortement avec une des études réalisées par AFM. Les DTF peuvent être reliés aux oligomères toxiques de type B observés par FRET. Les dimères présentent un bien plus grand nombre de feuillets  $\beta$  parallèles que les monomères et notamment les DTF. Ce résultat s'accorde avec les observations de spectroscopie infrarouge à transformée de Fourier. Enfin l'étude de la distribution du rayon de giration des dimères a permis de montrer une grande hétérogénéité de conformations avec plusieurs pics et notamment une sous population plus compacte pour A30P.



# Wild-type $\alpha$ -synuclein and variants occur in different disordered dimers and pre-fibrillar conformations in early stage of aggregation

Adrien Guzzo<sup>1</sup>, Patrice Delarue<sup>1</sup>, Ana Rojas<sup>2</sup>, Adrien Nicolai<sup>1</sup>, Gia G. Maisuradze<sup>3</sup> and Patrick Senet<sup>1,3\*</sup>

<sup>1</sup>Laboratoire Interdisciplinaire Carnot de Bourgogne, UMR 6303 CNRS-Université de Bourgogne Franche-Comté, Dijon, France

<sup>2</sup>Schrödinger, Inc., New York, USA

<sup>3</sup>Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, USA

Correspondence\*:

Patrick Senet: Laboratoire Interdisciplinaire Carnot de Bourgogne, UMR 6303 CNRS-Université de Bourgogne Franche-Comté, 9 Avenue Alain Savary, BP 47870, 21078 Dijon Cedex, France  
psenet@u-bourgogne.fr

## 2 ABSTRACT

3  $\alpha$ -synuclein is a 140 amino-acid intrinsically disordered protein mainly found in the brain. Toxic  
4  $\alpha$ -synuclein aggregates are the molecular hallmarks of the Parkinson disease. In vitro studies  
5 showed that  $\alpha$ -synuclein aggregates in oligomeric structures of several tenth of monomers and  
6 into cylindrical structures (fibrils), comprising hundred to thousands proteins, with polymorphic  
7 cross- $\beta$ -sheet conformations. Oligomeric species, formed at the early stage of aggregation  
8 remain, however, poorly understood and are hypothesized to be the most toxic aggregates.  
9 Here, we studied the formation of wild-type (WT) and mutants (A30P, A53T, and E46K) dimers of  
10  $\alpha$ -synuclein using coarse-grained molecular dynamics. We identified two principal segments of  
11 the sequence with an higher propensity to aggregate in the early stage of dimerization: residues  
12 36-55 and residues 66-95. The transient  $\alpha$ -helices (residues 53-65 and 73-82) of  $\alpha$ -synuclein  
13 monomers are destabilized by A53T and E46K mutations which favors the formation of fibril  
14 native contacts in the N-terminal region, whereas the helix 53-65 prevents the propagation of fibril  
15 native contacts along the sequence for the WT in the early stages of dimerization. The present  
16 results indicate that dimers do not adopt the Greek-key motif of the monomer fold in fibrils but  
17 form a majority of disordered aggregates and a minority (9-15 %) of pre-fibrillar dimers both with  
18 intra-molecular and inter-molecular  $\beta$ -sheets. The percentage of residues in parallel  $\beta$ -sheets is  
19 by increasing order monomer < disordered dimers < pre-fibrillar dimers. Native fibril contacts  
20 between the two monomers are present in the NAC domain for WT, A30P and A53T and in the  
21 N-domain for A53T and E46K. Structural properties of pre-fibrillar dimers agree with rupture-force  
22 Atomic Force Microscopy and single-molecule Förster Resonance Energy Transfer available data.  
23 This suggests that the pre-fibrillar dimers might correspond to the smallest type B toxic oligomers.  
24 The probability density of the dimer gyration radius is multi-peaks with an average radius which is  
25 10 Å larger than the one of the monomers for all proteins. Present results indicate that even the  
26 elementary  $\alpha$ -synuclein aggregation step, the dimerization, is a complicated phenomena which  
27 does not only involved the NAC region.

28 **Keywords:**  $\alpha$ -synuclein, amyloid, Parkinson disease, molecular dynamics, dimers, CUTABI

## 1 INTRODUCTION

29  $\alpha$ -synuclein ( $\alpha$ -syn) is a soluble 140 amino-acid intrinsically disordered protein (IDP)[1] abundant in the  
30 brain[2, 3] the function of which remains unclear. Abnormal aggregation of  $\alpha$ -syn is central to onset of  
31 diseases, and its inhibition is an intensive research area[4]. In synucleopathies[5, 6, 7, 8, 9], as the Parkinson  
32 disease (PD),  $\alpha$ -syn is found in high concentration, as filamentous aggregates, in intraneuronal inclusions  
33 (Lewy Bodies) and in extracellular deposits (Lewy neurites)[10, 11, 12, 13]. In addition to sporadic PD,  
34 rare familial cases of PD with similar phenotypes are induced either by an overexpression of wild-type  
35 (WT)  $\alpha$ -syn due to  $\alpha$ -syn gene triplication or by pathogenic mutations in  $\alpha$ -syn gene corresponding to  
36 single amino-acid substitution[14, 15, 16, 17, 18, 19, 20].

37 The origin of aggregation of  $\alpha$ -syn is unknown. Formation of  $\alpha$ -syn amyloid shares many common  
38 properties with the toxic aggregation of other IDPs, as for example the aggregation of A $\beta$  in the Alzheimer's  
39 disease[21, 22, 23]. When incubated in physiological conditions *in vitro*,  $\alpha$ -syn aggregates spontaneously  
40 into fibrils with polymorphic cross- $\beta$ -sheet conformations, in which a core of  $\beta$ -strands is aligned perpen-  
41 dicularly to the fibril axis forming extended regular  $\beta$ -sheets with different arrangements[24, 25, 26]. These  
42 different *in vitro* regular  $\beta$ -sheet structures are believed to mimic the  $\alpha$ -syn aggregates in synucleopathies.  
43 Post-mortem X-ray diffraction of brains from patients with PD reveal patterns of  $\beta$ -sheet arrangements[27].  
44 The first *ex vivo* structures of  $\alpha$ -syn fibrils from patients with Multiple System Atrophy have topology  
45 similar to the recombinant *in vitro* fibrils but with structural differences and occur with non-proteinaceous  
46 moieties[28]. In all known structures of fibrils, the monomer chain folds according to a three-layered  
47 L-shaped motif bearing resemblance to a Greek-key motif[25]. New toxic polymorphs were recently  
48 detected for which no structures are yet available[29].

49 A recent study using solid-state nuclear magnetic resonance (NMR) of WT  $\alpha$ -syn aggregation, on  
50 phospholipid small unilamellar vesicles, identified pre-fibrillar and early fibrillar species[30]. NMR data  
51 suggest a pre-fibrillar segmental folding of the  $\alpha$ -syn monomer in a three-layered L-shaped motif resembling  
52 to the one of mature fibrils. The folding occurs with about the first half of the residues of the N-terminal  
53 region of the protein bounded to the phospholipid in an  $\alpha$ -helical conformation[30]. Another study of  $\alpha$ -syn  
54 aggregation in solution identifies helix-rich intermediates in the transition of unstructured conformation  
55  $\alpha$ -syn into  $\beta$ -sheet rich fibril formation during the growth of fibrils (elongation phase)[31]. Interestingly,  
56  $\alpha$ -syn may also form ribbon aggregates[32].

57  $\alpha$ -syn fibrils are composed of thousands of  $\alpha$ -syn monomers[33]. Kinetic studies of  $\alpha$ -syn aggregation  
58 showed that the onset of fibril formation is preceded by a long lag phase indicating that the initial rate  
59 of aggregation is controlled by the nucleation of fibrils. It is hypothesized that pre-fibrillar aggregates  
60 of  $\alpha$ -syn,  $\alpha$ -syn protofibrils, i.e. oligomers composed of several dozen monomers occurring in the lag  
61 phase, might be the most toxic aggregates[34, 35, 36]. These  $\alpha$ -syn oligomers are transient intermediate  
62 species with heterogeneous structures which are consumed as formed fibrils[22]. Structural properties  
63 of  $\alpha$ -syn oligomers were observed by different techniques, including single-molecule Förster Resonance  
64 Energy Transfer (FRET), atomic force microscopies (AFM), small angle x-ray scattering (SAXS), small  
65 angle neutron scattering (SANS), transmission electron microscopy (TEM) and cryo-electron microscopy  
66 (cryo-EM)[35, 37, 38, 39, 40, 41, 42, 22, 43, 44, 45].

67 Single-molecule FRET provides information on the distance probability distribution between two fluo-  
68 rophores covalent linked to  $\alpha$ -synuclein molecule. FRET study of pre-fibrillar oligomerization of  $\alpha$ -syn

69 (oligomers composed of 2 to 150 monomers) revealed two types of oligomers: type A with a low FRET effi-  
70 ciency (large internal distances) and type B with a high FRET efficiency (low internal distances)[41, 46, 42].  
71 Type A occurred the first in the kinetic of aggregation. Primary nucleation of  $\alpha$ -syn may result in the creation  
72 of oligomers of type A from monomeric protein molecules which can grow through monomer addition, but  
73 they can also convert into type-B oligomers[41]. The type B oligomers are toxic, more compact than the  
74 initially formed type A oligomers, and more proteinase-K-resistant than the type A and the monomer[41].  
75 Unlike type B, type A oligomers dissociate in distilled water[46]. Single-molecule FRET showed that  
76 missense mutations (A30P, A53T and E46K) observed in familial forms of PD influence the kinetics of  
77 formation of oligomers and induce structural differences compared with the WT[42]. The concentration of  
78 oligomeric species of variants (A30P, A53T and E46K) in the lag phase of fibril formation is similar to the  
79 one of WT which indicates that the effects of the single amino acid substitutions related on the oligomer  
80 structure might be more relevant for their neurotoxicity[42]. Both types A and B oligomers are observed  
81 for mutants with a conversion rate from A to B higher for A53T and lower for A30P compared to WT[42].  
82 Interestingly, for E46K mutant, it is not possible to separate clearly the two types of oligomeric species.  
83 Electron microscopy analysis of  $\alpha$ -syn aggregation showed that A30P mutant promotes the formation of  
84 pore-like protofibrils, whereas A53T promotes formation of annular and tubular protofibrillar structures[47].  
85 The E46K variant generates annular oligomers similar to those observed for the A53T and A30P mutants  
86 [38].

87 Spherical metastable oligomers, with no significant secondary structures, with a diameter of 100 Å ,  
88 were identified by SANS[45]. Oligomers composed of about 30 monomers forming a compact core  
89 with a flexible shell were characterized by SAXS[40]. These ellipsoidal oligomers, of the order of 100  
90 Å size, inhibit both the primary nucleation and the subsequent elongation steps of  $\alpha$ -syn fibrils[40]. The  
91 structure of two main sub-groups of small oligomers, composed of 18 and 29 monomers in average, with  
92 a pore like shape were identified by cryo-EM[48]. The  $\beta$ -sheet structure in these oligomeric species is  
93 predominantly antiparallel (in opposite to the parallel  $\beta$ -sheet structure of fibrils) and amounts to 35%  
94 of the residues. Interestingly, spherical oligomers observed by AFM-IR are also mainly composed of  
95 anti-parallel  $\beta$ -sheet[44]. AFM studies of  $\alpha$ -syn pre-fibrillar formation on different solid surfaces reported  
96 the observation of heterogenous oligomers with spherical, elongated and annular shapes [35, 44]. Chains  
97 of spherical oligomers were also observed[35]. Oligomers are formed in the primary nucleation phase of  
98 fibril formation but can be also generated upon fibril disaggregation[32, 22].

99 Observation of annular oligomers for WT and mutants suggests that  $\alpha$ -syn may form a pore-like  $\beta$ -barrel  
100 structure in membranes which disrupt the membrane integrity[37, 38, 49]. Monomeric  $\alpha$ -syn and fibrils did  
101 not show membrane permeabilization activity[37]. Exposure of hydrophobic side chains of oligomers to  
102 solvent may facilitate membrane permeabilization and be a source of cellular dysfunction[41]. Excessive  
103 production of reactive oxygen species, which may lead to cell apoptosis, is induced by  $\alpha$ -syn oligomers  
104 [41]. The selective binding between the transmembrane protein LAG3 (lymphocyte-activation gene 3)  
105 and  $\alpha$ -syn preformed fibril species is also suggested as a mechanism of neurotoxicity in which LAG3  
106 facilitates the endocytosis of toxic oligomeric species and the neuron-to-neuron transmission of pathological  
107  $\alpha$ -synuclein aggregates[50]. In addition, observation of  $\alpha$ -syn helical tetramer[51, 52, 53, 54, 55] stabilized  
108 by hydrophobic interactions and salt bridges between the monomers[55] suggests another mechanism of  
109 toxicity related to the equilibrium between the tetrameric form and other ( $\beta$ -sheet) oligomeric species[52].  
110 Replica exchange Molecular Dynamics (MD) biased by NMR data showed that helix rich and  $\beta$ -strand  
111 rich trimers and tetramers are stable and may represent a minor population of  $\alpha$ -syn in solution[56].

112 Although oligomers composed of dozens of  $\alpha$ -syn monomers are identified as toxic (type B oligomers),  
113 the role of dimers in the neurotoxicity of  $\alpha$ -syn cannot be discarded. It has been hypothesized that the  
114 critical rate-limiting step in the primary nucleation is the oxidative formation and accumulation of a  
115 dityrosine cross-linked dimer[57]. The peak of accumulation of dimers coincides with the rapid onset  
116 of fibrillation for WT and mutants (A53T and A30P) proteins. Dimer formation is accelerated for the  
117 A30P and A53T variants[57]. It is worth noting that dimers have a detectable membrane permeabilization  
118 activity[38]. As dimerization is the most initial step in self-assembling of monomers, dimers may play  
119 an important role in the different pathways of aggregation leading to type B oligomers.  $\alpha$ -syn dimers are  
120 systematically observed in the lag phase of  $\alpha$ -syn aggregation and were characterized by FRET[41, 46, 42],  
121 fluorescence[38, 58, 57], AFM[59, 60] and circular dichroism[61].

122 Dimerization between fluorophore-free WT and mutants (A30P, A53T and E46K)  $\alpha$ -syn monomers on a  
123 substrate and fluorophore-labeled monomers in solution was analyzed by internal reflection fluorescence  
124 microscopy[58]. Two types of dimers were identified with differ by an order of magnitude in their lifetimes.  
125 For WT, E46K and A53T, the less stable (type 1) and most stable (type 2) dimers have a lifetime of the  
126 order of about 200-300 ms and about 3s, respectively[58]. The variant A30P is significantly different with  
127 lifetimes of the order of 700 ms (type 1) and 5 s (type 2), suggesting a more stable dimeric structure[58].  
128 Type 2 dimers were also observed in single-molecule AFM spectroscopy findings of WT  $\alpha$ -syn[59]. The  
129 lifetimes (types 1 and 2) of the mutants are systematically longer than the one of WT, which indicates that  
130 the missense mutations seem to increase the stability of the dimer aggregate. Most dimeric structure are of  
131 type 1 but A30P and A53T single amino-acid substitutions increase significantly the type 2 population. The  
132 type 2 dimer might be related to B type toxic oligomer[42]. The  $\alpha$ -syn dimers have a low FRET efficiency  
133 but might be a mixture of type A and B which cannot be resolved experimentally[46]. The fraction of  
134 dimers formed in the lag phase of aggregation is by increasing order WT < A53T < A30P as measured  
135 by FRET. A30P variant generates the largest population of B oligomers at the end of the lag phase of  
136 fibrillation [42]. These data[42, 58] taken together indicate a significant structural difference between A30P  
137 dimer compared to WT and A53T dimers.

138 Differences between WT and mutants dimers were also observed by single molecule AFM force spec-  
139 troscopy at low pH (promoting aggregation)[62]. Multiple segments interactions between the monomers  
140 are favored in A53T and E46K compared to WT as reflected by a larger number of multiple rupture force  
141 events[62] whereas A30P favored single segment dimerization compared to WT[62].  $\alpha$ -syn monomer and  
142 dimer structures and dynamics were also measured at neutral pH by high-speed AFM for proteins adsorbed  
143 on a solid surface immersed in a aqueous solution[60]. The monomer adopts mainly a compact and stable  
144 spherical structure but one tail and two tails transient structures were also identified[60]. Conformational  
145 transitions between different conformations occurred on second to minute time-scale. Dimers formed  
146 by association of two globular monomers (major population) or of one globular monomer and one-tail  
147 monomer (minor population) were observed[60]. However,  $\alpha$ -syn structure and dynamics on a surface  
148 might be different from those of  $\alpha$ -syn in solution, as the confinement of a polymer on a surface is known  
149 to modify the dynamics and stability of the polymer conformations. For example,  $\alpha$ -syn monomer adopts  
150 an  $\alpha$ -helical structure on the negatively charged (phospholipid) membrane[63].

151 No experimental technique so far provided atomistic description of the dimer heterogeneity. Detailed  
152 structural information on the influence of the missense mutations on the fundamental dimerization step  
153 of  $\alpha$ -synuclein is missing. The structural properties and the role of  $\alpha$ -syn dimers in the different parallel  
154 pathways of the formation of larger oligomers are still unclear. MD simulation is a complementary means  
155 to single-molecule experimental techniques which may give insights on the dimerization process. Previous

156 simulations of  $\alpha$ -synuclein dimers were limited to ultra-simplified model (discrete MD[60] or small  
157 fragments of  $\alpha$ -synuclein[64]) which does not take into account properly the dynamics of the polymer or  
158 biased the initial conditions of the simulations by using docking of monomer native structure adopted by  
159  $\alpha$ -synuclein on a phospholipid membrane[65]. Here, we are going a step further by using unbiased replica  
160 exchange MD simulations of two  $\alpha$ -synuclein molecules in implicit solvent by using a physics-based  
161 coarse-grained UNited-RESidue (UNRES) force field[66, 67] on a time-scale of 29.7 milliseconds (72  
162 replicas of 412  $\mu$ s each for each variant studied) which is three order of magnitude larger than typical  
163 all-atom MD simulations[68]. The force field was calibrated to reproduce the structure and thermodynamics  
164 of small model proteins and applied with success to simulate protein folding[66, 69, 70] and large scale  
165 conformational dynamics[71]. In the present MD simulations, most of the  $\alpha$ -syn molecules do not aggregate  
166 and remain thus in a monomeric conformation. These monomers for WT, A30P, A53T and E46K were  
167 recently described elsewhere[72]. Here, we describe the structure and populations of the different dimeric  
168 states found in the MD simulations for WT and the same mutants and compare to data extracted from  
169 experiments.

## 2 MATERIALS AND METHODS

170 All structures of  $\alpha$ -syn (WT and A30P, A53T and E46K variants) were extracted from replica exchange  
171 MD trajectories generated with the coarse-grained UNRES force field[66, 67]. In the UNRES force field, a  
172 polypeptide chain is represented as a sequence of  $C^\alpha$  atoms with united peptide groups located halfway of  
173 the virtual  $C^\alpha$ - $C^\alpha$  bonds and united side chains (SC) attached to the  $C^\alpha$  atoms. The SC-SC interaction  
174 potentials implicitly include the contribution from solvation[66, 73]. Descriptions of the UNRES force  
175 field and its parametrization are available in the reference[67] and at <http://www.unres.pl>.

176 A total of 72 trajectories were computed for *each* protein: 32 trajectories at 300 K and 8 trajectories at  
177 each of the following temperatures, 310 K, 323 K, 337 K, 353 K, 370 K using replica exchange MD as  
178 described in Ref. [72]. Each trajectory was started with 2 fully-unfolded monomers separated by a distance  
179 of 25 Å. The integration time step in UNRES is 4.9 fs corresponding to an effective actual time step of  
180 about 4.9 ps[68]. The convergence of each trajectory was monitored by computing the probability density  
181 of contacts of each residue resulting in 30 million steps (effective time scale of 147 microseconds) of  
182 consolidated data out of 84 million steps (effective time scale of 412 microseconds) of simulation for *each*  
183 trajectory. The Cartesian coordinates of  $C^\alpha$  and SC beads were saved every 1000 integration steps. Only  
184 structures at 300 K and 310 K (40 trajectories) are reported here as they are close to the physiological  
185 temperature.

186 Since the simulations are performed on two monomers, both isolated non interacting monomer conforma-  
187 tions and aggregated monomers were observed in the converged MD trajectories. A dimeric conformation  
188 was defined as two monomers with more than 10 pairs of residues with at least one inter-molecular distance  
189 between their  $C^\alpha$  atoms smaller than 5 Å. This cut-off value was chosen because the average distance  
190 between two residues in an inter-molecular  $\beta$ -sheet in an  $\alpha$ -syn fibril is 4.8 Å (PDB ID: 2n0a). With this  
191 dimer definition, the fraction of dimers out of all the conformers simulated at 300 K and 310 K, are 31 %  
192 for the WT, 23 % for A30P, 29 % for A53T and 35 % for E46K, respectively. The formation of dimers  
193 is increased for the E46K variant and is reduced for A30P (significantly) and A53T (weakly) mutants.  
194 The dimeric state of each protein is finally described here by about 300000 structures extracted every  
195 1000 integration steps from the converged part of the replica exchange MD trajectories at 300 and 310 K,  
196 representing a sampling on an effective time-scale of 1.47 milliseconds.

197 Analysis of the secondary structures of  $\alpha$ -synuclein conformations was performed directly from the  
198  $C^\alpha$  coordinates with the CUTABI (CURvature and Torsion based Alpha and Beta-sheet Identification)  
199 algorithm recently developed in our group[72]. CUTABI is 10 to 30 faster than the commonly used DSSP  
200 algorithm (Dictionary of Secondary Structure of Proteins)[74, 75] because it avoids the construction of  
201 an all-atom description of the protein backbone from the coarse-grained UNRES structure, as required to  
202 apply DSSP. In CUTABI, the minimal size of a helix is set to 4 residues. Helices with less than 3 residues,  
203 such as short  $3_{10}$  helices, are thus not counted. The minimal size of a  $\beta$ -strand is set to 2 residues, i.e.,  
204 a  $\beta$ -sheet cannot be smaller than 4 residues. The  $\beta$ -strands of 1 residue forming  $\beta$ -bridges are thus not  
205 considered. Detailed description of the MD trajectories and of the CUTABI algorithm can be found in our  
206 previous work on monomeric state[72].

207 The statistics of contacts between amino acids (Figures 1 to 3) were computed by defining a contact as a  
208 pair of two  $C^\alpha$  atoms belonging to different residues at a distance smaller than 6 Å. By definition, the  $C^\alpha$   
209 atom of a residue may form several contacts. The mean contact of a residue (Figure 2) is computed as the  
210 average of all *inter-molecular* contacts made by this residue in all snapshots of all trajectories at 300 K and  
211 310 K.

### 3 RESULTS

#### 212 3.1 Propensity of $\alpha$ -syn dimerization

213 A dimensionless (effective) free-energy landscape of the  $\alpha$ -syn dimers,  $-\ln \left[ \frac{P(n_{inter}, n_{intra})}{P_{max}} \right]$ , was com-  
214 puted from the two-dimensional probability density  $P$  of the number of inter-chain ( $n_{inter}$ ) and intra-chain  
215 ( $n_{intra}$ ) contacts between the residues for the WT and the mutants (Figure 1).  $P_{max}$  is the maximum  
216 value of the probability for each protein. For all proteins,  $n_{intra}$  varies between about 200 and 400. The  
217 minimum value of  $n_{inter}$  is 11 by definition of a dimer (see Material and Methods) and the maximum  
218 about 150 for WT, A30P, A53T and 200 for E46K. As for the monomer, the  $\alpha$ -syn dimer has no unique  
219 native conformation but is represented by a large variety of dimer conformations. The structural diversity  
220 of the dimer conformations is shown in Figure 1 where one structure was randomly selected from each  
221 sub-population of the minima of the WT effective free-energy.

222 The numbered local minima correspond to different sub-populations of dimers with similar free-energies  
223 (Table 1). These minima are separated by low barriers for WT and A30P (violet to red regions between the  
224 minima in Figure 1). A larger barrier exists between the minimum 1 and 5 of A53T and the minimum 3 and  
225 the minima 1 and 2 for E46K. It is worth noting that only the WT has a map without large barriers, whereas  
226 the maps of A30P, A53T and E46K show small islands separated from the rest by large barriers. The  
227 inter-conversions between different sub-populations of dimer structures seem easier in the WT compared  
228 to the variants.

229 Figure 1 is muted on how the numerous inter-molecular contacts are spread along the sequence. To  
230 answer this question, we first calculate the mean inter-molecular contacts of residues along the sequence as  
231 shown in Figure 2. For the WT, we observe seven representative peaks with maximal at A17, A27, Y39,  
232 V49-A53 in the N-terminal, at A76, I88-A90 in the NAC and at L113 in the C-terminal. All these positions  
233 correspond to hydrophobic residues. In particular, the two largest peaks correspond to segments of three,  
234 A76-V77-A78, and four, I88-A89-A90-A91, hydrophobic residues, respectively. As mentioned previously,  
235 alanine plays a particular role in the  $\alpha$ -syn dynamics[55]. It is worth noting that the mean contact variable  
236 counts both inter-molecular random coil contacts and inter-molecular  $\beta$ -sheet contacts. However, in *in vitro*

237 fibrils, as in dimers computed from MD, the inter-molecular contacts are mainly from residues in  $\beta$ -sheets  
238 (see next subsection). Therefore, we compare the mean contact curves with the location of intermolecular  
239  $\beta$ -sheets computed by CUTABI in two different experimental *in vitro* structures of WT  $\alpha$ -syn fibrils (PDB  
240 IDs: 6ssx and 2n0a) in Figure 2. The peaks at Y39 and V49-A53 in MD are within a region of  $\beta$ -sheets in  
241 both experimental WT fibril structures as well as the peaks at A76 and I88-A90. A more precise comparison  
242 between inter-molecular  $\beta$ -sheets found in MD and these structures will be examined below in Figures 6  
243 and 8.

244 As shown in Figure 2, it is remarkable to observe that a single mutation has huge effects on the relative  
245 weights of the different local regions having high aggregation propensity in the WT. The effects of a single  
246 amino-acid substitution is not limited to residues close to the mutation: the single amino-acid substitution  
247 has long-range effects on the mean contact. Compared to WT, the main differences are : (i) a larger  
248 propensity to aggregate in the N-terminal region for A53T and E46K and smaller for A30P; (ii) a larger  
249 propensity to aggregate in the NAC region for A30P and lower for A53T; (iii) E46K has a propensity to  
250 aggregate larger than WT all along the sequence excepted in the region A76-I88 of the NAC. The largest  
251 peak of A30P is located at G69-V70-V71 corresponding to a succession of three hydrophobic residues. The  
252 highest propensity to aggregate is maximum at L38 for A53T and at V49 for E46K. The contribution of the  
253 C-term is modest and only significant close to L113 (except for A30P) and to Y136 (mainly for A30P).

254 How the different regions of the protein interact with each other is summarized by the intra-molecular and  
255 inter-molecular contact maps in Figure 3 for the WT and the mutants. In a such map, colored lines parallel  
256 to the diagonal indicate contacts between residues in parallel segments whereas colored lines perpendicular  
257 to the diagonal show contacts between residues in anti-parallel segments. The parallel lines are the most  
258 probable for the inter-molecular contacts for all proteins. In the opposite, perpendicular lines are the most  
259 probable and numerous for the intra-molecular contacts for the WT and mutants.

260 For the WT, most probable inter-molecular contacts are on the diagonal of the map in the region 75-93  
261 (NAC) which means that this region in one chain interacts with the similar region in the other chain. Such  
262 contacts are similar to those in fibrils. Using (A) and (B) for chains A and B, we find more precisely  
263 contacts between residues 75-85 (A) and 74-84 (B) and between 83-93 (A) and 83-93 (B). A similar  
264 region in the NAC has contact along a parallel to the diagonal, between residues 75-91 (A) and 84-100 (B).  
265 Other long regions of contacts shifted by several residues occur between residues 36-51 (A) and 2-17 (B),  
266 32-52 (A) and 63-83 (B), 52-63 (A) and 88-99 (B) (parallel segments) and between 52-66 (A) and 37-50  
267 (B) (anti-parallel segments). There is also a large number of small contacts between parallel segments in  
268 different regions. An important characteristic of the WT map is the absence of inter-molecular contacts  
269 along the diagonal in the region 50-65 which is favorable to helical structure in the monomer [72]. As  
270 mentioned, the contact map of intra-molecular contacts is characterized by a large number of anti-parallel  
271 segments in contacts for the WT. Using the numbers in the horizontal axis of the maps, the most probables  
272 are 2-23, 12-39, 14-47, 33/36-50/51 (N-terminal) and 67-77, 87-101 (NAC). The WT map shows a lot of  
273 weakly probable contacts in distant regions.

274 For the A30P, the most probable inter-molecular contact are along the diagonal or parallel to the diagonal  
275 between residues 38 and 95 (end of the N-term and NAC). Contacts strictly similar to those in fibrils are  
276 found between residues 80 and 95. Contacts between parallel segments close to the diagonal (shifted by  
277 several residues) are between residues 69-79 (A) and 73-83 (B), 38-67 (A) and 50-78 (B) and 17-25 (A).  
278 Probable anti-parallel inter-molecular contacts are observed mainly between residues 3-23 (A) and 32-51  
279 (B), 8-25 (A) and 87-105 (B). Compared to WT, the number of regions interacting with each other are less

280 numerous. The intra-molecular contacts are mainly between anti-parallel segments in the N-term, as in the  
281 WT, with an extended region of high probability for the segments 8-53 and 77-93.

282 For the A53T, the most probable inter-molecular contacts are in the N-term region [residues 1-60 (A)]  
283 with many contacts between shifted segments as for example 15-29 (A) and 41-55 (B), 26-35 (A) and  
284 36-45 (B), 37-55 (A) and 27-45 (B). On the contrary to A30P, WT and E46K (discussed next), there are  
285 only short regions of inter-molecular contacts parallel to the diagonal, namely in the NAC region (85-92)  
286 and in the C-term (111-116). The intra-molecular contacts in the N-term are similar to those of the WT,  
287 whereas contacts in the NAC are similar to those of A30P.

288 The E46K mutant is characterized by a large region of high probability of aligned inter-molecular contacts  
289 (on the diagonal) from the N-term to the NAC (regions 36-81 and 87-90) as well as a large region of  
290 parallel inter-molecular contacts between shifted segments: between the region 15-95 (A) and 35-109 (B).  
291 Intra-molecular contacts between anti-parallel segments are found in the regions 7-20, 12-38 and 72-94.

### 292 3.2 Heterogeneity of secondary structures elements in $\alpha$ -syn dimers

293 For each protein, the algorithm CUTABI[72] was applied to the ensemble of  $\alpha$ -syn dimers to compute  
294 the sum of residues in  $\alpha$ -helix ( $\alpha$ ) and in  $\beta$ -sheet ( $\beta$ ). Each dimer conformation has thus  $(\alpha, \beta)$  coordinates.  
295 The resulting probability densities in the  $(\alpha, \beta)$  space are represented in Figure 4. In these maps, only the  
296 residues from N-terminal and NAC regions were considered for the calculations as the C-terminal region  
297 does not contribute to secondary structure differences between WT and mutants[72].

298 For the isolated monomers, a major observation was that the conformations of the monomers were  
299 divided into two distinct states for the N-terminal+NAC region: an ensemble of conformations with no  
300 residue in helix (state B) and the rest of conformations (state HB). The highest probability of observing a  
301 conformation in state B was an order of magnitude larger than that of state HB[72]. In Figure 4, the states  
302 B are clearly visible for the monomers forming A30P and E46K dimers with a probability only twice larger  
303 than the state HB. On the opposite, states B disappeared in the WT and A53T dimers where the probability  
304 to find a dimer with no helix is completely negligible. For A30P, the maximum number of residues in  
305  $\beta$ -sheets is 155 and the highest probable number is 139, i.e. half of the residues of the dimer. A sub-state B  
306 is found at about 90 for A30P. For E46K, the maximum number of residues in state B is also large (133)  
307 with a peak at 108 and a second peak at 66.

308 The states B and HB can be distinguished from the function presented in Figure 5 showing the fraction of  
309 conformations within an effective free-energy difference cut-off from the global minimum of the state B for  
310 each protein[72]. With  $P_{max}$  the maximum of probability at  $(0, \beta)$  (in the B state), and  $P$  the probability  
311 at  $(\alpha, \beta)$  ( $\alpha \geq 0$ , in the B or HB states), the effective free-energy cut-off is computed as  $-\ln(\frac{P_{max}}{P})$  in  
312  $kT$  units, where  $k$  is the Boltzmann constant and  $T$  is the temperature. By definition, the derivative of  
313 the curves represented in Figure 5 represents the Density Of conformations or micro-States (DOS). For  
314 A30P and E46K, a change of slope (DOS) is observed at about 1.2 kT corresponding to about 10 % and  
315 45 % of the conformations, respectively. The change of slope at 1.2 kT points up the separation between  
316 main states, i.e. the onset of state HB, i.e. a state with a mixture of  $\alpha$ -helices and  $\beta$ -sheets. The case of  
317 E46K is special. We observe in fact two changes of slope at about 0.8 kT and 1.2 kT corresponding to the  
318 major peak at 110 residues and the minor peak at 70 residues in the one-dimensional probability density  
319 of the state B in Figure 4. The HB state of E46K (Figure 4) is also very different from the HB states of  
320 the other proteins: they are rare dimers with helical regions larger than about 15 residues. Conformations  
321 with a maximum number of residues in helix are a few on the time-scale of the present simulations with a  
322 maximum number of residues in helix of 59, 50, 63 and 61 for WT, A30P, A53T and E46K, respectively.



323 As for the monomers, A53T is the structure with the largest number of residues in helix [72]. It is worth  
324 noting that the free-energy map of the HB state of A53T is also the most diffuse. The global minima of the  
325 HB maps are (13,80), (13,98), (31,70) and (8,91), for the WT, A30P, A53T and E46K, respectively. The  
326 positions of these minima reflect also the largest propensities to form helical segments and  $\beta$ -sheets for  
327 A53T and E46K, respectively.

328 The probability to find the different secondary structures along the amino-acid sequence of the WT  
329 and mutants of  $\alpha$ -syn dimers were analyzed with CUTABI. As shown in Figure 6, helices are found in  
330 four main regions: two in the C-terminal region (residues 119-125 and 127-130), one in the NAC region  
331 (residues 75-82) and one overlapping the N-terminal and NAC regions (residues 53-65) both for WT and  
332 mutants. There is no significant difference between the propensities of helices for WT and mutants in  
333 the C-terminal region. These results for the helical propensity are similar to those found for the isolated  
334 monomers[72]. The main differences are that the probability to form the helix 53-65 is twice larger in the  
335 dimer than in an isolated monomer for the WT and that the probability to form the 75-82 helix is about  
336 twice smaller in the dimer than in an isolated monomer for A30P and E46K.

337 The intra-molecular and inter-molecular  $\beta$ -sheets were analyzed separately. The probability for a residue  
338 to pertain to an intra-molecular  $\beta$ -sheet has peaks at the same positions than for an isolated monomer. For  
339 the WT, the maximum of the peaks observed in Figure 6 are (in brackets for an isolated monomer[72]):  
340  $P_{F4} = 0.44(0.57)$ ,  $P_{K10} = 0.53(0.61)$ ,  $P_{A17} = 0.71(0.87)$ ,  $P_{T22} = 0.74(0.76)$ ,  $P_{E28} = 0.60(0.65)$ ,  
341  $P_{K34} = 0.64(0.61)$ ,  $P_{38} = 0.65(0.69)$  and  $P_{Y39} = 0.63(0.69)$ ,  $P_{K43} = 0.36(0.50)$ ,  $P_{V49} = 0.26(0.4)$ ,  
342  $P_{53} = 0.19$  and  $P_{V55} = 0.13(0.69)$ ,  $P_{V63} = 0.14(0.64)$ ,  $P_{V70} = 0.15(0.32)$ ,  $P_{81} = 0.18(0.35)$  and  
343  $P_{K80} = 0.17(0.41)$ ,  $P_{88} = 0.43$  and  $P_{A90} = 0.42(0.58)$ , and  $P_{V95} = 0.20(0.35)$  and  $P_{97} = 0.27$ . These  
344 peaks are at the same locations for the mutants as it was the case in isolated monomers. These maxima are  
345 located at or close to valine residues, which is the most frequently found amino-acid in  $\beta$ -sheets[76]. The  
346 empirical prediction of the dimensionless propensity of formation of  $\beta$ -sheet computed from the scale of  
347 Chou and Fasman (CF)[76] is shown in Figure 6 for WT. The CF simple algorithm, using only the primary  
348 sequence, predicts some peaks at similar locations than predicted by MD, namely at 39, 52, 71, and 94.  
349 Of course, there are quantitative differences in the shape and relative heights predicted by CF compare to  
350 MD (they are not exactly the same quantities) but both methods identify qualitatively similar local regions  
351 of sequences which are important for  $\beta$ -sheets. It is worth noting that the CF algorithm predicts only a  
352 relatively small local change of the  $\beta$ -sheet propensity for the mutants, as the influence of a residue is  
353 limited to a few residues in the vicinity of its position.

354 Compared to WT, the increase of probability of intra-molecular  $\beta$ -sheets in the region 26-35 for A30P  
355 and its decrease in the region 53-65 for A53T were found also in the isolated monomers[72]. Compared  
356 to isolated monomers, the probability for a residue to pertain to an intra-molecular  $\beta$ -sheet decreases  
357 significantly from residue 43 to 100 for WT, A30P and A53T. Therefore the intra-molecular  $\beta$ -sheets form  
358 mainly in the N-terminal region for these three proteins rather they are found equally in the N-terminal and  
359 NAC region for E46K. Note that the CF algorithm underestimates the large propensity of formation of  
360  $\beta$ -sheets in the N-terminal region.

361 Maxima of probability to form inter-molecular  $\beta$ -sheets are located at the same positions than those  
362 of the probability to form intra-molecular  $\beta$ -sheets. The amplitude of these peaks are however different  
363 between the proteins. They are four major peaks for the WT and A30P:  $P_{Y39,WT} = 0.33$ ,  $P_{T75,WT} = 0.31$ ,  
364  $P_{T81,WT} = 0.29$ ,  $P_{I88,WT} = 0.36$  and  $P_{V70,A30P} = 0.40$ ,  $P_{T75,A30P} = 0.46$ ,  $P_{I88,A30P} = 0.40$ ,  
365  $P_{A90,A30P} = 0.52$ , respectively. Except for the peak at Y39, the larger propensities to form inter-molecular  
366  $\beta$ -sheets are located in the NAC for WT and A30P. For A53T, all the major peaks are in the N-terminal

367 region:  $P_{F4,A53T} = 0.36$ ,  $P_{E28,A53T} = 0.37$ ,  $P_{L38,A53T} = 0.60$ ,  $P_{K43,WT} = 0.34$ . For the variant  
368 E46K, the propensity to form inter-molecular  $\beta$ -sheets are mainly in the N-terminal region with peaks  
369 at  $P_{A27,E46K} = 0.35$ ,  $P_{Y39,E46K} = 0.45$ ,  $P_{T44,E46K} = 0.37$ ,  $P_{V49,E46K} = 0.40$ ,  $P_{V53,E46K} = 0.35$   
370 but also in the NAC region with peaks at  $P_{V71,E46K} = 0.31$ ,  $P_{A90,E46K} = 0.40$ . A single amino-acid  
371 substitution modifies significantly the most probable regions of formation of inter-molecular  $\beta$ -sheets:  
372 highest in the NAC for WT and A30P, in the N-terminal region for A53T, and both in the N-terminal region  
373 and in the NAC for E46K. These results compare well with the mean inter-molecular contact profiles  
374 (Figure 2). As shown in Figure 6 for the WT, the mean contact curve follows quite well the probability  
375 of inter-molecular contact (except for the 80-85 region). Finally, the probability to form inter-molecular  
376  $\beta$ -sheets between the same residues in both monomers (diagonal in the contact maps in Figure 3), named  
377 Native fibril like contacts (Nfc), are also shown for comparison in Figure 6. For all proteins, the probability  
378 of Nfc is low. The sub-ensemble of dimers showing these contacts is analyzed and discussed next.

### 379 3.3 Dimers with fibril native contacts (Dfnc)

380 Contact maps of WT and mutants show contacts along their diagonal as in the experimental structures  
381 of fibrils (PDB IDs: 6ssx and 2n0a). To identify conformations belonging to the diagonal of the contact  
382 maps, we extracted from the dimer statistics a sub-ensemble of dimers with Nfc. We define a Dimer with  
383 fibril like native contacts (Dfnc) as a dimer conformation which has at least 5 consecutive inter-molecular  
384 contacts made between residues with the same indices at a contact distance less than 5Å. This value was  
385 chosen as the distances between native contacts in fibrils are 4.9Å and 4.7Å in the measured structures  
386 with PDB IDs 6ssx and 2n0a. We may consider the Dfnc ensemble as the conformations which are the  
387 most probable pre-formed fibrils in the present work.

388 The dimensionless (effective) free-energy landscape of the Dfnc was computed from the two-dimensional  
389 probability density  $P$  of the number of inter-chain ( $n_{inter}$ ) and intra-chain ( $n_{intra}$ ) contacts between the  
390 residues for the WT and the mutants (Figure 7). For the WT and A53T, the Dfnc are located in one basin  
391 whereas for A30P and E46K two distinct clusters are visible. Some of the fibril clusters are located at  
392 local minima of the global free-energy landscape (Figure 1), except for A53T. For A30P and E46K, one  
393 of two clusters of Dfnc occur at the global minimum of the global free-energy landscape (Figure 1). This  
394 indicates that Dfnc are more probable to be formed for these two mutants. In fact, the sub-ensemble of  
395 Dfnc represents 14,04 % and 15,73 % for A30P and E46K, respectively. For WT and A53T, the population  
396 of Dfnc is only 8.33 % and 10.65 %, respectively. It is worth noting that these numbers indicate that the  
397 majority of dimers are disordered aggregates at least on the effective time-scale (millisecond) of the present  
398 simulations .

399 The calculation of the probability to find the different secondary structures elements along the amino-acid  
400 sequence of the WT and mutants of  $\alpha$ -syn dimers was repeated for the sub-ensemble of Dfnc using CUTABI  
401 (Figure 8). Compared to the ensemble of dimers, the helical regions of Dfnc show significant differences  
402 for the WT and A53T: the region 53-65 can only form an helix in WT preventing the dimerisation of this  
403 region whereas this helical region disappears in A53T as well as the helical region 75-82 disappears in WT.

404 The peaks of probability for a residue to pertain in an intra-molecular  $\beta$ -sheet occur at the same positions  
405 in Dfnc sub-ensemble and in the ensemble of the dimers for all proteins (see Figure 6). For the WT, the  
406 propensity of intra-molecular  $\beta$ -sheets is strongly increased in the N-terminal region of Dfnc with the  
407 peaks with a probability near 1 meaning residues pertaining to this type of secondary structure in all Dfnc  
408 conformations. On the contrary, all residues above the position 50, have a totally negligible probability to  
409 form an intra-molecular  $\beta$ -sheet in Dfnc. Unlike the ensemble of WT dimers for which inter-molecular

410  $\beta$ -sheets can be found nearly everywhere along the sequence (Figure 6), in the Dfnc sub-ensemble, they are  
411 found only for residues 43-51 (N-terminal region), 69-95 (NAC region), 99-100 and 134-135 (C-terminal  
412 region). We observe a propensity to form Nfc, as in fibrils, only for specific regions in the NAC: residues  
413 75-76, 80-82, 85, 87-93. For the segment 87-93, the superposition of the black and red curves in the Figure  
414 8 means that only Nfc are formed there. Considering the limited time-scale of the present simulations,  
415 these residues are probably nucleation centers of pre-fibril like dimers in WT.

416 For A30P, the probability to form intra-molecular  $\beta$ -sheet is similar in Dfnc and in the full ensemble  
417 of dimers (Figure 6) but with an increase in the region 36-58. On the contrary to the ensemble of A30P  
418 dimers, for which inter-molecular  $\beta$ -sheets can be found nearly everywhere between residue 3 to 104 along  
419 the sequence (Figure 6), in Dfnc they are found in the NAC for residues 70-83 and 87-96 as well as with  
420 smaller probabilities in the C-terminal region at 103-105 and in the N-terminal region 18-20 and 26-28.  
421 The propensity to form Nfc occurs only in the NAC at 87-88 and 90-95 which are probably the nucleation  
422 centers for this type of structure for this variant. Clearly, WT and A30P share the same region of formation  
423 of Nfc in the NAC.

424 For A53T, the propensity for a residue to pertain to an intra-molecular  $\beta$ -sheet is increases in the NAC  
425 and C-terminal regions in Dfnc compared to the complete ensemble of dimers (Figure 6). The formation of  
426 inter-molecular  $\beta$ -sheets occur all along the sequence: residues 3-11, 22-30, 35-60 (N-terminal region),  
427 61-65, 71, 87-92 (NAC region) and 112-116 (C-terminal region). The A53T differs thus from WT and  
428 A30P by a large number of regions of inter-molecular  $\beta$ -sheets including the N-terminal region. In addition,  
429 the probability of segments 44-65 and 87-92 in the NAC to form Nfc is identical to the probability of  
430 inter-molecular  $\beta$ -sheets, meaning that these residues form only Nfc in the Dfnc sub-ensemble and are  
431 probably nucleation centers for A53T pre-fibril like dimers.

432 For E46K, the probability of intra-molecular  $\beta$ -sheets is similar between Dfnc and the ensemble of dimers.  
433 The exception is the peak at residues 38 and 39 for which the probability is 1 as well as the probability  
434 for these residues to form Nfc, meaning that these residues are involved in both Nfc inter-molecular and  
435 intra-molecular  $\beta$ -sheets. As for the other proteins, the presence of inter-molecular  $\beta$ -sheets is limited to  
436 specific regions on the contrary of the ensemble of dimers where they were found along all the sequence  
437 (Figure 6). In addition, nearly all inter-molecular  $\beta$ -sheets are Nfc (the black and red curves in Figure 8  
438 overlap), namely at 26-28 and 37-60 (N-terminal region) and 61-65, 75-78 and 88-92 (NAC).

439 Figure 8 shows that regions of high propensity of Nfc in A53T and E46K are significantly different  
440 than those of the WT and A30P. This can be more clearly seen for the Nfc in Figure 9. Figure 9 shows  
441 the different lengths of the Nfc regions and their probabilities along the sequence. A30P has the shortest  
442 consecutive segments of Nfc with the longest between residues 86 and 97 (NAC and C-terminal region)  
443 with the largest probability of Nfc at 87-88 (NAC). The maximum number of consecutive Nfc for the WT  
444 is 18 between residues 78 and 95 (NAC). As already mentioned, the formation of Nfc in A53T and E46K  
445 occur differently compared to the WT and A30P: they are mainly in the N-terminal region. These two  
446 variants have the largest segments of consecutive Nfc on the time-scale of the simulation: 25 between  
447 residues 43 and 66 for A53T and 37 between residues 34 and 70 for E46K. The structures of the proteins  
448 with the largest number of Nfc are represented in insets of Figure 9, they can be considered as the most  
449 probable nucleation structures of Dfnc on the time-scale of the present simulations (millisecond). As  
450 mentioned above, the number of Dfnc structures is the largest for A30P and E46K. One may deduce that  
451 the nucleation is easier for these two structures. On the other hand, if we consider the maximum number of  
452 Nfc possibly formed on the time-scale of the simulations as a criteria on the difficulty of the growth of

453 Nfc from a nucleation center, we find by increasing difficulty of expansion of the Nfc segment: A30P, WT,  
454 A53T and E46K.

455 As shown in Figure 9, some Dfnc of A53T and E46K have segments which does not start before residue  
456 39 and does not expand above residue 58. It is worth noting that residue 58 is the location of the maximum  
457 of probability to form an helix and residue 39 has a high propensity to form intra-molecular  $\beta$ -sheets in  
458 both disordered dimers (Figure 6) and in the monomers[72].

### 459 3.4 Selected local and global properties related to experimental techniques

460 In dimers, the residues may form both intra-molecular and inter-molecular parallel and anti-parallel  
461  $\beta$ -sheets which can be distinguished in infrared and Raman spectroscopies using amide bands. Table 2  
462 shows a significant variation in the amount of parallel  $\beta$ -sheets between the monomers and the dimers.  
463 The formation of disordered dimers and Dfnc is clearly characterized by an increase of the proportion  
464 of residues in parallel  $\beta$ -sheets compared to monomers: by increasing percentage, one finds monomer,  
465 disordered dimers and Dfnc. The amount of parallel  $\beta$ -sheets in Dfnc is the largest for the WT despite the  
466 fact that the length of segments of Nfc is quite short (Figure 9), which implies that most of the parallel  
467  $\beta$ -sheets are in segments of WT monomers which are not aligned as in fibrils. For the proportion of  
468 anti-parallel  $\beta$ -sheets, we do not observe a clear systematic variation between the monomers and the  
469 dimers. For A53T and E46K, the percentage of anti-parallel  $\beta$ -sheets is increased from the monomers to  
470 disordered dimers and Dfnc. No significant change is found for the WT. The case of A30 is special, the  
471 proportion of anti-parallel  $\beta$ -sheets is decreased in Dfnc and increased in disordered dimers compared to  
472 the monomers. It is worth noting that on the time-scale of the simulations, the Dfnc represents about 9-15  
473 % of the structures, the other dimers being disordered.

474 A global parameter which can be measured by SAXS is the gyration radius. The probability density  
475 functions of the gyration radius of the ensemble of dimers are represented in Figure 10. Each function  
476 can be represented by the sum of several sub-populations described by Gaussian functions. The Gaussian  
477 parameters, given in Table 3, were computed with the Gaussian Mixture Model (GMM) algorithm[77].  
478 However, the GGM clustering is misleading as the sub-populations cover large areas of the free-energy  
479 landscape maps of the contacts,  $(n_{inter}, n_{intra})$  (Figure 1), and of the secondary structures,  $(\alpha, \beta)$  (Figure  
480 4), as shown in Figures S1 and S2, respectively. Finally, the average gyration radius (Table 3) is the smallest  
481 for E46K and the largest for A30P.

482 Single-molecule FRET allows to extract informations on local properties. In references[41, 42, 46], the  
483 fluorophores (Alexa Fluor 488 and Alexa Fluor 647) were covalently linked at the position 90 of  $\alpha$ -syn  
484 mutants A90C. The FRET efficiencies depends on the distance between the fluorophores. Therefore, the  
485 FRET experimental procedure observed the aggregation of  $\alpha$ -syn using a specific local property which is  
486 directly correlated to the distance between residues 90 of interacting monomers. It is, therefore, interesting  
487 to compute the probability density functions of the distances between the  $C^\alpha$  of residues A90 in disordered  
488 dimers and in Dfnc. These functions are represented in Figure 11 (Panels A and B) for the WT and mutants.  
489 In disordered dimers, we observe a heterogeneity of these distances with a group of peaks below 15 Å and  
490 other peaks between 20 Å and 80 Å (panel A). In the sub-population of Dfnc, there is a drastic change of  
491 the distance probability distribution (panel B). All proteins, except E46K, have a peak around 3.8 Å. This  
492 can be explained by Figures 8 and 9, which show that the residue 90 is involved in Nfc for all proteins,  
493 except E46K. All other peaks present in disordered dimers have disappeared in Dfnc, except for E46K,  
494 for which we observe a peak at around 15 Å and a large background. Regarding this specific A90-A90

495 distance, the difference between the probability densities of disordered dimers and Dfnc is spectacular. In  
496 addition, this clearly shown that E46K has a different behavior regarding this local parameter.

## 4 DISCUSSION

497 A first important finding of the present simulations is that the ensemble of  $\alpha$ -syn conformations (WT  
498 and mutants) is strongly heterogeneous, as shown by the computed probability density functions of the  
499 gyration radius (Figure 10), the effective free-energy landscape of the contacts (Figure 1), the effective  
500 free-energy landscape of the secondary structures (Figure 4), and by the probability density function of  
501 the A90-A90 distance (Figure 11). A second significant finding is that the ensemble of heterogeneous  
502 dimer conformations is divided in a majority of disordered dimers and a complementary minority of Dfnc  
503 (Figure 9). A third important finding is that a single-amino acid substitution has a huge effect on the contact  
504 probability between (hydrophobic) residues of  $\alpha$ -syn, which is not limited to the vicinity of the mutated  
505 residue as shown in Figure 6 for disordered dimers and in Figures 8 and 9 for Dfnc. Keeping in mind that the  
506 effective time-scale of the present simulations (millisecond) represents the early lag phase of fibril growth,  
507 as it is three to four order of magnitudes smaller than the lag phase of fibrils observed experimentally, we  
508 will, however, attempt next to make qualitative comparisons between these three significant results and  
509 different experimental findings.

510 Unfortunately, there are no SAXS or SANS measurements of  $\alpha$ -syn dimers to compare directly with  
511 Figure 10 to the best of our knowledge. It is worth noting that the distribution of gyration radius of  $\alpha$ -syn  
512 monomers is single peak, except for the sub-ensemble of A30P and E46K monomers without alpha-helix  
513 (B monomer state) for which a small peak at  $R_g=18 \text{ \AA}$  was found[72]. For recombinant monomeric  $\alpha$ -syn,  
514 the average gyration radius, extrapolated at infinite dilution, is  $27.2 \pm 0.44 \text{ \AA}$ [78], comparable to the value  
515 calculated previously in our MD simulations ( $24.7 \text{ \AA}$ ) [72]. The experimental value of the gyration radius  
516 of a monomer varies among the experiments: for example, in Tris-Buffer, it is increased to  $42.7 \text{ \AA}$ [78]. We  
517 hypothesize that the presence of dimers in a monomeric solution may give an apparent larger gyration  
518 radius of the monomers because the dimer conformational ensembles have an average gyration radius of  
519 about  $10 \text{ \AA}$  larger than one of the monomer ensembles (Table 2).

520 The heterogeneity of the size distribution (Figure 10) and of the contacts (Figure 1) of  $\alpha$ -syn dimers can be  
521 related to the rupture-force AFM experiments on WT, A30P, A53T and E46K  $\alpha$ -syn dimers in solution[62].  
522 These experiments were performed at low pH to increase the aggregation propensity of  $\alpha$ -syn and with  
523 the additional mutation A40C needed to hang one of the monomers on the AFM tip and the other on the  
524 surface. Contour lengths of the dimers were extracted from force-distance curves. They were interpreted as  
525 total lengths of stretchable parts of  $\alpha$ -syn molecules which are not involved in dimeric interaction, and thus  
526 as the sum of the total length from the C-terminal anchor point to the first residue of a dimerization region  
527 in each monomer. The data show a multi-peaks distribution of contour lengths[62], which might reflect  
528 the multiple minima observed in Figure 1, as the contour length of a dimer conformation is related to the  
529 number of intra-molecular and inter-molecular contacts. Moreover, the distribution of contour lengths are  
530 different for each  $\alpha$ -syn variant in the AFM experiments, as in Figure 1 for the contacts. We hypothesize  
531 that the number of single and multiple rupture force events might be correlated with the propensity of the  
532 dimers to form Nfc. A larger number of multiple rupture force events for A53T and E46K compared to WT  
533 and A30P was interpreted as multiple interactions segments for the former[62]. This might be supported  
534 by Figure 8 (black and red curves) as WT and A30P form Nfc and inter-molecular  $\beta$ -sheets in shorter  
535 localized segments than A53T and E46K. Moreover, contour lengths can be estimated for residues not  
536 involved in Nfc from Figure 8 by assuming a  $C^\alpha - C^\alpha$  virtual bond distance of  $3.8 \text{ \AA}$  and evaluating the

537 contour length as twice (for the two monomers) the distance starting from the C-terminal to the first residue  
538 forming a Nfc[62]. For WT Dfnc, we find a contour length of 357 Å (residues 140 to 94). For A30P Dfnc,  
539 the distance is similar. For A53T Dfnc, two contour lengths are estimated: 380 Å (residues 140 to 91)  
540 and 585 Å (residues 140 to 64). This last value is the same for E46K Dfnc. These values are on the same  
541 order of magnitude that the 340 Å and 440 Å of the major peaks of the contour length distributions in the  
542 experiment[62]. The comparison, although qualitative, indicates that the heterogeneity of  $\alpha$ -syn dimers  
543 and the influence of a single amino-acid substitution on their structural properties, found in MD, agree  
544 qualitatively with the AFM force-distance data.

545 The heterogeneity of the ( $\alpha$ ,  $\beta$ ) map could be tested in the early lag phase of fibril formation by using  
546 single-molecule Raman spectroscopy[79], as  $\alpha$ -helices and  $\beta$ -sheets have been well described by Raman  
547 fingerprints. Such data are not available so far. In the present simulations, the ensemble of heterogeneous  
548 dimer conformations is divided in a majority of disordered dimers and a complementary minority of Dfnc  
549 both for WT and mutants. Dfnc could be identified by spectroscopy. Indeed, the percentage of parallel  
550 inter-molecular  $\beta$ -sheets is characteristic of Dfnc as it is significantly larger than the one in monomers  
551 and in disordered dimers (Table 2). This result agrees with FTIR spectroscopy measurements indicating  
552 that WT  $\alpha$ -syn fibrils have a majority of parallel  $\beta$ -sheets in opposite to oligomers showing a majority  
553 of anti-parallel  $\beta$ -sheets[48]. The total amounts of  $\beta$ -sheets estimated from FTIR[48] are  $35\% \pm 5\%$  for  
554 oligomers of 10-40 molecules and  $65 \pm 10\%$  for fibrils. In spite of the smaller size of dimers and the shorter  
555 time-scale of the present simulations, these values compare qualitatively with the sum of parallel and  
556 anti-parallel  $\beta$ -sheets in WT disordered dimers (44,4%) and in Dfnc (57,8%), respectively (Table 2).

557 Probably the best technique to identify Dfnc from disordered dimers would be single-molecule FRET. We  
558 tentatively compare the results of Figure 11 with FRET studies of  $\alpha$ -syn oligomerization[22, 41, 42, 46]  
559 in the early lag phase of fibril formation. A quantitative prediction of the FRET efficiencies from  $\alpha$ -syn  
560 conformations would require to simulate the protein labeled with the fluorescent molecules and to model  
561 the FRET mechanism[80]. Such an approach is difficult to apply for a large and complicated protein like  
562  $\alpha$ -syn. In addition, to predict protein structure modifications introduced by the cysteine mutations and the  
563 fluorescent organic molecules, it is necessary to build and test a new coarse-grained force field for the  
564 fluorescent molecules covalently linked to  $\alpha$ -syn. This task is extremely demanding in computer time and  
565 is well beyond the scope of the present work. For these reasons, we choose to compute an approximate  
566 efficiency  $E$  from the distance distributions between the  $C^\alpha$  atoms shown in Figure 11 (Panel A and B).  
567 From these distances, we extracted an approximated FRET efficiency  $E$  as  $E = 1/[1 + (d/R_0)^6]$ , where  $d$   
568 is the distance between the  $C^\alpha$  atoms increased by 20Å and  $R_0$  is the Förster distance with  $R_0 = 60\text{Å}$  for  
569 the fluorophores used[81]. Note that a single value of  $R_0$  is a severe approximation which assumes an  
570 averaging of the fluorophore orientation. The shift of 20Å takes into account of the estimated distance  
571 between the dye center of a fluorophore and the  $C^\alpha$  atom to which it is linked[82].

572 The probability distribution of  $E$  is shown in Figure 11 (Panels C and D). In disordered dimers, one  
573 observed two regions with peaks of low ( $< 0.15$ ) and of high efficiency ( $> 0.75$ ). In Dfnc, peaks are ob-  
574 served only at high efficiency ( $> 0.9$ ) with a small background at all efficiencies. It is tempting to associate  
575 the Dfnc to type B dimers (panel C). The disordered dimers are both A and B (panel D). It was proposed  
576 that the  $\alpha$ -syn dimers might be a mixture of type A and B which cannot be resolved experimentally[46].  
577 The time-scale simulated here is of course extremely short compared to FRET experiments but it is very  
578 interesting to observe that the formation of consecutive Nfc lead to an increase of the FRET efficiency.  
579 Comparison with Figure 9 shows that A90 belongs to all segments of consecutive Nfc except for E46K.  
580 However, Figure 8 shows that the Dfnc of E46K variant form other contacts in the A90 region. E46K

581 behaves thus differently from WT, A30P and A53T in the A90 region which may be a hint to explain  
582 the difficulty to reproduce the probability distribution of the FRET signal in the experiments[46]. If the  
583 population of dimers formed in the early lag phase measured by FRET for A30P is taken as 1 for reference,  
584 the population of WT and A53T dimers were 0.4 and 0.64, respectively[46]. In the present simulations,  
585 the populations of Dfnc using the same reference is 0.59 and 0.75 for WT and A53T. However the total  
586 populations of dimers out of all the conformations (monomers+dimers) found in the simulations are  
587 different. Taken A30P as a reference (=1), we found 1.34 and 1.26 for WT and A53T.

588 The effect of a single amino-acid substitution on the propensity to form contacts (Figure 1) and on the  
589 formation of Nfc (Figures 6, 8 9) is spectacular. How these results could be compared to various differences  
590 observed experimentally between the variants and the WT ? The aggregation of A30P in fibrils is slower  
591 than the WT[35] whereas E46K and A53T aggregate faster than WT[83, 84]. However, the A30P monomer  
592 was consumed at a comparable rate or slightly more rapidly than the WT monomer, whereas A53T was  
593 consumed even more rapidly[35]. This might be explained by the early nucleation steps of the fibrils which  
594 might be represented by Figure 8. Indeed E46K and A53T form *larger* regions of Nfc than the A30P and  
595 WT, which might be interpreted as they are "faster" to form pre-fibrils. For A30P the region of Nfc is  
596 extremely short. However, the sub-ensemble of Dfnc in the dimer ensemble is larger for A30P (14.04%)  
597 and E46K (15.73 %) in comparison with WT (8.33 %) and A53T (10.65%). On the opposite, the number of  
598 dimers found in the MD simulations is the lowest for A30P (23%), followed by increasing order by A53T  
599 (29%), WT (31%), E46K (35%). To make more concrete these percentages, it is better to take an example.  
600 For 123 monomers in solution, A30P forms 20 disordered dimers, 3 Dfnc and 77 monomers remain free.  
601 For 121 monomers in solution, the WT form 28 disordered dimers, 3 Dfnc and 69 monomers remain free.  
602 For 129 monomers, A53T forms 26 disordered dimers, 3 Dfnc and 71 monomers remains free. Finally, for  
603 135 monomers, E46K forms 30 disordered dimers and 5 to 6 Dfnc and 65 monomers remain free. Thus, all  
604 proteins will form approximately the same number of Dfnc, except E46K, but the size of Dfnc on the same  
605 time-scale is much larger for A53T and E46K (Figure 9).

606 In order to compare present results to WT mutagenesis experiments, we list the key residues which  
607 play a role in  $\alpha$ -syn dimerization in the present simulations in the Table 4 for WT and mutants. These  
608 residues were selected from the maxima of probabilities of mean contacts (Figure 1) and of the formation  
609 of inter-molecular  $\beta$ -sheets and Nfc (Figures 6 and 8).

610 For WT, the residue I88 is of paramount importance. The segment S87-G93 forms only Nfc in the  
611 sub-ensemble of Dfnc (black and red curves are superposed in Figure 8). This agrees with the observation  
612 that the removal of the segment A85-E94 reduces the  $\alpha$ -syn polymerization[85]. Truncation of V71-V82  
613 prevents the polymerization of fibrils[85]. This agrees with the fact that K80-V82 forms Nfc. We found that  
614 the pair T75-A76 has non-negligible probabilities to form Nfc and other inter-molecular  $\beta$ -sheets (Figure  
615 8). However, removal of residue A76 or V77 alone has no effect on the polymerization but the missing  
616 pair A76-V77 prevents the polymerization in fibrils[85]. The role of A76 cannot be neglected. It is worth  
617 noting that Table 4 suggests that mutations or cutting segments in the N-terminal of A53T and E46K could  
618 provide information on the key residues promoting the aggregation of these variants.

619 The role of the N-terminal region in  $\alpha$ -syn in the dimerization has not to be underestimated as shown by  
620 the formation of Nfc for A53T and E46K and by the high probability to form intra-molecular  $\beta$ -sheets for  
621 WT and A30P in this region (Figures 8 and 9). In a recent work, two segments in the N-terminal region  
622 which regulate the  $\alpha$ -syn polymerization have been identified: G36-S42 (named P1) and K45-E57 (named  
623 P2)[86, 87]. Removal of P1 at pH 7.5 prevents  $\alpha$ -syn aggregation but not at pH 4.5. Removal of P1 and P2  
624 prevents aggregation at both pH. The present simulations are calibrated at pH7 for which the force-field

625 was developed. As a single amino-acid substitution has huge effect on the propensity of aggregation, it is  
626 difficult to compare the results of P1 truncated protein with the present results. However, the dimerization  
627 of A53T and E46K show that a mutation in this region changes drastically the Nfc propensity. Moreover,  
628 for the WT, the region G36-S42 is a region with high propensity of intra-molecular  $\beta$ -sheets in Dfnc which  
629 precedes a region of inter-molecular  $\beta$ -sheets (43-51) (black curve in Figure 8). It is possible that the  
630 formation of intra-molecular  $\beta$ -sheets may be coupled to the formation of fibrils. Finally, the region P2  
631 overlaps the region of the main helical segment of WT monomers[72] and dimers (Figures 6 and 8). Clearly,  
632 the high probability of this helical segment must play a role in the polymerization.

633 In conclusion, present MD simulations show that the dimer conformations are largely heterogeneous with  
634 both disordered and pre-fibrillar like dimers which differ between the WT and the variants A30P, A53T and  
635 E46K. Despite the limitations inherent to any MD simulations (accuracy of the force field and the limited  
636 time scale), the present findings agree quite well with available experimental data and suggest possible  
637 further spectroscopic and mutagenesis experiments.

## CONFLICT OF INTEREST STATEMENT

638 The other authors declare that the research was conducted in the absence of any commercial or financial  
639 relationships that could be construed as a potential conflict of interest.

## DATA AVAILABILITY STATEMENT

640 The program CUTABI and the  $\alpha$ -syn conformations analyzed in the present work are available upon request  
641 from the corresponding author.

## AUTHOR CONTRIBUTIONS

642 P.S. and G.G.M. designed research. A.G. performed research. A.G., P.S. and G.G.M. analyzed the data.  
643 A.R. provided technical assistance with the UNRES package. A.N. performed the Gaussian clustering of  
644 gyration radii. P.S. wrote the paper and A.G. did the figures. All authors discussed the results and reviewed  
645 the manuscript.

## FUNDING

646 The calculations were performed using HPC resources from DSI-CCuB (Centre de Calcul de l'Université  
647 de Bourgogne) and 588-processor Beowulf cluster at the Baker Laboratory of Chemistry and Chemical  
648 Biology, Cornell University. This work is part of the project NANO-NEURO-MED (2019–2022) supported  
649 by the EIPHI Graduate School (contract ANR-17-EUR-0002), the Conseil Régional de Bourgogne Franche-  
650 Comté and the European Union through the PO FEDER-FSE Bourgogne 2014/2020 programs. This work  
651 is supported by a grant from the National Institutes of Health grant no. R01GM14312.

## SUPPLEMENTARY MATERIAL

## REFERENCES

652 [1] Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function  
653 paradigm. *Journal of Molecular Biology* **293** (1999) 321–331. doi:10.1006/jmbi.1999.3110.



- 654 [2] Jakes R, Spillantini MG, Goedert M. Identification of two distinct synucleins from human brain.  
655 *FEBS letters* **345** (1994) 27–32. doi:10.1016/0014-5793(94)00395-5.
- 656 [3] Mollenhauer B, Cullen V, Kahn I, Krastins B, Outeiro TF, Pepivani I, et al. Direct quantification  
657 of CSF  $\alpha$ -synuclein by ELISA and first cross-sectional study in patients with neurodegeneration.  
658 *Experimental Neurology* **213** (2008) 315–325. doi:10.1016/j.expneurol.2008.06.004.
- 659 [4] Teil M, Arotcarena ML, Faggiani E, Laferriere F, Bezard E, Dehay B. Targeting  $\alpha$ -Synuclein for PD  
660 Therapeutics: A Pursuit on All Fronts. *Biomolecules* **10** (2020) 391. doi:10.3390/biom10030391.
- 661 [5] Stefanis L.  $\alpha$ -Synuclein in Parkinson's disease. *Cold Spring Harbor Perspectives in Medicine* **2** (2012)  
662 a009399. doi:10.1101/cshperspect.a009399.
- 663 [6] Chiti F, Dobson CM. Protein Misfolding, Functional Amyloid, and Human Disease. *Annual Review*  
664 *of Biochemistry* **75** (2006) 333–366. doi:10.1146/annurev.biochem.75.101304.123901.
- 665 [7] Chiti F, Dobson CM. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of  
666 Progress Over the Last Decade. *Annual Review of Biochemistry* **86** (2017) 27–68. doi:10.1146/annurev-  
667 biochem-061516-045115.
- 668 [8] Soto C. Unfolding the role of protein misfolding in neurodegenerative diseases. *Nature Reviews*  
669 *Neuroscience* **4** (2003) 49–60. doi:10.1038/nrn1007.
- 670 [9] Tanudjojo B, Shaikh SS, Fenyi A, Bousset L, Agarwal D, Marsh J, et al. Phenotypic manifestation of  $\alpha$ -  
671 synuclein strains derived from Parkinson's disease and multiple system atrophy in human dopaminergic  
672 neurons. *Nature Communications* **12** (2021) 3817. doi:10.1038/s41467-021-23682-z.
- 673 [10] Spillantini MG, Schmidt ML, Lee VMY, Trojanowski JQ, Jakes R, Goedert M.  $\alpha$ -Synuclein in Lewy  
674 bodies. *Nature* **388** (1997) 839–840. doi:10.1038/42166.
- 675 [11] Breydo L, Wu JW, Uversky VN. A-synuclein misfolding and Parkinson's disease. *Biochimica Et*  
676 *Biophysica Acta* **1822** (2012) 261–285. doi:10.1016/j.bbadis.2011.10.002.
- 677 [12] Lashuel HA. Do Lewy bodies contain alpha-synuclein fibrils? and Does it matter? A brief  
678 history and critical analysis of recent reports. *Neurobiology of Disease* **141** (2020) 104876.  
679 doi:10.1016/j.nbd.2020.104876.
- 680 [13] Trinkaus VA, Riera-Tur I, Martínez-Sánchez A, Bäuerlein FJB, Guo Q, Arzberger T, et al. In  
681 situ architecture of neuronal  $\alpha$ -Synuclein inclusions. *Nature Communications* **12** (2021) 2110.  
682 doi:10.1038/s41467-021-22108-0.
- 683 [14] Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, et al. Mutation in the  
684  $\alpha$ -Synuclein Gene Identified in Families with Parkinson's Disease. *Science* **276** (1997) 2045–2047.  
685 doi:10.1126/science.276.5321.2045.
- 686 [15] Zarranz JJ, Alegre J, Gómez-Esteban JC, Lezcano E, Ros R, Ampuero I, et al. The new mutation,  
687 E46K, of  $\alpha$ -synuclein causes parkinson and Lewy body dementia. *Annals of Neurology* **55** (2004)  
688 164–173. doi:https://doi.org/10.1002/ana.10795.
- 689 [16] Fuchs J, Tichopad A, Golub Y, Munz M, Schweitzer KJ, Wolf B, et al. Genetic variability in the SNCA  
690 gene influences  $\alpha$ -synuclein levels in the blood and brain. *The FASEB Journal* **22** (2008) 1327–1334.  
691 doi:https://doi.org/10.1096/fj.07-9348com.
- 692 [17] Krüger R, Kuhn W, Müller T, Woitalla D, Graeber M, Kösel S, et al. AlaSOPro mutation in the gene  
693 encoding  $\alpha$ -synuclein in Parkinson's disease. *Nature Genetics* **18** (1998) 106–108. doi:10.1038/ng0298-  
694 106.
- 695 [18] Appel-Cresswell S, Vilarino-Guell C, Encarnacion M, Sherman H, Yu I, Shah B, et al. Alpha-synuclein  
696 p.H50Q, a novel pathogenic mutation for Parkinson's disease. *Movement Disorders* **28** (2013) 811–813.  
697 doi:https://doi.org/10.1002/mds.25421.

- 698 [19] Pasanen P, Myllykangas L, Siitonen M, Raunio A, Kaakkola S, Lyytinen J, et al. A novel  $\alpha$ -synuclein  
699 mutation A53E associated with atypical multiple system atrophy and Parkinson's disease-type pathol-  
700 ogy. *Neurobiology of Aging* **35** (2014) 2180.e1–2180.e5. doi:10.1016/j.neurobiolaging.2014.03.024.
- 701 [20] Petrucci S, Ginevrino M, Valente EM. Phenotypic spectrum of alpha-synuclein mutations: New  
702 insights from patients and cellular models. *Parkinsonism & Related Disorders* **22** (2016) S16–S20.  
703 doi:10.1016/j.parkreldis.2015.08.015.
- 704 [21] Nguyen PH, Ramamoorthy A, Sahoo BR, Zheng J, Faller P, Straub JE, et al. Amyloid Oligomers:  
705 A Joint Experimental/Computational Perspective on Alzheimer's Disease, Parkinson's Disease,  
706 Type II Diabetes, and Amyotrophic Lateral Sclerosis. *Chemical Reviews* **121** (2021) 2545–2647.  
707 doi:10.1021/acs.chemrev.0c01122.
- 708 [22] Cremades N, Chen SW, Dobson CM. Chapter Three - Structural Characteristics of  $\alpha$ -Synuclein  
709 Oligomers. Sandal M, editor, *International Review of Cell and Molecular Biology* (Academic  
710 Press), *Early Stage Protein Misfolding and Amyloid Aggregation*, vol. 329 (2017), 79–143.  
711 doi:10.1016/bs.ircmb.2016.08.010.
- 712 [23] Iadanza MG, Jackson MP, Hewitt EW, Ranson NA, Radford SE. A new era for understanding  
713 amyloid structures and disease. *Nature Reviews Molecular Cell Biology* **19** (2018) 755–773.  
714 doi:10.1038/s41580-018-0060-8.
- 715 [24] Guerrero-Ferreira R, Taylor NM, Arteni AA, Kumari P, Mona D, Ringler P, et al. Two new polymorphic  
716 structures of human full-length alpha-synuclein fibrils solved by cryo-electron microscopy. *eLife* **8**  
717 (2019) e48907. doi:10.7554/eLife.48907.
- 718 [25] Guerrero-Ferreira R, Kovacic L, Ni D, Stahlberg H. New insights on the structure of alpha-  
719 synuclein fibrils using cryo-electron microscopy. *Current Opinion in Neurobiology* **61** (2020) 89–95.  
720 doi:10.1016/j.conb.2020.01.014.
- 721 [26] Tuttle MD, Comellas G, Nieuwkoop AJ, Covell DJ, Berthold DA, Kloepper KD, et al. Solid-state  
722 NMR structure of a pathogenic fibril of full-length human  $\alpha$ -synuclein. *Nature Structural & Molecular*  
723 *Biology* **23** (2016) 409–415. doi:10.1038/nsmb.3194.
- 724 [27] Araki K, Yagi N, Aoyama K, Choong CJ, Hayakawa H, Fujimura H, et al. Parkinson's disease is  
725 a type of amyloidosis featuring accumulation of amyloid fibrils of  $\alpha$ -synuclein. *Proceedings of the*  
726 *National Academy of Sciences* **116** (2019) 17963–17969. doi:10.1073/pnas.1906124116.
- 727 [28] Schweighauser M, Shi Y, Tarutani A, Kametani F, Murzin AG, Ghetti B, et al. Structures of  $\alpha$ -synuclein  
728 filaments from multiple system atrophy. *Nature* **585** (2020) 464–469. doi:10.1038/s41586-020-2317-6.
- 729 [29] De Giorgi F, Laferrière F, Zinghirino F, Faggiani E, Lends A, Bertoni M, et al. Novel self-replicating  
730  $\alpha$ -synuclein polymorphs that escape ThT monitoring can spontaneously emerge and acutely spread in  
731 neurons. *Science Advances* **6** (2020) eabc4364. doi:10.1126/sciadv.abc4364.
- 732 [30] Antonschmidt L, Dervişoğlu R, Sant V, Movellan KT, Mey I, Riedel D, et al. Insights into the  
733 molecular mechanism of amyloid filament formation: Segmental folding of  $\alpha$ -synuclein on lipid  
734 membranes. *Science Advances* **7** (2021) eabg2174. doi:10.1126/sciadv.abg2174.
- 735 [31] Ghosh D, Singh PK, Sahay S, Jha NN, Jacob RS, Sen S, et al. Structure based aggregation studies  
736 reveal the presence of helix-rich intermediate during  $\alpha$ -Synuclein aggregation. *Scientific Reports* **5**  
737 (2015) 9228. doi:10.1038/srep09228.
- 738 [32] Bousset L, Pieri L, Ruiz-Arlandis G, Gath J, Jensen PH, Habenstein B, et al. Structural and func-  
739 tional characterization of two alpha-synuclein strains. *Nature Communications* **4** (2013) 2575.  
740 doi:10.1038/ncomms3575.

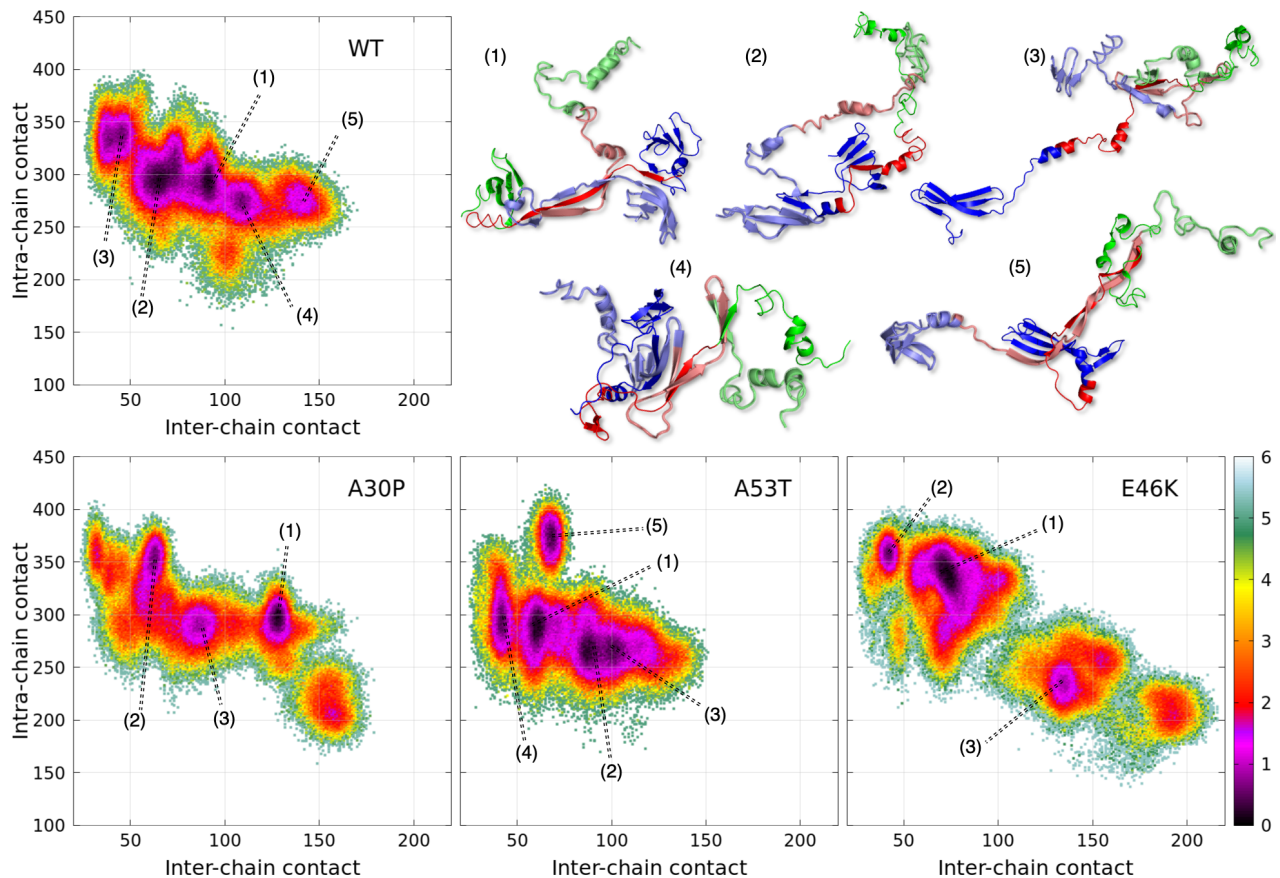
- 741 [33] Iadanza MG, Jackson MP, Hewitt EW, Ranson NA, Radford SE. A new era for understanding  
742 amyloid structures and disease. *Nature Reviews Molecular Cell Biology* **19** (2018) 755–773.  
743 doi:10.1038/s41580-018-0060-8.
- 744 [34] Winner B, Jappelli R, Maji SK, Desplats PA, Boyer L, Aigner S, et al. In vivo demonstration that  
745  $\alpha$ -synuclein oligomers are toxic. *Proceedings of the National Academy of Sciences* **108** (2011)  
746 4194–4199. doi:10.1073/pnas.1100976108.
- 747 [35] Conway KA, Lee SJ, Rochet JC, Ding TT, Williamson RE, Lansbury PT. Acceleration of oligomer-  
748 ization, not fibrillization, is a shared property of both alpha-synuclein mutations linked to early-onset  
749 Parkinson's disease: implications for pathogenesis and therapy. *Proceedings of the National Academy*  
750 *of Sciences of the United States of America* **97** (2000) 571–576. doi:10.1073/pnas.97.2.571.
- 751 [36] Lázaro DF, Rodrigues EF, Langohr R, Shahpasandzadeh H, Ribeiro T, Guerreiro P, et al. Systematic  
752 Comparison of the Effects of Alpha-synuclein Mutations on Its Oligomerization and Aggregation.  
753 *PLOS Genetics* **10** (2014) e1004741. doi:10.1371/journal.pgen.1004741.
- 754 [37] Volles MJ, Lee SJ, Rochet JC, Shtilerman MD, Ding TT, Kessler JC, et al. Permeabilization by  
755 protofibrillar alpha-synuclein: implications for the pathogenesis and treatment of Parkinson's disease.  
756 *Biochemistry* **40** (2001) 7812–7819. doi:10.1021/bi0102398.
- 757 [38] Fredenburg RA, Rospigliosi C, Meray RK, Kessler JC, Lashuel HA, Eliezer D, et al. The impact  
758 of the E46K mutation on the properties of alpha-synuclein in its monomeric and oligomeric states.  
759 *Biochemistry* **46** (2007) 7107–7118. doi:10.1021/bi7000246.
- 760 [39] Giehm L, Svergun DI, Otzen DE, Vestergaard B. Low-resolution structure of a vesicle disrupting  
761  $\alpha$ -synuclein oligomer that accumulates during fibrillation. *Proceedings of the National Academy of*  
762 *Sciences* **108** (2011) 3246–3251. doi:10.1073/pnas.1013225108.
- 763 [40] Lorenzen N, Nielsen SB, Buell AK, Kaspersen JD, Arosio P, Vad BS, et al. The Role of Stable  
764  $\alpha$ -Synuclein Oligomers in the Molecular Events Underlying Amyloid Formation. *Journal of the*  
765 *American Chemical Society* **136** (2014) 3859–3868. doi:10.1021/ja411577t.
- 766 [41] Cremades N, Cohen SA, Deas E, Abramov A, Chen A, Orte A, et al. Direct Observation of  
767 the Interconversion of Normal and Toxic Forms of  $\alpha$ -Synuclein. *Cell* **149** (2012) 1048–1059.  
768 doi:10.1016/j.cell.2012.03.037.
- 769 [42] Tosatto L, Horrocks MH, Dear AJ, Knowles TPJ, Dalla Serra M, Cremades N, et al. Single-molecule  
770 FRET studies on alpha-synuclein oligomerization of Parkinson's disease genetically related mutants.  
771 *Scientific Reports* **5** (2015) 16696. doi:10.1038/srep16696.
- 772 [43] Li X, Dong C, Hoffmann M, Garen CR, Cortez LM, Petersen NO, et al. Early stages of aggregation  
773 of engineered  $\alpha$ -synuclein monomers and oligomers in solution. *Scientific Reports* **9** (2019) 1734.  
774 doi:10.1038/s41598-018-37584-6.
- 775 [44] Zhou L, Kurouski D. Structural Characterization of Individual  $\alpha$ -Synuclein Oligomers Formed  
776 at Different Stages of Protein Aggregation by Atomic Force Microscopy-Infrared Spectroscopy.  
777 *Analytical Chemistry* **92** (2020) 6806–6810. doi:10.1021/acs.analchem.0c00593.
- 778 [45] Bhak G, Lee S, Kim TH, Lee JH, Yang JE, Joo K, et al. Morphological Evaluation of Meta-stable  
779 Oligomers of  $\alpha$ -Synuclein with Small-Angle Neutron Scattering. *Scientific Reports* **8** (2018) 14295.  
780 doi:10.1038/s41598-018-32655-0.
- 781 [46] Horrocks MH, Tosatto L, Dear AJ, Garcia GA, Iljina M, Cremades N, et al. Fast Flow Microfluidics and  
782 Single-Molecule Fluorescence for the Rapid Characterization of  $\alpha$ -Synuclein Oligomers. *Analytical*  
783 *Chemistry* **87** (2015) 8818–8826. doi:10.1021/acs.analchem.5b01811.

- 784 [47] Lashuel HA, Petre BM, Wall J, Simon M, Nowak RJ, Walz T, et al. Alpha-synuclein, especially the  
785 Parkinson's disease-associated mutants, forms pore-like annular and tubular protofibrils. *Journal of*  
786 *Molecular Biology* **322** (2002) 1089–1102. doi:10.1016/s0022-2836(02)00735-0.
- 787 [48] Chen SW, Drakulic S, Deas E, Ouberai M, Aprile FA, Arranz R, et al. Structural characterization of  
788 toxic oligomers that are kinetically trapped during  $\alpha$ -synuclein fibril formation. *Proceedings of the*  
789 *National Academy of Sciences* **112** (2015) E1994–E2003. doi:10.1073/pnas.1421204112.
- 790 [49] Laganowsky A, Liu C, Sawaya MR, Whitelegge JP, Park J, Zhao M, et al. Atomic View of a Toxic  
791 Amyloid Small Oligomer. *Science* (2012). doi:10.1126/science.1213151.
- 792 [50] Mao X, Ou MT, Karuppagounder SS, Kam TI, Yin X, Xiong Y, et al. Pathological  $\alpha$ -synuclein  
793 transmission initiated by binding lymphocyte-activation gene 3. *Science* **353** (2016) aah3374.  
794 doi:10.1126/science.aah3374.
- 795 [51] Wang W, Perovic I, Chittuluru J, Kaganovich A, Nguyen LTT, Liao J, et al. A soluble  $\alpha$ -synuclein  
796 construct forms a dynamic tetramer. *Proceedings of the National Academy of Sciences* **108** (2011)  
797 17797–17802. doi:10.1073/pnas.1113260108.
- 798 [52] Bartels T, Choi JG, Selkoe DJ.  $\alpha$ -Synuclein occurs physiologically as a helically folded tetramer that  
799 resists aggregation. *Nature* **477** (2011) 107–110. doi:10.1038/nature10324.
- 800 [53] Lucas H, Fernández R. Navigating the dynamic landscape of alpha-synuclein morphology: a review  
801 of the physiologically relevant tetrameric conformation. *Neural Regeneration Research* **15** (2020) 407.  
802 doi:10.4103/1673-5374.265792.
- 803 [54] Dettmer U, Newman AJ, von Saucken VE, Bartels T, Selkoe D. KTKEGV repeat motifs are key  
804 mediators of normal  $\alpha$ -synuclein tetramerization: Their mutation causes excess monomers and neuro-  
805 toxicity. *Proceedings of the National Academy of Sciences of the United States of America* **112** (2015)  
806 9596–9601. doi:10.1073/pnas.1505953112.
- 807 [55] Cote Y, Delarue P, Scheraga HA, Senet P, Maisuradze GG. From a Highly Disordered to a Metastable  
808 State: Uncovering Insights of  $\alpha$ -Synuclein. *ACS Chemical Neuroscience* **9** (2018) 1051–1065.  
809 doi:10.1021/acschemneuro.7b00446.
- 810 [56] Gurry T, Ullman O, Fisher CK, Perovic I, Pochapsky T, Stultz CM. The Dynamic Structure  
811 of  $\alpha$ -Synuclein Multimers. *Journal of the American Chemical Society* **135** (2013) 3865–3872.  
812 doi:10.1021/ja310518p.
- 813 [57] Krishnan S, Chi EY, Wood SJ, Kendrick BS, Li C, Garzon-Rodriguez W, et al. Oxidative Dimer  
814 Formation Is the Critical Rate-Limiting Step for Parkinson's Disease  $\alpha$ -Synuclein Fibrillogenesis.  
815 *Biochemistry* **42** (2003) 829–837. doi:10.1021/bi026528t.
- 816 [58] Lv Z, Krasnoslobodtsev AV, Zhang Y, Ysselstein D, Rochet JC, Blanchard SC, et al. Direct Detection  
817 of  $\alpha$ -Synuclein Dimerization Dynamics: Single-Molecule Fluorescence Analysis. *Biophysical Journal*  
818 **108** (2015) 2038–2047. doi:10.1016/j.bpj.2015.03.010.
- 819 [59] Yu J, Malkova S, Lyubchenko YL.  $\alpha$ -Synuclein Misfolding: Single Molecule AFM Force Spectroscopy  
820 Study. *Journal of Molecular Biology* **384** (2008) 992–1001. doi:10.1016/j.jmb.2008.10.006.
- 821 [60] Zhang Y, Hashemi M, Lv Z, Williams B, Popov KI, Dokholyan NV, et al. High-speed atomic  
822 force microscopy reveals structural dynamics of  $\alpha$ -synuclein monomers and dimers. *The Journal of*  
823 *Chemical Physics* **148** (2018) 123322. doi:10.1063/1.5008874.
- 824 [61] Roostae A, Beaudoin S, Staskevicius A, Roucou X. Aggregation and neurotoxicity of recombi-  
825 nant  $\alpha$ -synuclein aggregates initiated by dimerization. *Molecular Neurodegeneration* **8** (2013) 5.  
826 doi:10.1186/1750-1326-8-5.

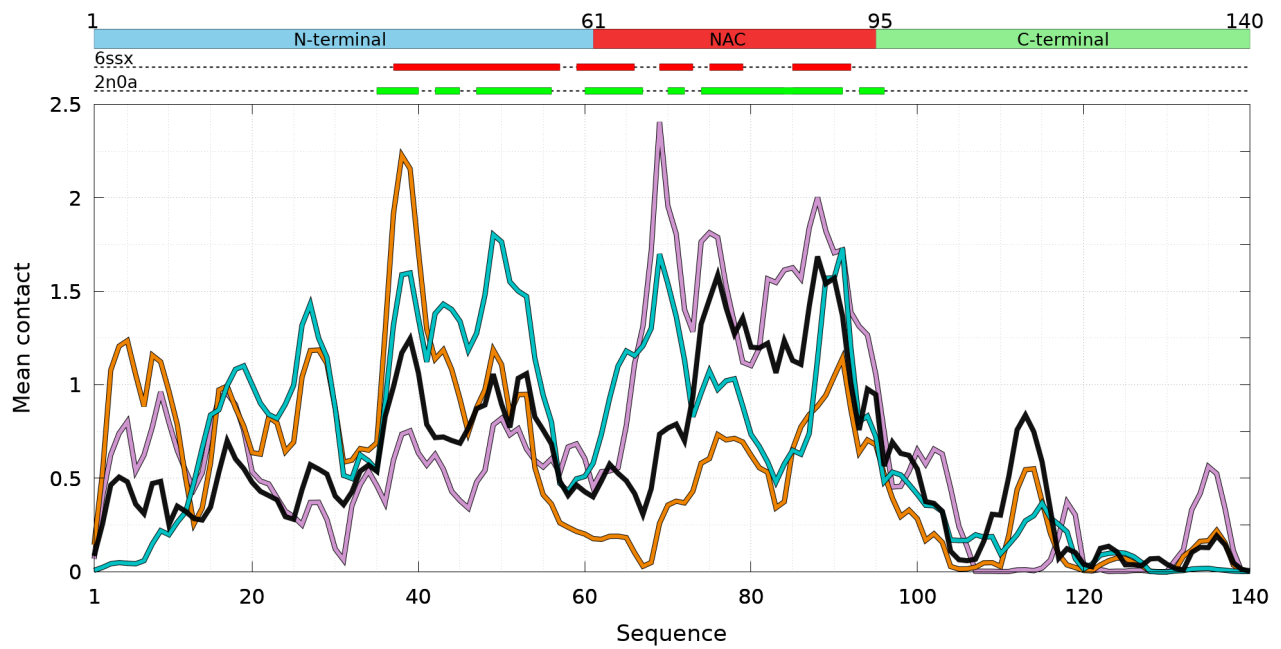
- 827 [62] Krasnoslobodtsev AV, Volkov IL, Asiago JM, Hindupur J, Rochet JC, Lyubchenko YL.  $\alpha$ -Synuclein  
828 Misfolding Assessed with Single Molecule AFM Force Spectroscopy: Effect of Pathogenic Mutations.  
829 *Biochemistry* **52** (2013) 7377–7386. doi:10.1021/bi401037z.
- 830 [63] Fusco G, Sanz-Hernandez M, De Simone A. Order and disorder in the physiological membrane binding  
831 of  $\alpha$ -synuclein. *Current Opinion in Structural Biology* **48** (2018) 49–57. doi:10.1016/j.sbi.2017.09.004.
- 832 [64] Yamauchi M, Okumura H. Dimerization of  $\alpha$ -Synuclein Fragments Studied by Isothermal–Isobaric  
833 Replica-Permutation Molecular Dynamics Simulation. *Journal of Chemical Information and Modeling*  
834 **61** (2021) 1307–1321. doi:10.1021/acs.jcim.0c01056.
- 835 [65] Sahu KK, Woodside MT, Tuszynski JA.  $\alpha$ -synuclein dimer structures found from computational  
836 simulations. *Biochimie* **116** (2015) 133–140. doi:10.1016/j.biochi.2015.07.011.
- 837 [66] Maisuradze GG, Senet P, Czaplewski C, Liwo A, Scheraga HA. Investigation of Protein Folding  
838 by Coarse-Grained Molecular Dynamics with the UNRES Force Field. *The Journal of Physical*  
839 *Chemistry A* **114** (2010) 4471–4485. doi:10.1021/jp9117776.
- 840 [67] Liwo A, Sieradzan AK, Lipska AG, Czaplewski C, Joung I, Żmudzińska W, et al. A general method  
841 for the derivation of the functional forms of the effective energy terms in coarse-grained energy  
842 functions of polymers. III. Determination of scale-consistent backbone-local and correlation potentials  
843 in the UNRES force field and force-field calibration and validation. *The Journal of Chemical Physics*  
844 **150** (2019) 155104. doi:10.1063/1.5093015.
- 845 [68] Khalili M, Liwo A, Jagielska A, Scheraga HA. Molecular Dynamics with the United-Residue Model  
846 of Polypeptide Chains. II. Langevin and Berendsen-Bath Dynamics and Tests on Model  $\alpha$ -Helical  
847 Systems. *The Journal of Physical Chemistry B* **109** (2005) 13798–13810. doi:10.1021/jp058007w.
- 848 [69] Zhou R, Maisuradze GG, Suñol D, Todorovski T, Macias MJ, Xiao Y, et al. Folding kinetics of  
849 WW domains with the united residue force field for bridging microscopic motions and experimen-  
850 tal measurements. *Proceedings of the National Academy of Sciences* **111** (2014) 18243–18248.  
851 doi:10.1073/pnas.1420914111.
- 852 [70] Sieradzan AK, Korneev A, Begun A, Kachlishvili K, Scheraga HA, Molochkov A, et al. Investiga-  
853 tion of Phosphorylation-Induced Folding of an Intrinsically Disordered Protein by Coarse-Grained  
854 Molecular Dynamics. *Journal of Chemical Theory and Computation* **17** (2021) 3203–3220.  
855 doi:10.1021/acs.jctc.1c00155.
- 856 [71] Gołaś E, Maisuradze GG, Senet P, Ołdziej S, Czaplewski C, Scheraga HA, et al. Simulation of the  
857 Opening and Closing of Hsp70 Chaperones by Coarse-Grained Molecular Dynamics. *Journal of*  
858 *Chemical Theory and Computation* **8** (2012) 1750–1764. doi:10.1021/ct200680g.
- 859 [72] Guzzo A, Delarue P, Rojas A, Nicolai A, Maisuradze GG, Senet P. Missense Mutations Modify the  
860 Conformational Ensemble of the  $\alpha$ -Synuclein Monomer Which Exhibits a Two-Phase Characteristic.  
861 *Frontiers in Molecular Biosciences* **8** (2021) 1104. doi:10.3389/fmolb.2021.786123.
- 862 [73] Liwo A, Czaplewski C, Pillardy J, Scheraga HA. Cumulant-based expressions for the multibody terms  
863 for the correlation between local and electrostatic interactions in the united-residue force field. *The*  
864 *Journal of Chemical Physics* **115** (2001) 2323–2347. doi:10.1063/1.1383989.
- 865 [74] Kabsch W, Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-  
866 bonded and geometrical features. *Biopolymers* **22** (1983) 2577–2637. doi:10.1002/bip.360221211.
- 867 [75] Touw WG, Baakman C, Black J, te Beek TA, Krieger E, Joosten RP, et al. A series of  
868 PDB-related databanks for everyday needs. *Nucleic Acids Research* **43** (2015) D364–D368.  
869 doi:10.1093/nar/gku1028.
- 870 [76] Chou PY, Fasman GD. Conformational parameters for amino acids in helical,  $\beta$ -sheet, and random  
871 coil regions calculated from proteins. *Biochemistry* **13** (1974) 211–222. doi:10.1021/bi00699a001.

- 872 [77] Reynolds D. Gaussian Mixture Models. Li SZ, Jain A, editors, *Encyclopedia of Biometrics* (Boston,  
873 MA: Springer US) (2009), 659–663. doi:10.1007/978-0-387-73003-5\_196.
- 874 [78] Araki K, Yagi N, Nakatani R, Sekiguchi H, So M, Yagi H, et al. A small-angle X-ray scatter-  
875 ing study of alpha-synuclein from human red blood cells. *Scientific Reports* **6** (2016) 30473.  
876 doi:10.1038/srep30473.
- 877 [79] Dai X, Fu W, Chi H, Mesias VSD, Zhu H, Leung CW, et al. Optical tweezers-controlled hotspot for  
878 sensitive and reproducible surface-enhanced Raman spectroscopy characterization of native protein  
879 structures. *Nature Communications* **12** (2021) 1292. doi:10.1038/s41467-021-21543-3.
- 880 [80] Hoefling M, Lima N, Haenni D, Seidel CAM, Schuler B, Grubmüller H. Structural Heterogeneity and  
881 Quantitative FRET Efficiency Distributions of Polyprolines through a Hybrid Atomistic Simulation  
882 and Monte Carlo Approach. *PLOS ONE* **6** (2011) e19791. doi:10.1371/journal.pone.0019791.
- 883 [81] dos Remedios CG, Moens PDJ. Fluorescence Resonance Energy Transfer Spectroscopy Is a Reliable  
884 "Ruler" for Measuring Structural Changes in Proteins: Dispelling the Problem of the Unknown  
885 Orientation Factor. *Journal of Structural Biology* **115** (1995) 175–185. doi:10.1006/jsbi.1995.1042.
- 886 [82] Nicolai A, Delarue P, Senet P. Conformational dynamics of full-length inducible human Hsp70 derived  
887 from microsecond molecular dynamics simulations in explicit solvent. *Journal of Biomolecular  
888 Structure and Dynamics* **31** (2013) 1111–1126. doi:10.1080/07391102.2012.726190.
- 889 [83] Conway KA, Harper JD, Lansbury PT. Accelerated in vitro fibril formation by a mutant alpha-synuclein  
890 linked to early-onset Parkinson disease. *Nature Medicine* **4** (1998) 1318–1320. doi:10.1038/3311.
- 891 [84] Greenbaum EA, Graves CL, Mishizen-Eberz AJ, Lupoli MA, Lynch DR, Englander SW, et al. The  
892 E46K Mutation in  $\alpha$ -Synuclein Increases Amyloid Fibril Formation \*. *Journal of Biological Chemistry*  
893 **280** (2005) 7800–7807. doi:10.1074/jbc.M411638200.
- 894 [85] Waxman EA, Mazzulli JR, Giasson BI. Characterization of Hydrophobic Residue Requirements for  
895  $\alpha$ -Synuclein Fibrillization. *Biochemistry* **48** (2009) 9427–9436. doi:10.1021/bi900539p.
- 896 [86] Doherty CPA, Ulamec SM, Maya-Martinez R, Good SC, Makepeace J, Khan GN, et al. A short motif  
897 in the N-terminal region of  $\alpha$ -synuclein is critical for both aggregation and function. *Nature Structural  
898 & Molecular Biology* **27** (2020) 249–259. doi:10.1038/s41594-020-0384-x.
- 899 [87] Tripathi T. A Master Regulator of  $\alpha$ -Synuclein Aggregation. *ACS chemical neuroscience* **11** (2020)  
900 1376–1378. doi:10.1021/acscemneuro.0c00216.

## FIGURES AND TABLES

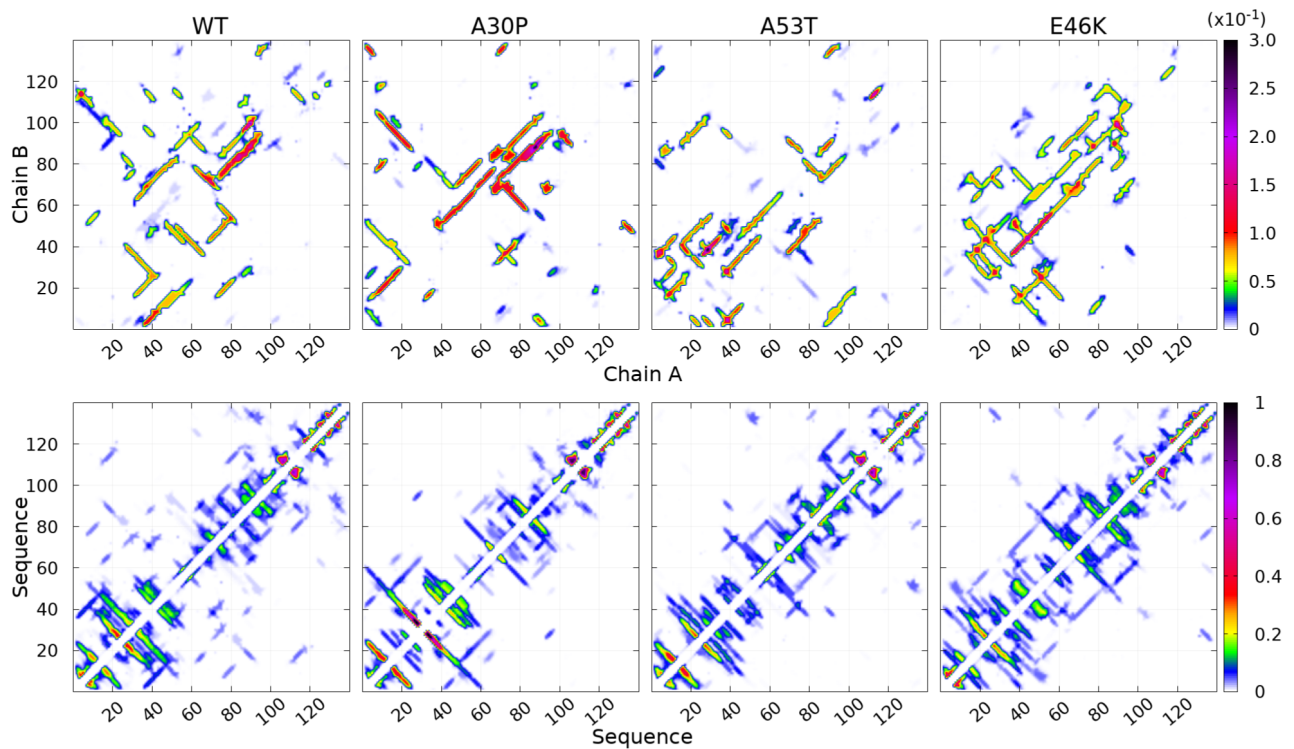


**Figure 1.** Color maps of  $-\ln \left[ \frac{P(n_{inter}, n_{intra})}{P_{max}} \right]$  computed from the 2D probability density function  $P$  of the number of inter-chain ( $n_{inter}$ ) and intra-chain ( $n_{intra}$ ) contacts between the residues of  $\alpha$ -syn for the complete ensemble of dimers found in molecular dynamics trajectories at 300 K and 310 K of the WT and variants. In each map,  $P_{max}$  is the maximum value of the probability of the map. The local minima within one unit from the global minimum of each map are numbered. Examples of 3D structures associated to each local minimum of WT are represented with N-terminus in blue, NAC in red and the C-term in green. Light and dark colors differentiate the structure of the two monomers within the dimers.

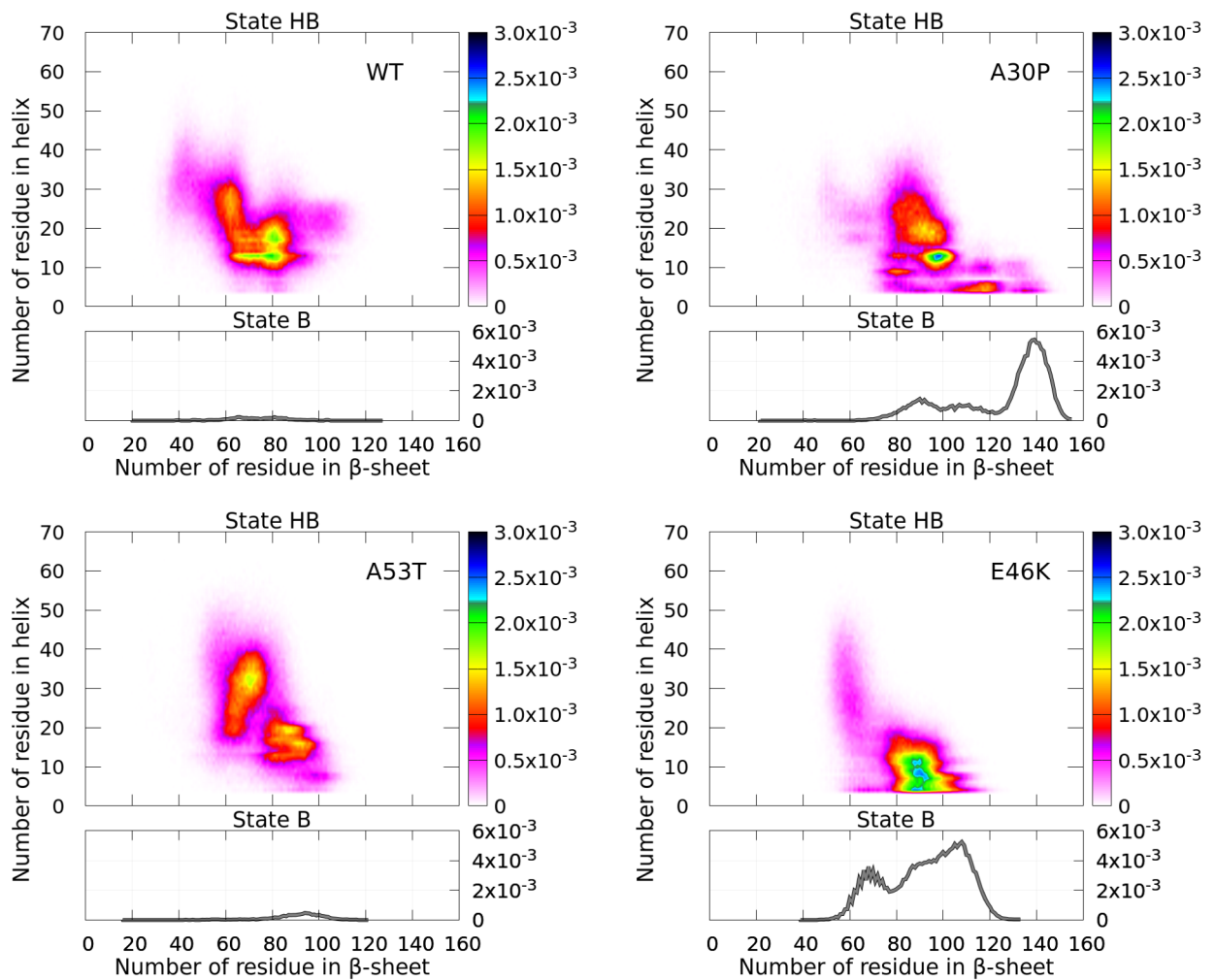


**Figure 2.** . Mean inter-molecular contact of a residue of  $\alpha$ -syn WT and mutants for the complete ensemble of dimers found in molecular dynamics trajectories at 300 K and 310 K. The color code is black (WT), violet (A30P), orange (A53T), turquoise (E46K). The upper bar represents the N-terminal (blue), NAC (red) and C-terminal (green) regions of the sequence. The red and green rectangles on the upper dotted bars describe the locations of the inter-molecular  $\beta$ -sheets identified by CUTABI in the experimental structures of  $\alpha$ -syn fibrils with the PDB IDs: 6ssx and 2n0a, respectively.

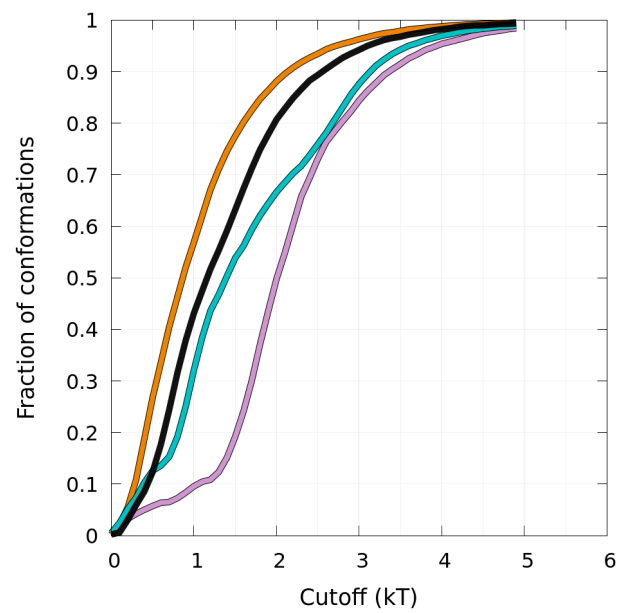




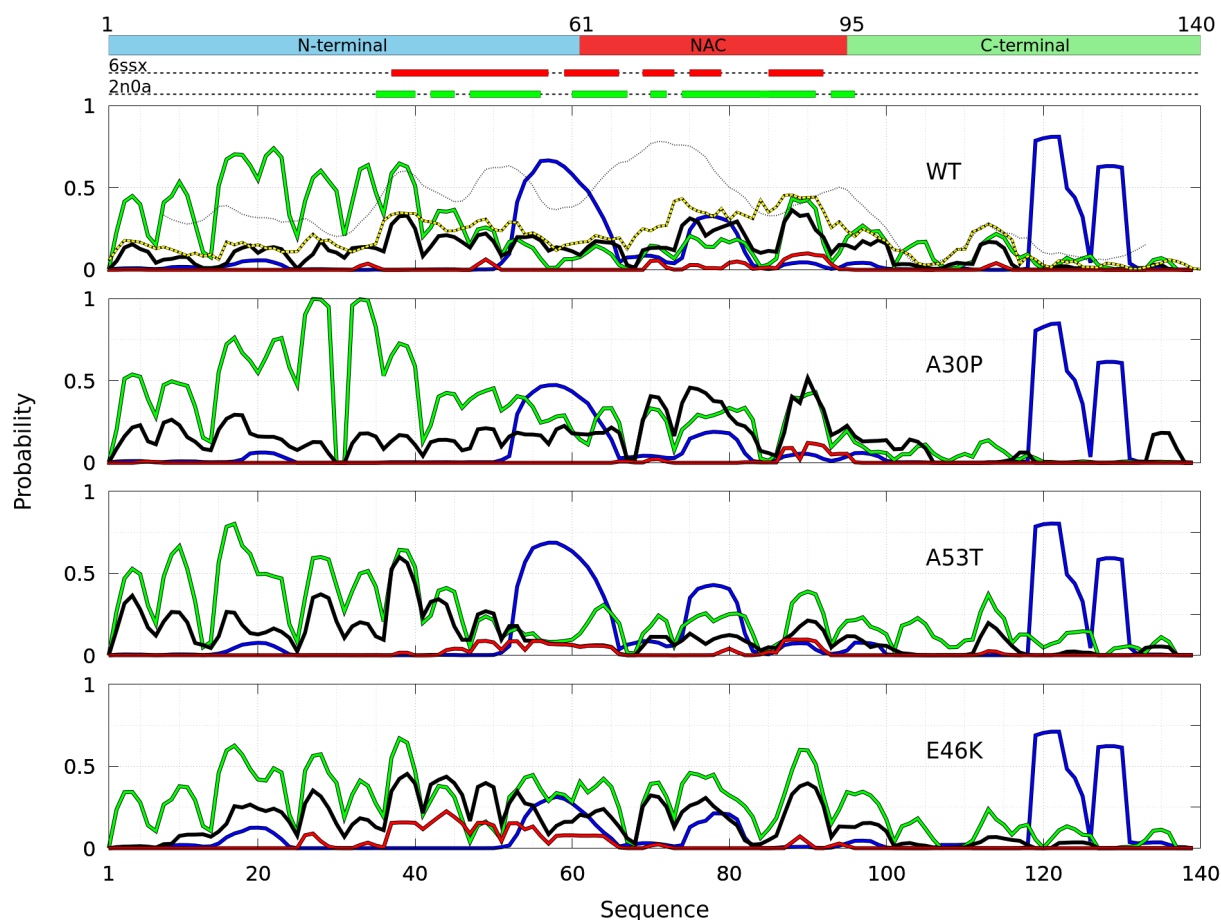
**Figure 3.** Contact maps computed for the complete ensemble of  $\alpha$ -syn dimers found in molecular dynamics simulations for the WT and the mutants. The upper panels show inter-molecular contacts between chains A and B of the dimers and the lower panels show intra-molecular contacts within the monomers A+B. The color bars indicate the probability. Note the different scales for intra- and inter-molecular contact maps.



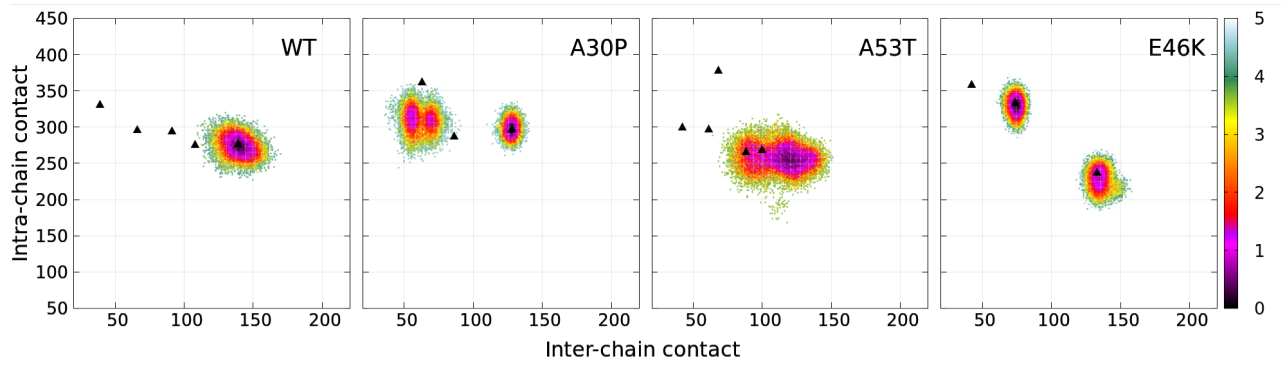
**Figure 4.** Probability density of the number of residues in  $\alpha$ -helix and  $\beta$ -sheet for WT and mutants of  $\alpha$ -syn. The probability density of the state B (no helix) is represented by a function (gray) (right vertical axis) and the probability density of state HB is represented by a two-dimensional map (right color bar).



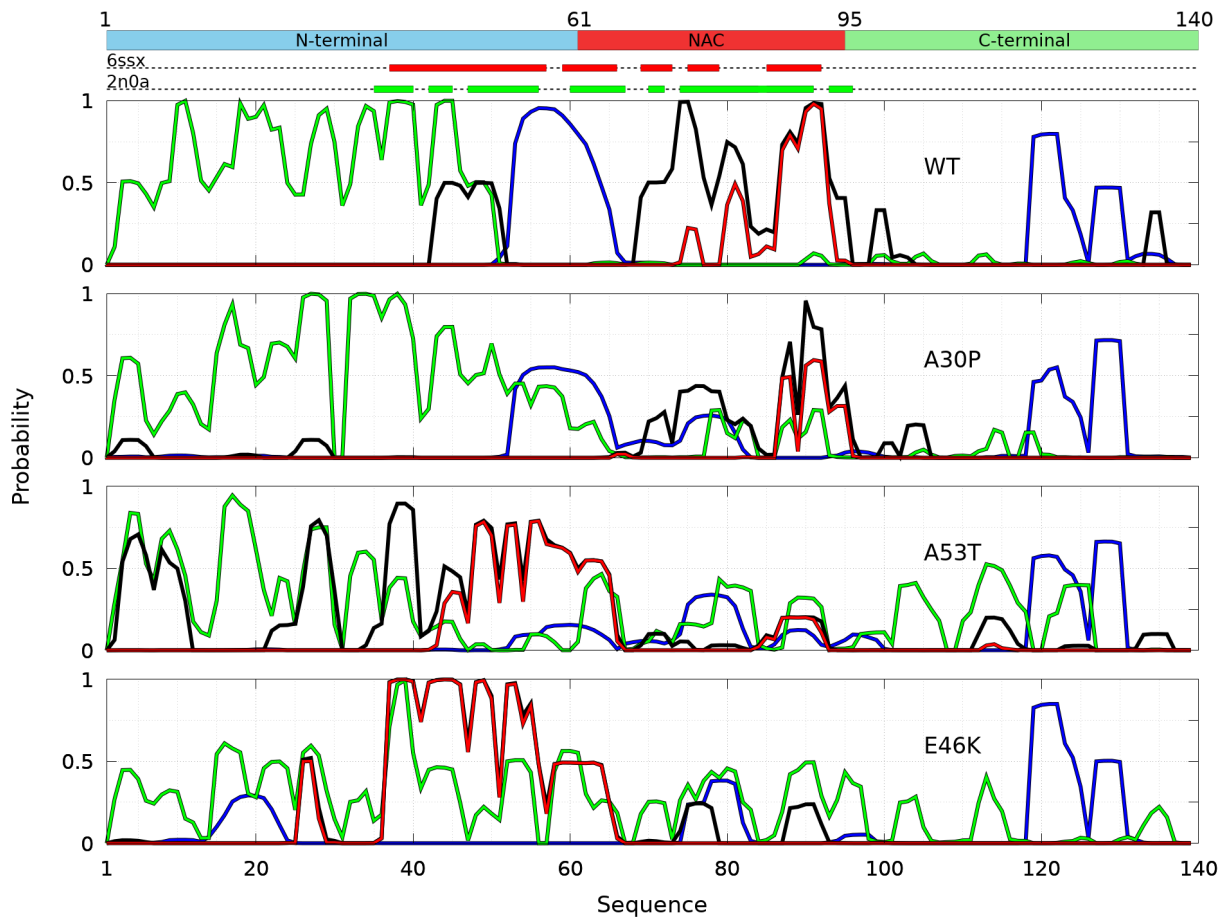
**Figure 5.** Cumulative fraction of the total number of conformations simulated as a function of an effective free-energy cut-off from the global minimum of the B state (see text) represented in Figure 4. Curves are for WT (black), A30P (purple), E46K (turquoise) and A53T (orange).



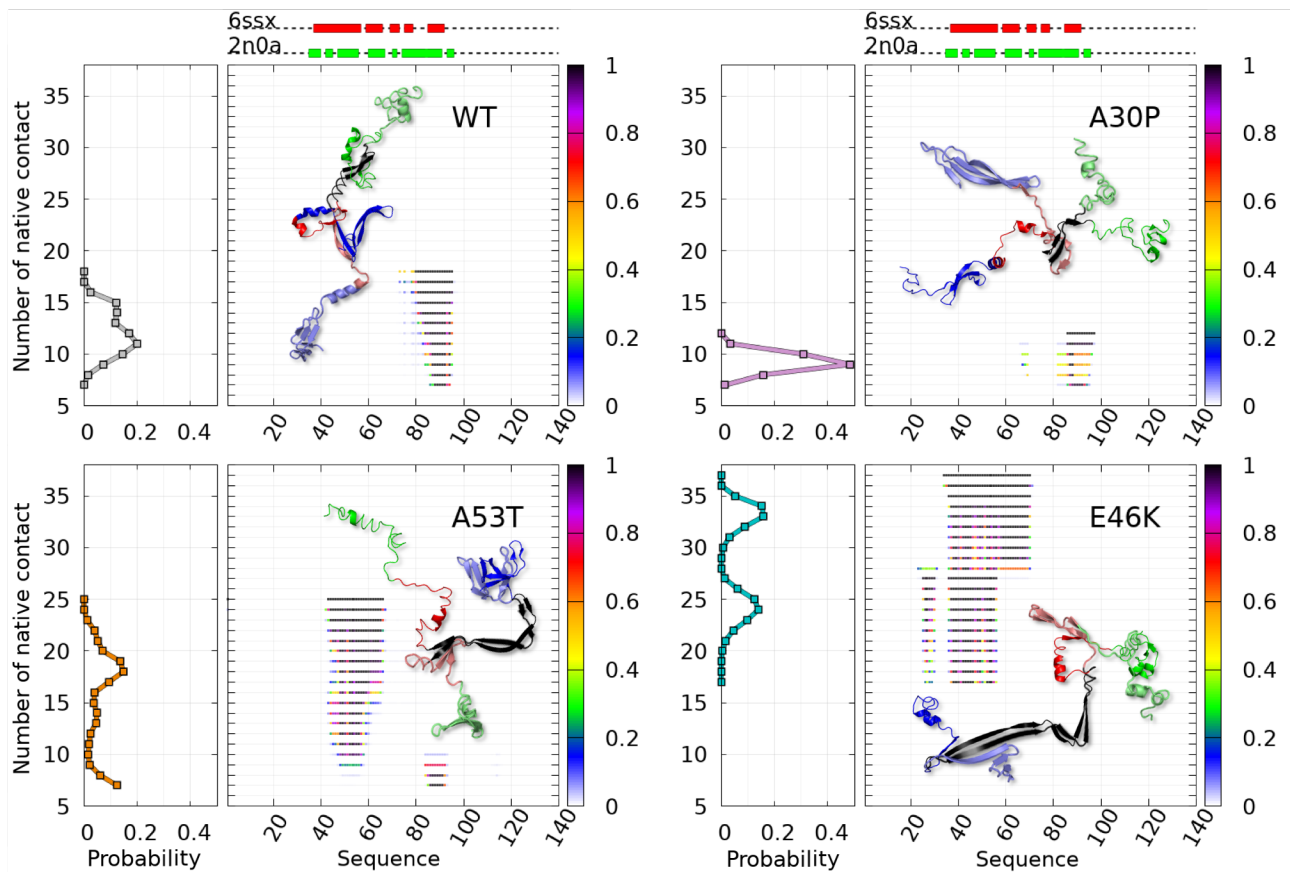
**Figure 6.** Analysis of the secondary structures for the complete ensemble of  $\alpha$ -syn dimers found in molecular dynamics simulations for the WT and the mutants using CUTABI algorithm[72]. Probabilities for each residue to pertain to an  $\alpha$ -helix (blue), to an intra-molecular  $\beta$ -sheet (green), to an inter-molecular  $\beta$ -sheet (black) and to a native inter-molecular  $\beta$ -sheet fibril-type (red). The probability for each residue to form an inter-molecular contact (Figure 2) is shown for the WT (black/yellow broken line) for comparison. The propensity to form a  $\beta$ -sheet calculated from the empirical scale of Chou and Fasman with the ProtScale online tool (<https://web.expasy.org/protscale/>) using a window of 15 residues and a linear interpolation with a weight of 0.10 for the window edges is also represented (dotted line) for the WT. The red and green rectangles on the upper dotted bars describe the locations of the inter-molecular  $\beta$ -sheets identified by CUTABI in the experimental structures of  $\alpha$ -syn fibrils with the PDB IDs: 6ssx and 2n0a, respectively.



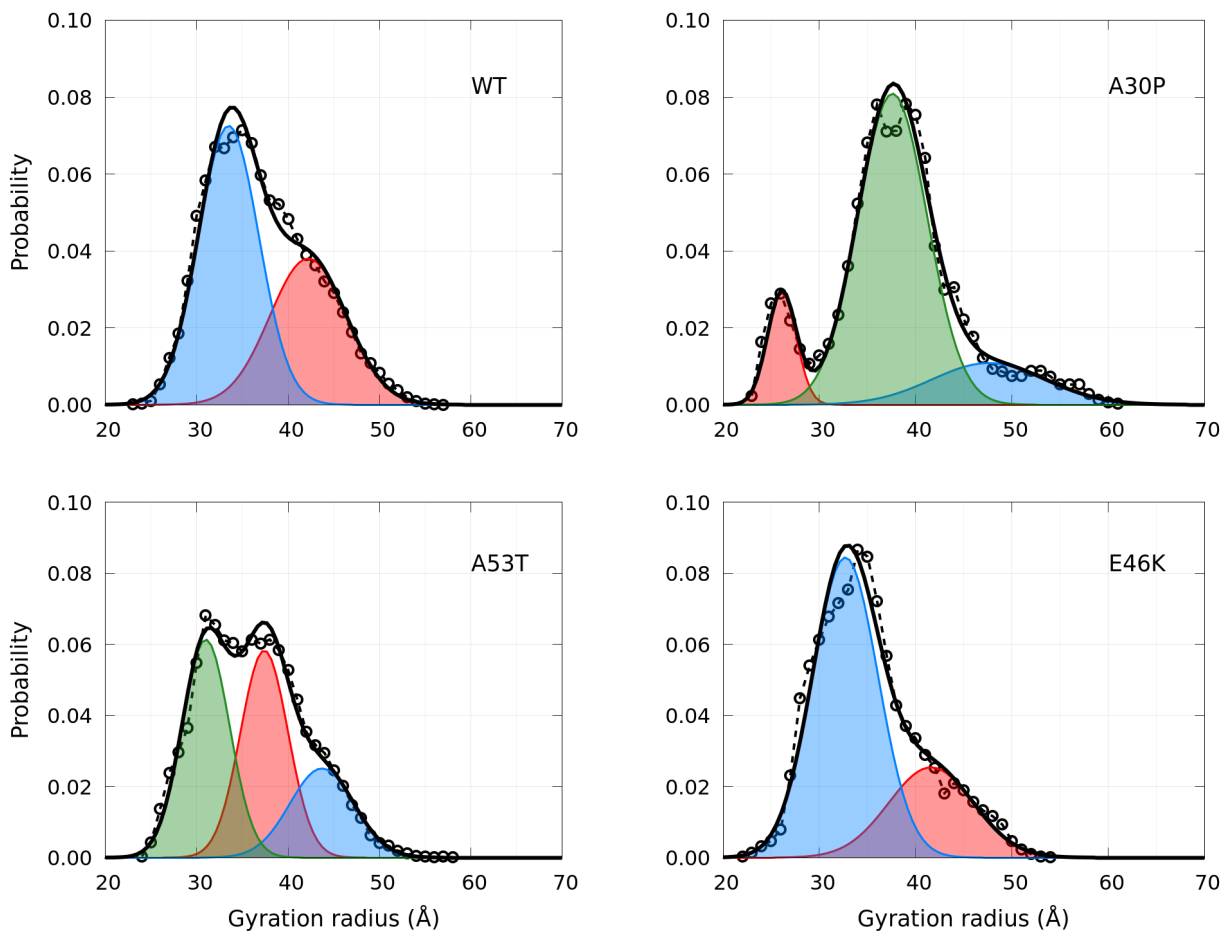
**Figure 7.** Color maps of  $-\ln \left[ \frac{P(n_{inter}, n_{intra})}{P_{max}} \right]$  computed from the 2D probability density function  $P$  of the number of inter-chain ( $n_{inter}$ ) and intra-chain ( $n_{intra}$ ) contacts between the residues of  $\alpha$ -syn for the sub-ensemble of dimers with fibril native contacts found in molecular dynamics trajectories at 300 K and 310 K of the WT and variants. In each map,  $P_{max}$  is the maximum value of the probability of the map. The local minima numbered in Figure 1 are shown by black triangles.



**Figure 8.** Analysis of the secondary structures in the sub-ensemble of  $\alpha$ -syn dimers with native fibril-type contacts found in molecular dynamics simulations for the WT and the mutants at 300 K and 310 K. Probabilities for each residue to pertain to an  $\alpha$ -helix (blue), to an intra-molecular  $\beta$ -sheet (green), to an inter-molecular  $\beta$ -sheet (black) and to a native inter-molecular  $\beta$ -sheet fibril-type (red).

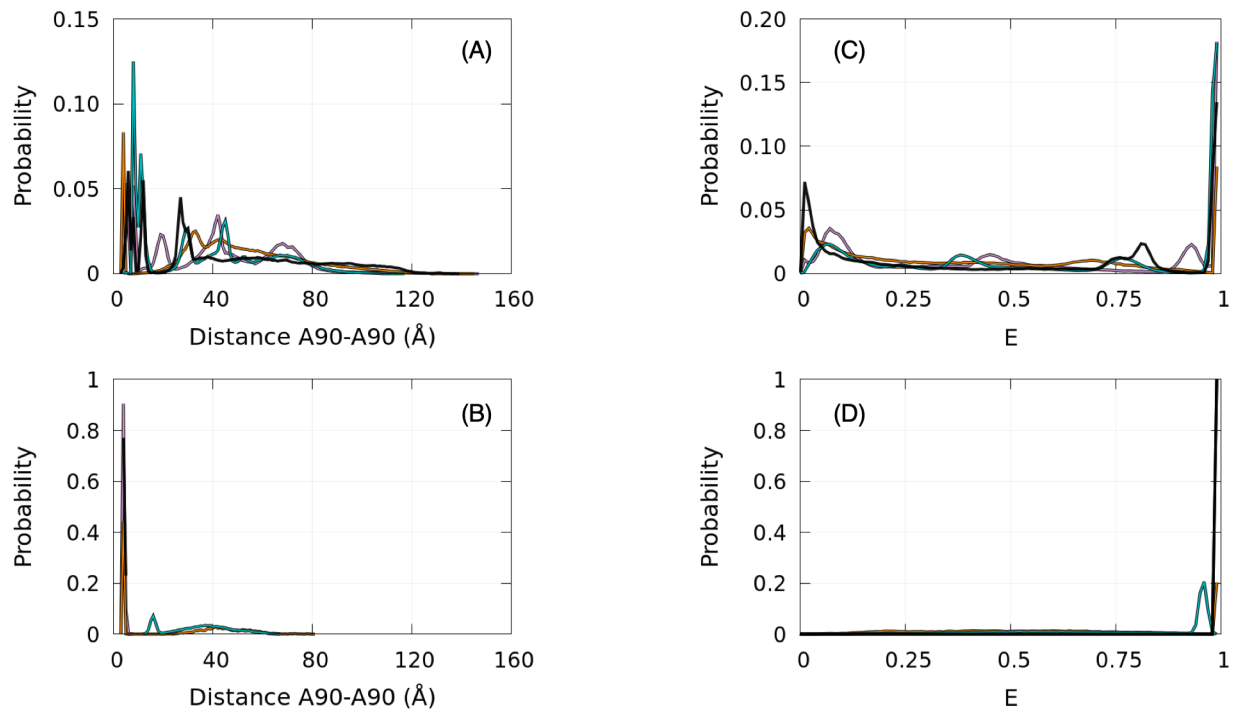


**Figure 9.** Analysis of the mean contact probability along the amino-acid sequence for the sub-ensemble of Dfnc found in MD trajectories for the WT and variants. For each protein, the left panel represents the probability density of the number of contacts and the left panel its distribution along the sequence for each point of the distribution with a color code giving the probability at each residue according to the color bar on the right. 3D structures representative of the maximum number of native fibril-type contacts are shown with the following color: N-terminal (blue), NAC (red), C-terminal (green) and native contact region (black)



**Figure 10.** . Probability density functions of the gyration radius for the WT and mutants (circle symbols and dashed lines) computed from MD simulations. Each distribution is represented by a set of Gaussians (colored areas) the number and the parameters of which were found by applying the GMM algorithm. The sum of gaussians is represented by a full black line.





**Figure 11.** Probability density functions of the distance between the  $C^\alpha$  atoms of residues A90 in the ensembles of disordered dimers (panel A) and of Dfnc (panel B) and the corresponding probability distribution functions of the approximated FRET efficiency E (see text) for the ensembles of disordered dimers (panel C) and of Dfnc (panel D). Color code is WT (black), A30P (purple), E46K (turquoise) and A53T (orange)

Protein	min 2	min 3	min 4	min 5
WT	0.08	0.39	0.40	0.84
A30P	0.42	0.54	-	-
A53T	0.04	0.12	0.14	0.24
E46K	0.32	0.72	-	-

**Table 1.** Effective (dimensionless) free-energy difference ( $-\ln \left[ \frac{P_1}{P_i} \right]$ ) where  $P_1$  and  $P_i$  are the probabilities of the minimum 1 and the of  $i$ th minima shown in Figure 1 for the WT and the variants .

Protein	Dimer		Dfnc		Monomer	
	Parallel	Anti-parallel	Parallel	Anti-parallel	Parallel	Anti-parallel
WT	16.5	27.9	29.1	28.7	10.6	28.4
A30P	18.2	35.1	21.0	27.0	10.9	31.1
A53T	17.1	29.4	22.8	30.8	11.0	22.4
E46K	18.9	31.8	23.1	30.0	10.3	25.8

**Table 2.** Percentages of residues in parallel and anti-parallel intra- and inter-molecular  $\beta$ -sheets computed with CUTABI for the ensemble of dimers, the sub-ensemble of Dfnc and for isolated monomers.

Protein	Cluster 1	Cluster 2	Cluster 3	Average value
WT	33.5 Å (60%)	42.2 Å (40%)	-	37.0 Å (100 %)
A30P	37.6 Å (73%)	47.6 Å (16%)	26.1 Å (11%)	37.9 Å (100 %)
A53T	31.1 Å (41%)	37.6 Å (38%)	43.8 Å (21%)	36.2 Å (100 %)
E46K	32.7 Å (72%)	41.7 Å (28%)	-	35.2 Å (100 %)

**Table 3.** Clustering of the gyration radius probability density using the GMM algorithm. The values in brackets are the corresponding % of the ensemble of the conformations.

Protein	Mean contact	Inter-molecular $\beta$ -sheet (all dimers)	Nfc in Dfnc
WT	I88	I88	T75, A76, K80, T81, V82, A85, S87-G93
A30P	V70	A90	S87, I88, A90-V95
A53T	L38	L38	T44-N65, S87-T92
E46K	V49	Y39	V26-E28, V37-N65

**Table 4.** Main residues or segments identified by MD as important for the dimerization of  $\alpha$ -syn from the maxima of propensity for the mean contact, inter-molecular  $\beta$ -sheets and Nfc.



# Chapitre 5

## Résultats : Etudes du segment 38-53 de l' $\alpha$ -synucléine

### 5.1 Comportement des angles diédraux des résidus du segment

Comme explicité dans l'introduction la région 38-53 de la séquence présente un intérêt tout particulier dans la protéine  $\alpha$ -syn. En effet, cette région a été décrite grâce à l'utilisation de la dynamique moléculaire tous atomes comme étant propice à former une hairpin de deux brins, 38-44 et 47-53, dont la boucle se constitue de 2 résidus en 45-46 (voir chapitre Introduction [section 1.3](#) page [13](#)) [\[68\]](#). Cependant les simulations présentant cette conclusion s'appuient sur un faible pourcentage de formation de ces structures. Cette faible statistique laisse donc une grande part d'inconnue sur cette région. Nous avons donc réalisé une analyse de cette région à partir de l'ensemble des monomères isolés extraits de nos simulations calculée avec le programme UNRES (voir [chapitre 2](#) [section 2.1](#) page [23](#)). Le but de notre analyse est de comparer nos résultats avec ceux déjà publiés mais également d'apporter d'autres éléments sur le comportement de cette région lorsqu'elle ne forme pas une hairpin. Dans un premier temps, en utilisant cette région comme point de départ, une analyse globale des couples d'angles  $(\theta, \gamma)$  (voir [chapitre 2](#) [figure 2.5](#)) formés par chaque résidu présent dans la région 38-53 a permis de mettre en évidence les différentes possibilités de conformations de ce segment ([figure 5.1](#)).

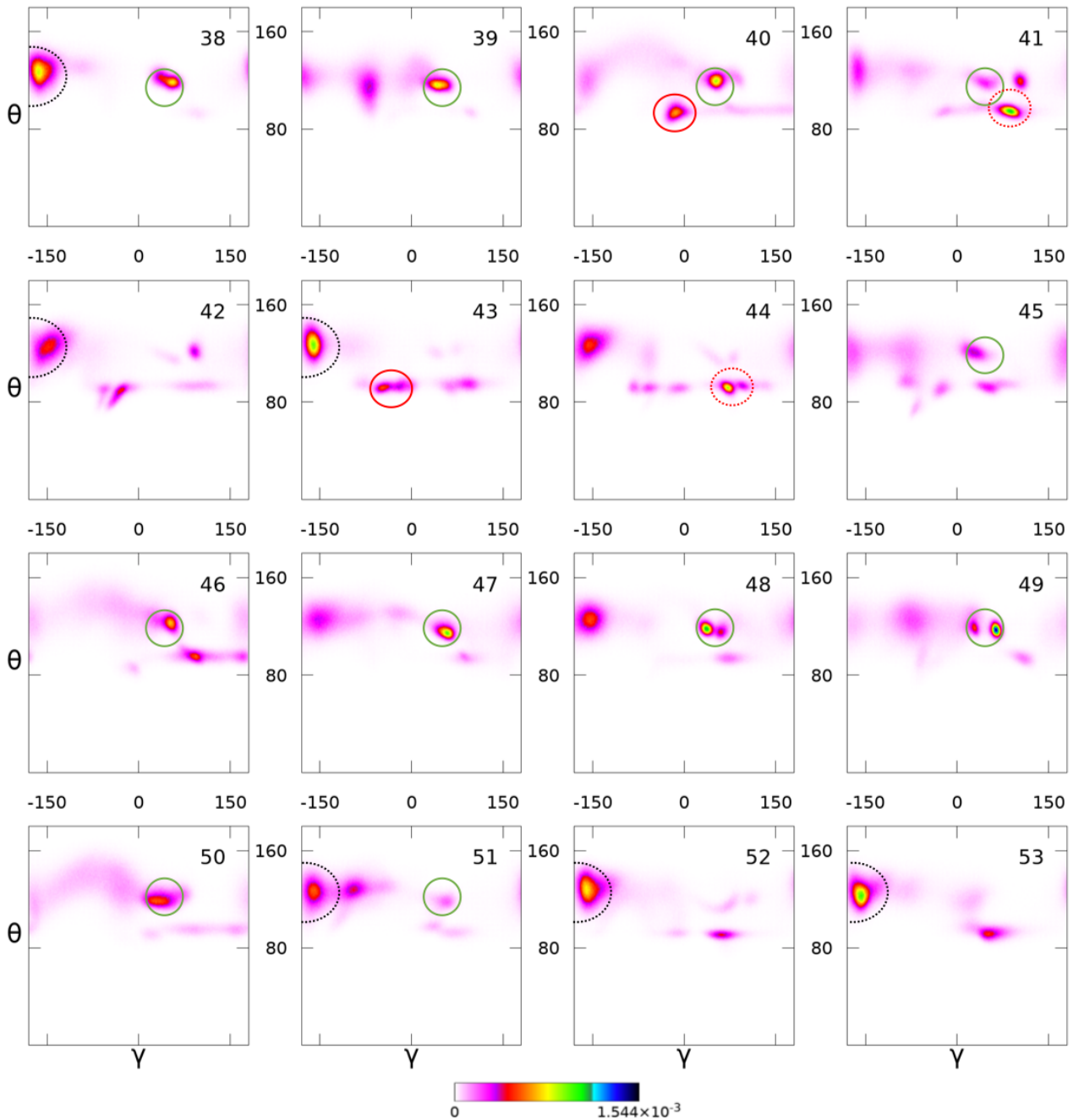


FIGURE 5.1 – Paysage d'énergie libre des couples d'angles diédraux  $\gamma$  et de liaisons  $\theta$  pour les AA 38 à 53 sur l'ensemble des monomères isolés pour le WT.

La majorité des AA présentent un état (état 1) entre  $-150^\circ$  et  $-180^\circ$  pour l'angle  $\gamma$  (cercles discontinus noirs [figure 5.1](#)), cette région du paysage d'énergie libre ( $\gamma, \theta$ ) est très favorable à la formation de feuillets  $\beta$ . Les AA 38, 42, 43, 51, 52, et 53 ont pour états majoritaires cette région.

Un second état (état 2) est identifié dans la région  $110^\circ < \theta < 120^\circ$  et  $30^\circ < \gamma < 60^\circ$  (cercles continus verts [figure 5.1](#)). Tous les AA sont concernés par cet état exceptés les AA 42, 43, 44, 52 et 53 où il est très minoritaire. L'intervalle d'angle  $\gamma$   $30^\circ - 60^\circ$  recoupe celui des hélices défini par CUTABI mais l'intervalle de l'angle  $\theta$  est cependant plus grand que celui des hélices. La figure 8 dans notre publication [\[95\]](#) montre bien l'absence d'hélices dans cette région des angles  $\theta$ . Il s'agit donc d'un autre type de structure atypique encore jamais discuté dans la littérature à notre

connaissance que nous avons appelé une tresse. Ce type de structure possède des caractéristiques proches de l'hélice par sa forme hélicoïdale ce qui impose aux résidus la constituant des angles  $\gamma$  semblables à ceux d'une hélice. Cependant, le fort encombrement stérique impliqué dans cette structure impose une élongation de la chaîne conduisant à un angle  $\theta$  légèrement supérieur.

Un troisième état (état 3)  $80^\circ < \theta < 100^\circ$  et  $-50^\circ < \gamma < 0^\circ$  est occupé par les AA 40,43 mais également plus faiblement par le 42 (représenté par un cercle rouge continu sur la [figure 5.1](#)). Cette région est un marqueur de la présence d'une boucle proche de ce résidu. Cet état est couplé avec un état possédant un angle  $\gamma$  au delà de  $50^\circ$  et un angle  $\theta$  compris entre  $80^\circ$  et  $110^\circ$ . Ce quatrième état (état 4) se retrouve avec une grande probabilité chez les résidus 41 et 44 (cercles rouges discontinus [figure 5.1](#)). En couplant la détection des état 3 et 4, on peut définir la position d'une boucle composée de deux résidus dans une hairpin (voir [chapitre 2](#) [section 2.2](#) page [28](#))

Un grand nombre d'états différents minoritaires sont également présents sur cette carte et confirment la grande flexibilité de la protéine et de ce segment en particulier. On donne à la [figure 5.2](#) une représentation 3D des 3 possibilités majoritaires de repliement du segment 38-53 chez l' $\alpha$ -syn : la hairpin, la tresse et la structure aléatoire.

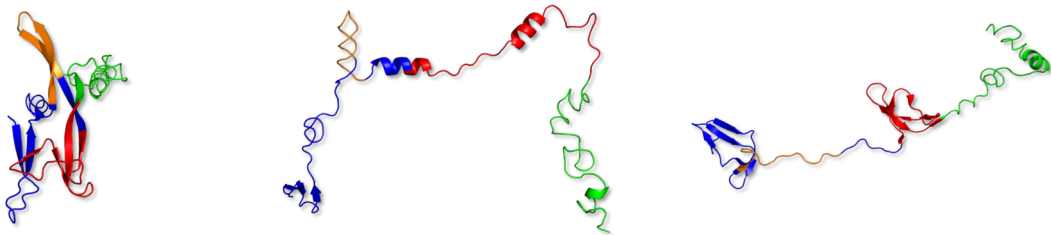


FIGURE 5.2 – Conformations 3D des 3 types de structures identifiées dans la région 38-53 dans notre ensemble de monomères. La région 38-53 est colorée en orange., le Nterm en bleu (1-60), le NAC en rouge (61-95) et le Cterm en vert (96-140). De gauche à droite : la hairpin, la tresse et la structure aléatoire.

## 5.2 Mise en évidence des trois types de structures principales

Nous avons étudié la présence des trois types de conformations détectées dans la partie précédente sur le segment 38-53. Dans le cas de la hairpin, nous détectons les boucles contenues dans cette portion de séquence. Les hairpins formées par ces boucles peuvent former des feuillets  $\beta$  bien au-delà des résidus du segment pris en compte. Pour des raisons explicitées plus loin, nous avons étendu l'analyse des boucles formant des hairpins sur le segment 35-53.

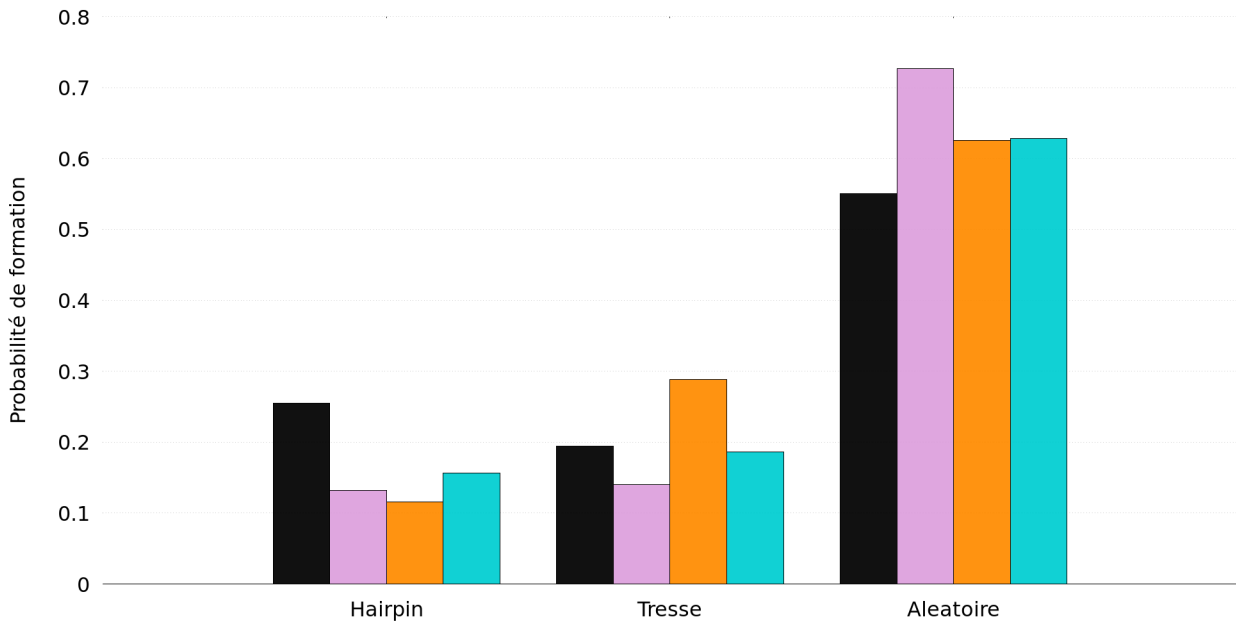


FIGURE 5.3 – Histogramme de probabilité de formation de boucle sur le segment 35-53 formant une hairpin, la tresse et le repliement aléatoire pour le WT (noir), A30P (rose), A53T (orange) et E46K (cyan) dans l'ensemble des monomères.

Les proportions des 3 types de structures retrouvées sur le segment 35-53 sont données [figure 5.3](#). Le résultat marquant de cette analyse est que le WT favorise la formation d'hairpins par rapport aux mutants, ce qui réduit sa probabilité de former des repliements aléatoires. La tresse est la conformation la plus représentée pour le mutant A53T. Le mutant A30P favorise beaucoup plus la formation de repliements aléatoires que les autres protéines.

## La classe hairpin : les différentes conformations possibles

En considérant les boucles des hairpins comme étant des couples de résidus, on définit des histogrammes de probabilités de formation de couple de boucles au sein d'une hairpin à deux résidus noté  $H_{x/x+1}$ . La comparaison donnée à la [figure 5.4](#) démontre une bien plus grande probabilité de formation de la hairpin  $H_{41-42}$  chez le WT ( $p=0.20$ ) comparé aux mutants avec un rapport de 2 entre le WT et les deux mutants A30P ( $p=0.09$ ) et A53T ( $p=0.1$ ) et un rapport de 4 avec le mutant E46K ( $p=0.05$ ).

La  $H_{45-46}$  retrouvée dans les simulations de la ref [\[68\]](#) n'apparaît que très rarement dans nos conformations ( $p<0.01$ ). La hairpin ayant une probabilité non négligeable dans notre ensemble est la  $H_{46-47}$  avec une probabilité de présence de 0.05 pour le WT. E46K est nettement plus favorable à sa formation avec une probabilité de présence 0.09. La boucle est constituée du motif E46G47 pour le WT, A30P et A53T mais de K46G47 pour E46K. La lysine est plus favorable à la formation de boucles par rapport à l'acide glutamique ce qui pourrait expliquer ces écarts de probabilité de formation.

Un intérêt particulier peut être également porté à la  $H_{35-36}$  qui possède une forte probabilité de formation chez le WT ( $p_{WT}=0.18$ ) contrairement aux mutants ( $p_{A30P}=0.01$ ,  $p_{A53T}=0.04$ ,

$p_{E46K}=0.08$ ). Lorsque cette hairpin se forme pour le WT, elle est couplée très souvent avec la  $H_{41-42}$  ( $p=0.08$ ). Ce phénomène est beaucoup plus rare pour les mutants.

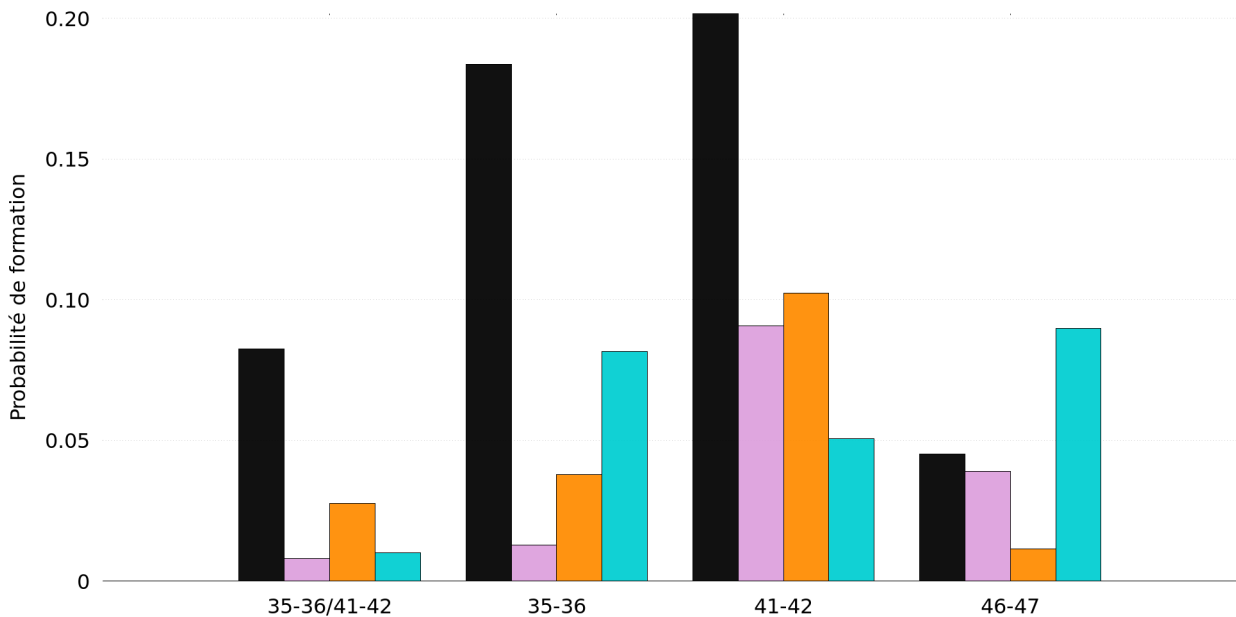


FIGURE 5.4 – Histogramme de probabilité de formation des hairpins sur le segment 35-53 suivant la position de la boucle constituée de deux résidus dans l'ensemble des monomères. Le couple de résidus appartenant à la boucle de la hairpin est donné avec  $X, X+1$  sur l'axe des abscisses. Le cas 35-36/41-42 donne la probabilité de formation des deux boucles simultanément.

Dans les analyses suivantes nous chercherons à comprendre le comportement générale de chaque hairpin et son impact sur le reste de la protéine. Pour cela nous avons regroupé toutes les conformations contenant chaque type de hairpin dans des classes. Nous avons ensuite réalisé les CM de l'ensemble des conformations contenu dans chaque classe. La méthodologie de construction des CM est donnée dans la partie Méthode. (voir [section 2.3](#) page [32](#))

### - Etude de la Hairpin 46-47 ( $H_{46-47}$ )

Les CM des conformations possédant le hairpin  $H_{46-47}$  sont données [figure 5.5](#). En s'intéressant dans un premier temps au WT, on remarque une bande perpendiculaire à la diagonale de la CM de probabilité proche de 1, centrée sur les résidus 46 et 47 correspondant à la hairpin recherchée dans cette classe. Son premier brin se forme du résidu 38 au 45 et le second du 48 au 55 soit 18 résidus en totalité en incluant les deux résidus composant la boucle. Dans ces conditions on est très proche de la configuration retrouvée dans la référence [\[68\]](#). Avec ce type d'hairpin de taille fixe le WT possède 3 configurations majeures, représentées à la [figure 5.5](#), qui sont :

Conformation A : Une multitude de feuillets  $\beta$  anti-parallèles successifs ( $A1, A2, H_{46-47}$  et  $A3$ ) se forment du Nterm jusqu'au début du NAC créant un plan  $\beta$  dans lequel la  $H_{46-47}$  fait interagir son premier brin avec le Nterm et son second avec le NAC.



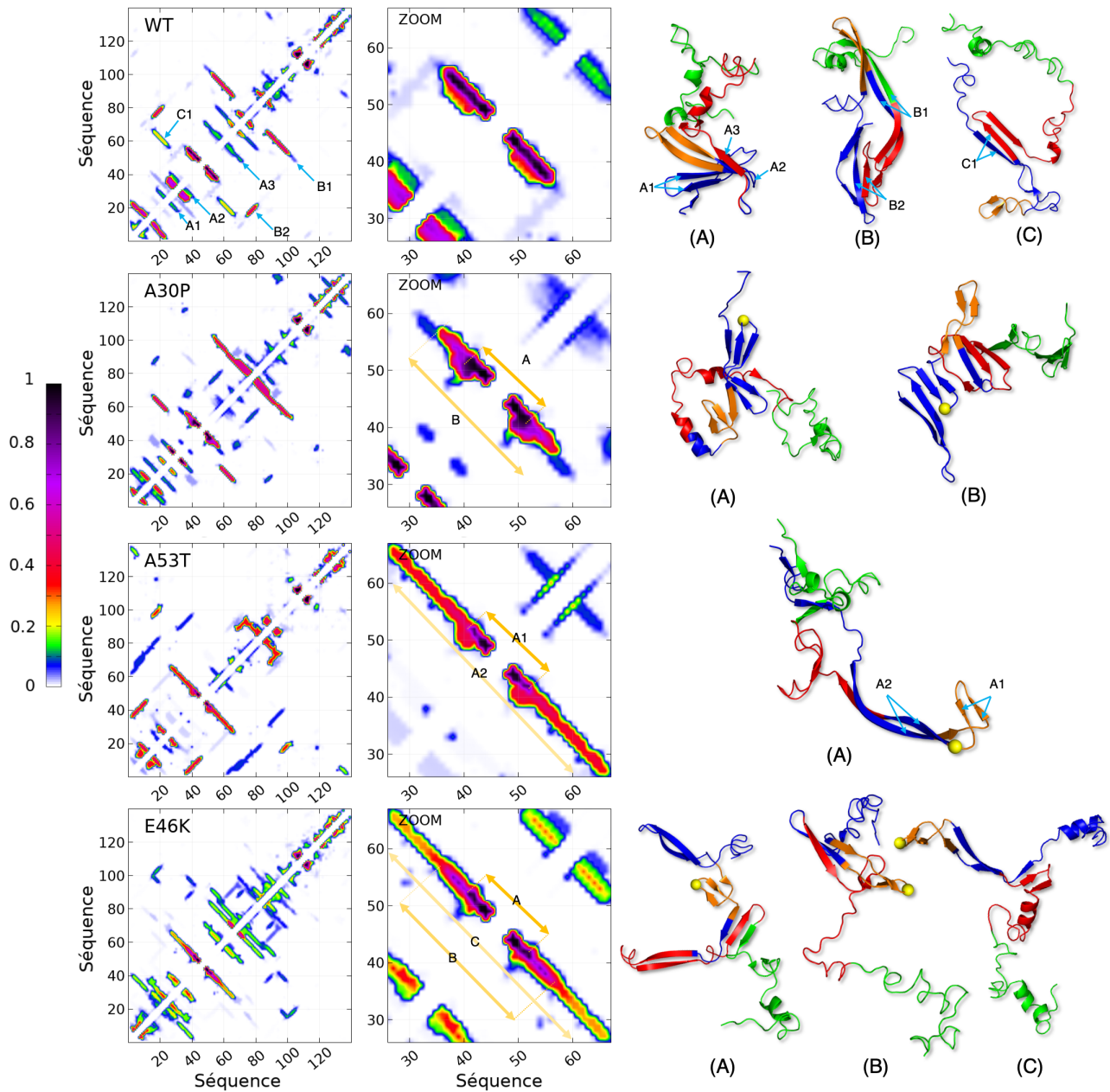


FIGURE 5.5 – Chaque ligne de figures correspond à chaque protéine (WT, A30P, A53T et E46K). Cartes portant le nom des protéines : CM de la protéine complète pour la classe des conformations possédant la hairpin  $H_{46-47}$ . A côté de chaque carte, on retrouve un zoom de ces cartes pour le segment 26-67. Des représentations 3D sont données pour illustrer les cartes avec pour code couleur en bleu le Nterm 1-61, en rouge le NAC 61-95, en vert le Cterm 96-140 et en orange le segment 38-53. La position des mutations est identifié par une sphère jaune.

Conformation B : La protéine est cette fois plus étirée avec un très grand feuillet  $\beta$  (B1). Ce feuillet  $\beta$  fait interagir le debut du Cterm avec le premier brin de la  $H_{46-47}$ . Une petite portion dans le NAC du feuillet  $\beta$  (B1) modifie légèrement sa structure pour former un feuillet  $\beta$  parallèle avec le Nterm (B2) permettant de rendre la structure globale très rectiligne.

Conformation C : Une conformation hybride entre la tresse et la  $H_{46-47}$  mettant en avant le lien étroit entre ces deux configurations. A cette configuration s'ajoute un repliement du NAC

avec le Nterm sous forme d'un feuillet  $\beta$  anti-parallèle contenu dans un plan  $\beta$  constitué de 3 brins (C1).

La longueur de la hairpin  $H_{46-47}$  semble ne varier que très peu chez le WT contrairement aux mutants où l'on observe plusieurs longueurs possibles avec différentes probabilités. Une hairpin très courte 41-52 (12 résidus) est observable chez les 3 mutants (notée A). Les figures 3D associées pour A30P et E46K montrent une formation d'une succession de 3 brins dont deux sont parallèles et un anti-parallèle (partie orange conformation (A)). Cette petite hairpin chez A53T est présente à l'extrémité d'un très grand feuillet A2. Cette configuration correspond également à la conformation (C) de E46K. E46K et A30P possèdent également une  $H_{46-47}$  plus longue du résidus 38-55 semblable à celle du WT (conformation (B)). Les contacts sont présents mais la formation du feuillet  $\beta$  complet n'est pas validé par CUTABI dans les conformations (B) représentées pour ces mutants.

### - Etude de la Hairpin 41-42 ( $H_{41-42}$ )

Ce type d'hairpin est la plus probable chez les 4 types de protéines étudiées et plus particulièrement chez le WT. Les CM de chaque protéine pour cette classe sont données [figure 5.6](#). Le WT forme une très grande variété de structures contenant cette hairpin. Comme pour  $H_{46-47}$ , différentes tailles d'hairpin sont possibles. On en dénombre trois pour la  $H_{41-42}$  chez le WT, deux pour A30P et A53T et une pour E46K. La plus petite est constituée de 10 résidus allant du 37 au 46 (conformation (A) pour les 4 protéines). Cette petite hairpin se forme associée à la  $H_{35-36}$  pour les 4 protéines au sein d'un plan  $\beta$  plus ou moins grand. En particulier pour A30P, la conformation (A) montre une seconde boucle formée par la mutation permettant l'ajout d'un 4ème brin dans le plan  $\beta$ . Ce type de structure connaît une autre forme chez E46K (conformation B) avec un repliement parallèle du brin B2 sur le premier brin de la  $H_{41-42}$ . La configuration B pour A53T, A30P et le WT est une hairpin imparfaite plus longue de 16 résidus incluant les résidus 34 à 49 se liant une nouvelle fois à des brins anti-parallèles du Nterm dans un plan  $\beta$  de taille plus conséquente. La mutation A30P impose la formation d'une boucle et donne donc une typicité différente dans la formation du plan  $\beta$ . Enfin le WT prend une dernière configuration avec une hairpin de 30 résidus du 26 au 57 où cette fois on ne retrouve pas la formation de plan  $\beta$  mais une hairpin libre (configuration (C)).

On remarque une multitude de repliements possibles de la fin du Cterm avec le reste de la protéine (pour le WT visible notamment sur la conformation (C).)

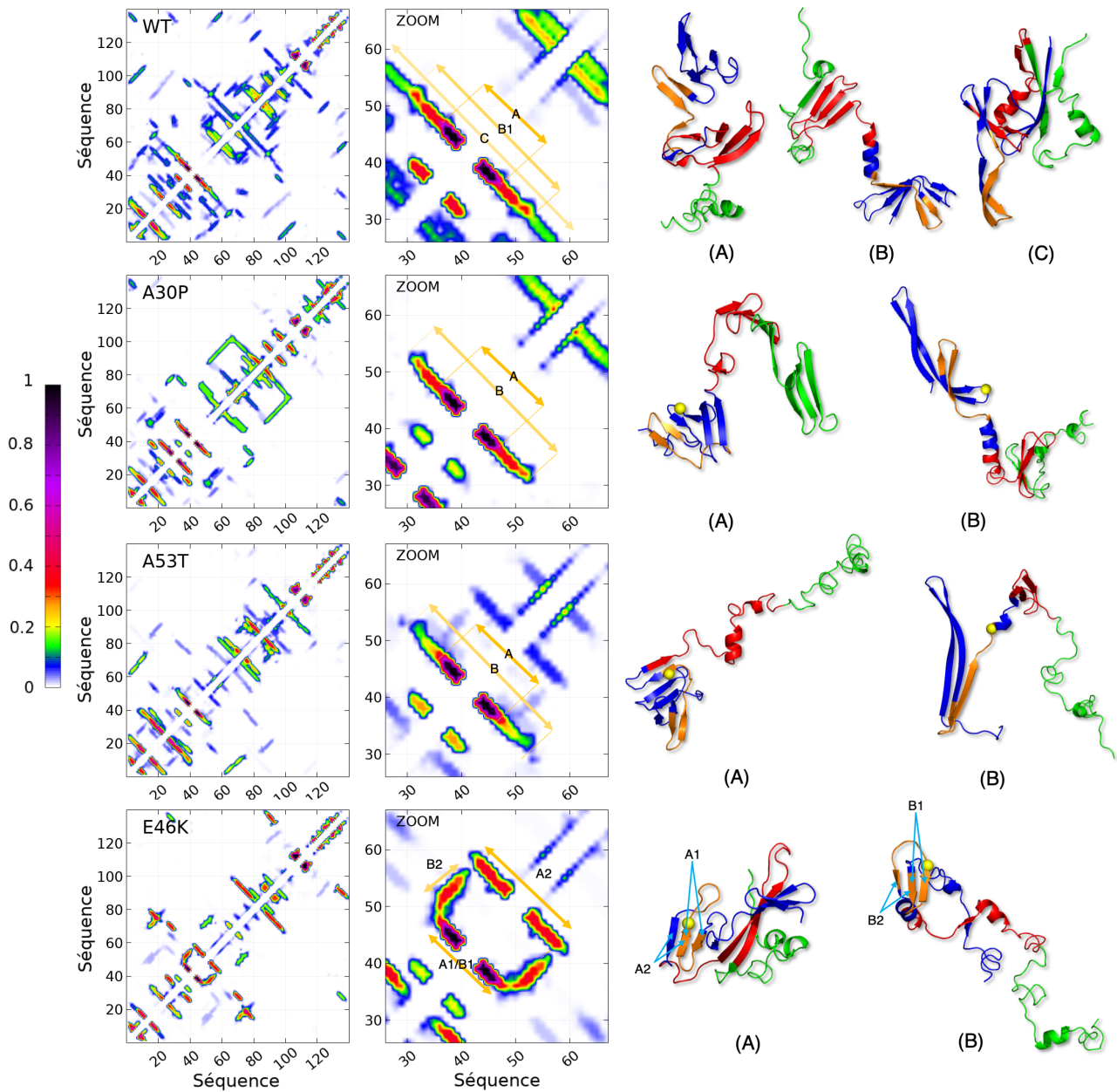


FIGURE 5.6 – Chaque ligne de figures correspond à chaque protéine (WT, A30P, A53T et E46K). Cartes portant le nom des protéines : cartes de contact de la protéine complète pour la classe des conformations possédant la hairpin  $H_{41-42}$ . A côté de chaque carte, on retrouve un zoom de ces cartes pour le segment 26-67. Des représentations 3D sont données pour illustrer les cartes avec pour code couleur en bleu le Nterm 1-61, en rouge le NAC 61-95, en vert le Cterm 96-140 et en orange le segment 38-53. La position des mutations est identifié par une sphère jaune.

## La classe tresse

La tresse est une structure atypique jamais discutée à notre connaissance dans la littérature. Ses caractéristiques à la fois de feuillet  $\beta$  et d'hélice lui confèrent des propriétés particulières, notamment au niveau des angles diédraux des résidus qui la constituent. La méthode de détection de cette structure est explicitée dans la partie Méthode ( [section 2.3](#) page [29](#) ).

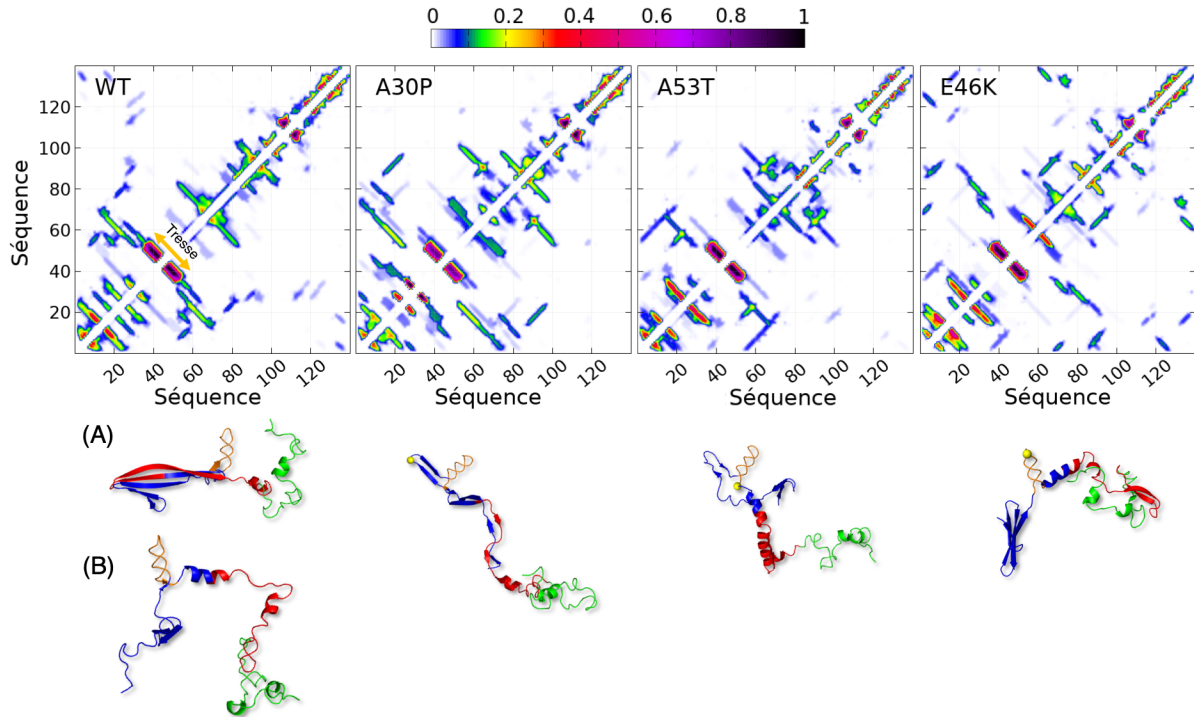


FIGURE 5.7 – Cartes de contacts des protéines  $\alpha$ -syn WT, A30P, A53T et E46K de la classe tresse. Des représentations 3D sont données pour illustrer les cartes avec pour code couleur en bleu le Nterm 1-61, en rouge le NAC 61-95, en vert le Cterm 96-140 et en orange le segment 38-53. La position des mutations est identifiée par une sphère jaune.

La probabilité de présence de la tresse pour les 4 protéines d'étude est de près de 0.3 pour A53T, environ 0.2 pour E46K et le WT et moins de 0.15 pour A30P. Sa présence n'est donc pas négligeable et son fort encombrement peut avoir un impact crucial sur d'éventuels repliements du reste de la chaîne sur cette région mais également sur l'agrégation avec un autre monomère d' $\alpha$ -syn.

La même étude que pour les hairpins a été effectuée sur le groupe des conformations possédant une tresse. Sur les CM données [figure 5.7](#), la tresse est facilement identifiable par un trace épaisse perpendiculaire à la diagonale car les résidus présents dans la structure vont former plusieurs contacts sous le rayon de coupure utilisé.

Les résidus impliqués vont du 37 au 51, avec une boucle constituée de 3 résidus 43,44 et 45. Les CM données [figure 5.7](#) montrent que la structure n'interagit que peu avec le reste de la protéine contrairement aux hairpins présentées précédemment. Les représentations 3D données sous les CM ([figure 5.7](#)) exposent la grande variété des conformations prises par les différentes protéines possédant la tresse. Le NAC et le Nterm du WT ont la capacité en présence de la tresse

de former de très grand feuillets  $\beta$  constitués de plusieurs brins dans un plan  $\beta$  (conformation (A)) ou à l'inverse d'être totalement distincts avec une formation d'hélice à la jointure des deux régions (conformation (B)). Pour A30P, le Nterm forme un très grand feuillet imparfait dont la boucle est positionnée au niveau de la mutation. A53T possède un type de conformation très riche en hélices. E46K peut avoir une conformation hybride avec un Nterm très agrégé en feuillets  $\beta$  et une hélice à la jointure Nterm-NAC séparée par la tresse. La tresse semble donc être une structure n'impactant que les résidus la constituant.

## La classe aléatoire

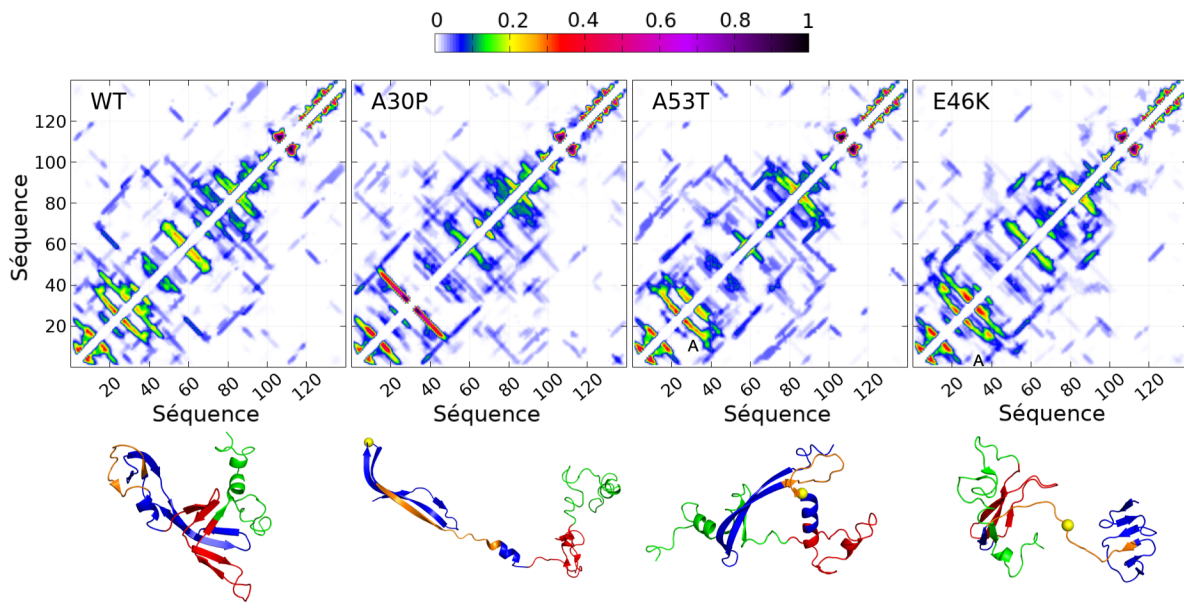


FIGURE 5.8 – Cartes de contacts des protéines  $\alpha$ -syn WT, A30P, A53T et E46K de la classe aléatoire. Des représentations 3D sont données pour illustrer les cartes avec pour code couleur en bleu le Nterm 1-61, en rouge le NAC 61-95, en vert le Cterm 96-140 et en orange le segment 38-53. La position des mutations est identifiée par une sphère jaune.

Les classes hairpin et tresse étudiées précédemment ne représentent pas statistiquement la majorité des conformations prises par le segment 38-53. En effet, la plupart des conformations des quatre protéines étudiées ne possèdent ni hairpin ni tresse dans cette région. Le segment est simplement déplié de manière aléatoire ou forme un feuillet à plus longue distance n'étant donc pas considéré comme une hairpin. Les CM présentées à la [figure 5.8](#) montrent l'immense variété de contacts possibles au sein des 4 différentes protéines, conduisant à de multiples possibilités de conformations différentes. Dans les 4 protéines, le Nterm avant le segment 38-53 semble pouvoir former plus de contacts. A30P diffère par la très longue trace perpendiculaire à la diagonale de la CM centrée sur le résidu 30 de sa mutation (en rouge). La représentation 3D d'une structure possédant cette propriété est donnée en dessous de la carte. On identifie facilement le long feuillet incluant une partie du segment 38-53 (en orange) avec la mutation dans sa boucle. Pour les autres mutants une structure 3D choisie aléatoirement est présentée sous chaque CM ([figure 5.8](#)). On observe sur les conformations choisies pour le segment 38-53

une forme de boucle élargie chez le WT et A53T et une forme très allongée chez E46K.

### 5.3 Etude approfondie de la tresse

#### Résidus favorables à la formation d'une tresse

L'étude du segment 38-53 nous a amené à repérer la présence de la tresse dans cette région de la protéine. On peut se demander si ce phénomène est propre à ce segment ou si il se produit également dans une autre région de l' $\alpha$ -syn . La gamme d'angles utilisée pour détecter une tresse peut être recherchée pour tous les résidus de la protéine de sorte à déterminer ceux étant favorables à sa formation. Cette méthode ne confirme pas la formation d'une tresse puisque celle-ci nécessite une coordination de plusieurs résidus, mais la capacité de certain résidu à prendre les angles propices à sa formation.

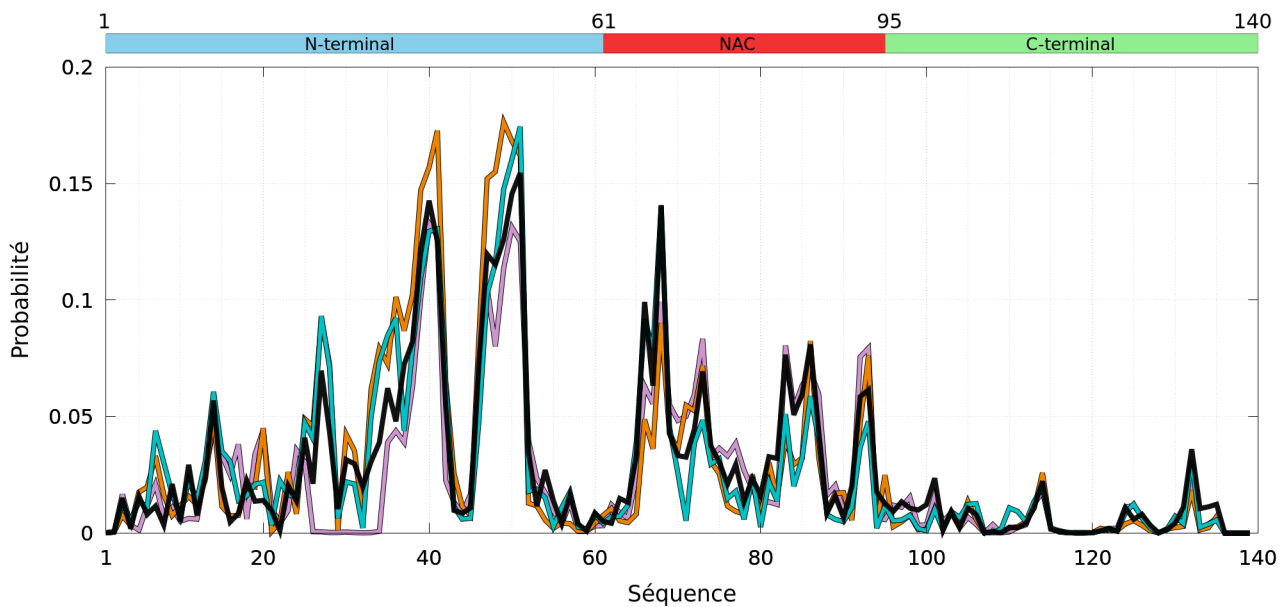


FIGURE 5.9 – Probabilité qu'un résidu possède la gamme d'angles favorables à la formation d'une tresse pour le WT (noir), A30P(rose), A53T (orange) et E46K(cyan) dans l'ensemble des monomères.

La [figure 5.9](#) montre que les résidus susceptibles de former une tresse sont majoritairement dans la région du Nterm pour le WT et les mutants. Pour le WT, on retrouve deux pics majeurs incluant les résidus 39YVG41 ( $p=0.12, 0.14, 0.12$ , respectivement) et 48GVVHG51 avec une probabilité supérieure à 0.1 pour tous ces résidus et un maximum pour G51 avec  $p=0.15$ . D'autres pics sont présents dans le NAC avec le segment 66VGG68 ( $p=0.1, 0.06, 0.14$ , respectivement) et des résidus ponctuels G73, E83, G86 ( $p=0.07, 0.08, 0.08$ , respectivement). On remarque que les résidus favorables à cette gamme d'angles sont majoritairement des glycines et des valines. La valine possède une chaîne latérale courte et la glycine n'en possède pas. Ce sont également des résidus hydrophobes au même titre que les alanines et que la tyrosine 39 dont les probabilités de formation sont également élevées. L'hydrophobicité ne semble pas être l'unique

condition puisque le NAC est bien plus hydrophobe que le Nterm cependant on y trouve moins de résidus favorables à la formation de la tresse. C'est la combinaison de résidus hydrophobes et de glycines qui apparaît comme un facteur important. En effet par exemple les résidus G51, G73 et G86 favorables à la formation d'une tresse possèdent tous deux un voisin hydrophobe (V52, V74 et A85 respectivement).

La comparaison entre les différentes protéines montrent un comportement quasi similaire entre le WT et A30P (excepté une décroissance forte pour A30P proche du résidu 30). Par rapport au WT, la mutation A53T favorise la formation de la tresse sur le segment possédant deux pics 39-51 et la diminue légèrement dans le NAC. Les probabilités de formation de la tresse pour E46K sont semblables au WT dans le Nterm mais plus faibles que celui-ci dans le NAC.

### Simulation du segment 38-53 seul avec UNRES

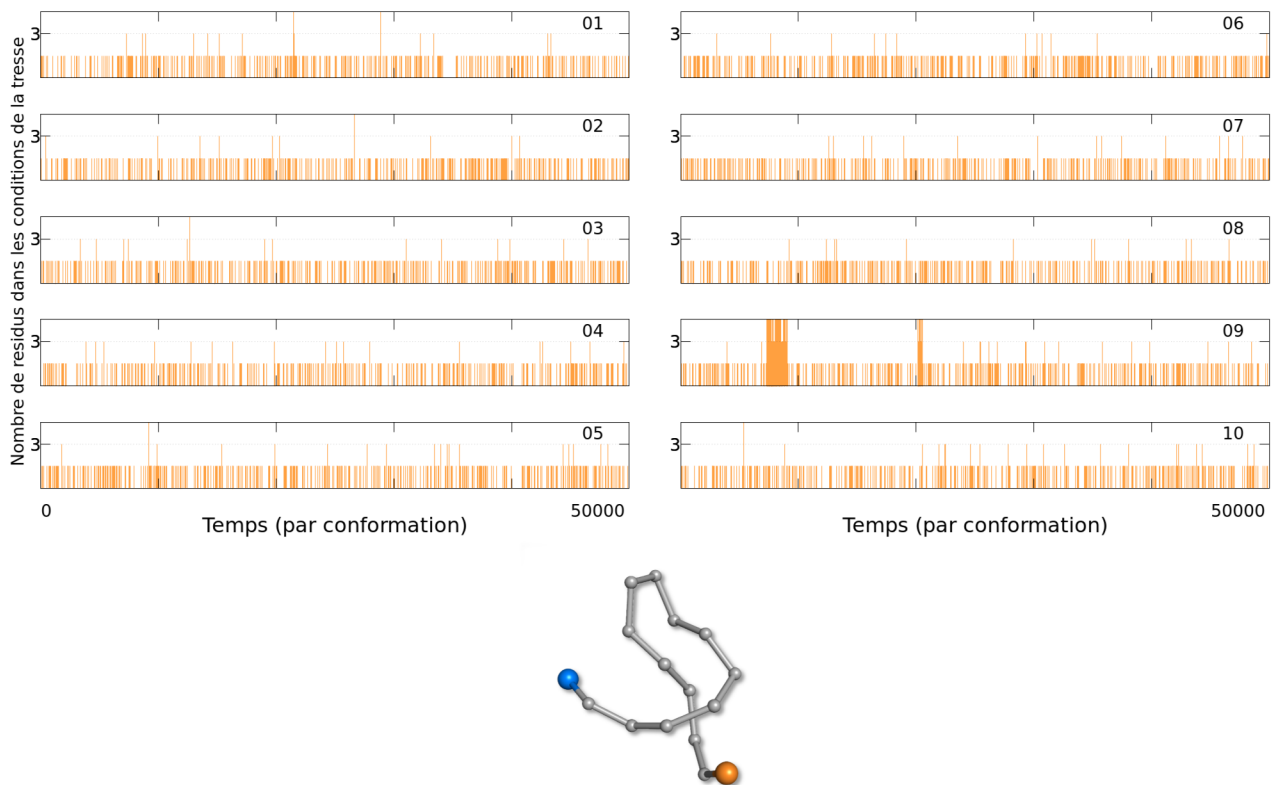


FIGURE 5.10 – Nombre de résidus possédant des angles favorables à la formation de tresse au sein des simulations du segment 38-53 seul. Une courbe est donnée pour chaque trajectoire en fonction du temps. Une représentation 3D du cas de figure possédant le plus de résidus favorables dans la trajectoire 09 est donnée sous la courbe. Le résidu 38 est matérialisé par une sphère bleu et le résidu 53 par une sphère orange.

On peut se demander si la tresse se forme en raison des résidus qui la compose ou si un effet d'environnement de la séquence est responsable de la formation de cette structure dans l' $\alpha$ -syn. Pour répondre à cette problématique, nous avons simulé le segment 38-53 seul. La séquence complète du peptide simulé est donc 38LYVGSKTKEGVVHGVA53. Dix trajectoires

de dynamique moléculaire dans l'ensemble NVT à 310K, partant d'une conformation rectiligne du peptide ont été simulées. Chaque trajectoire est constituée de 50 millions de pas équivalent à un temps effectif de 200 ms. Les coordonnées cartésiennes des atomes ont été sauvegardées tous les 1000 pas donnant ainsi 50000 conformations par trajectoire. Pour chaque trajectoire, nous avons calculé les angles diédraux et de liaison de chaque résidu qui ont ensuite été comparés à la gamme d'angles utilisée dans la détection des tresses. La [figure 5.10](#) donne en fonction du temps le nombre de résidus dans les conditions favorables à la formation d'une tresse au sein du peptide 38-53. On considère qu'à partir de 3 résidus, la structure du peptide commence à former une boucle. Une représentation 3D de ce cas de figure est donnée sous les courbes (extraite de la trajectoire 09). On remarquera que les conformations comprenant 3 résidus se forment de manière très sporadique dans les trajectoires. Aucune trajectoire n'arrive à maintenir un nombre de résidus au delà de 3 de manière stable sur la durée. Aucune tresse comparable à celle obtenue dans nos simulations de l' $\alpha$ -syn complète n'est observée dans ces trajectoires du segment 38-53 seul. La stabilité globale de la tresse semble donc être causée par un effet d'environnement du reste de la séquence de l' $\alpha$ -syn .

### Les différentes portions du segment 38-53 : impact de chaque portion sur la formation potentielle d'une tresse

Le comportement de ce segment seul n'étant pas propice à la formation de tresse, on peut se demander si dans un autre environnement, c'est à dire dans une autre protéine que l' $\alpha$ -syn ce segment réagit de la même façon. Le segment 38LYVGSKTKEGVVHGVA53 n'existe que chez la protéine  $\alpha$ -syn dans l'ensemble des protéines de la PDB à ce jour. En effet, cette protéine est la seule connue à posséder le motif KTKEGV. Cependant, les deux autres parties de la séquence LYVGS et VHGVA se retrouvent dans 393 et 35 autres protéines de la PDB respectivement. Parmi celles-ci, nous avons sélectionné à l'aide du logiciel BLAST un petit peptide possédant le motif LYVGS de 13 acides aminés 1CVLYVGSGINWVF13 dont le nom est "immunoglobulin light chain junction region" [\[96\]](#). Ce peptide a été simulé dans les mêmes conditions que le segment 38-53 seul. Nous avons également simulé ce peptide en changeant le motif LYVGS par KTKEGV et VHGVA. Nous avons donc simulé 3 peptides possédant les séquences suivantes :

- CVLYVGSINWVF
- CVKTKEGVINWVF
- CVVHGVAINWVF

L'évolution en fonction du temps du nombre de résidus possédant des angles favorables à la formation d'une tresse est donnée [figure 5.11](#). La simulation du peptide original possédant le motif LYVGS (courbes de gauche en rouge) montre une tendance à atteindre fréquemment 3 résidus mais la stabilité semble discutable. Le peptide contenant le motif KTKEGV (courbes centrale en orange) a un comportement semblable à celui du segment 38-53 ([figure 5.10](#)) avec très peu de conformations contenant 3 résidus. Avec le motif VHGVA (courbes de droite en vert), le



peptide a cette fois un comportement très favorable à la formation de tresse en atteignant très rapidement dans toutes les simulations un nombre de 4 résidus. La représentation 3D donnée permet de voir en vert le motif VHGV qui donne la forme de tresse au peptide. Ce motif est constitué de 3 résidus hydrophobes (2 valines et 1 alanine) ainsi que d'une glycine. Ces résidus ont été identifiés précédemment (partie [section 5.3](#)) comme étant les plus favorables à la formation de tresse lorsqu'ils sont combinés.

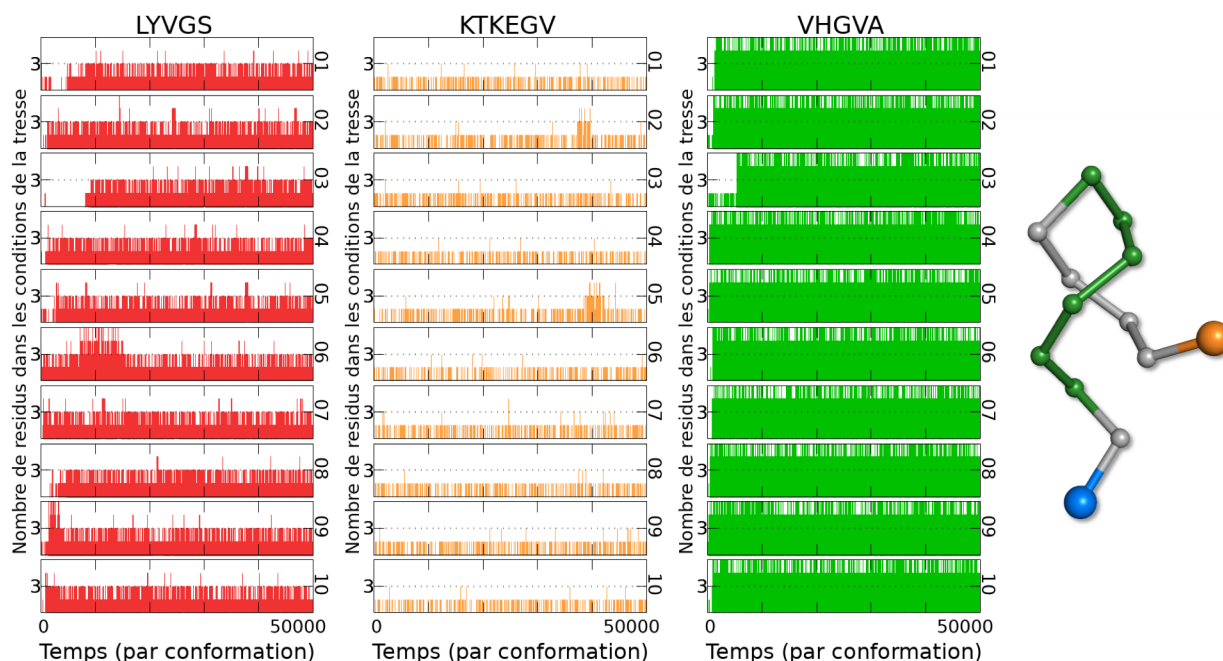


FIGURE 5.11 – Nombre de résidus possédant des angles favorables à la formation de tresse au sein des simulations du peptide 1CVLYVGS<sub>GINWVF13</sub> (panneau de gauche en rouge), 1CVKTKEGV<sub>GINWVF13</sub> (panneau au centre en orange) et 1CVHGV<sub>AGINWVF13</sub> (panneau de droite en vert). Une courbe est donnée pour chaque trajectoire en fonction du temps. Une représentation 3D du cas de figure possédant le plus de résidus favorables pour le peptide 1CVHGV<sub>AGINWVF13</sub> est donnée à droite. Le Nterm est matérialisé par une sphère bleu et le Cterm par une sphère orange. La partie en vert correspond aux résidus du motif VHGV.

D'après la [figure 5.11](#), le LYVGS et surtout VHGV semblent favorables à la formation d'une tresse dans un petit peptide dans les simulations réalisées avec UNRES. Afin d'examiner l'effet de l'environnement sur la formation d'une tresse, nous avons aussi réalisé une simulation de deux protéines contenant ces motifs. Elles sont toutes les deux des structures cristallisées extraites de la PDB. La première protéine contient le motif VHGV (code PDB : 2jfq), la structure complète se présente sous forme de dimère, dans notre cas uniquement un monomère a été simulé. Le RMSD du motif VHGV est donné [figure 5.12](#) panneau (A). La protéine n'étant pas sous sa forme dimérique, la stabilité a été affectée et UNRES a donc déplié la protéine. Le segment VHGV à l'origine dans un feuillet  $\beta$  (partie orange sur les représentations 3D) a un RMSD très fluctuant et se retrouve donc au cours de la simulation dans diverses conformations aléatoires mais ne forme à aucun moment une tresse. Cette protéine possède dans sa séquence 4 motifs GV (dont un dans VHGV) et 1 motif VG. Ces motifs favorables à la formation de tresse chez l' $\alpha$ -syn n'ont pas provoqué la formation de tresse chez 2jfq.

La seconde protéine contenant le motif LYVGS (code PDB : 3c6u) cristallisée sous forme monomérique ne connaît pas de fort changement structurel au cours de la simulation. Le RMSD représenté [figure 5.12](#) panneau (B) montre très peu de variation pour le segment contenant le motif LYVGS. Ce segment à l'origine dans un feuillet  $\beta$  conserve sa forme. Cette protéine contient 1 fois le motif GV et 4 fois le motif GV (dont une fois dans LYVGS), on remarquera également une portion de séquence VHGVY très semblable à VHGVA. Malgré ces nombreux éléments propices à la formation de tresse la protéine n'en a en aucun cas formé lors des simulations réalisées avec UNRES.

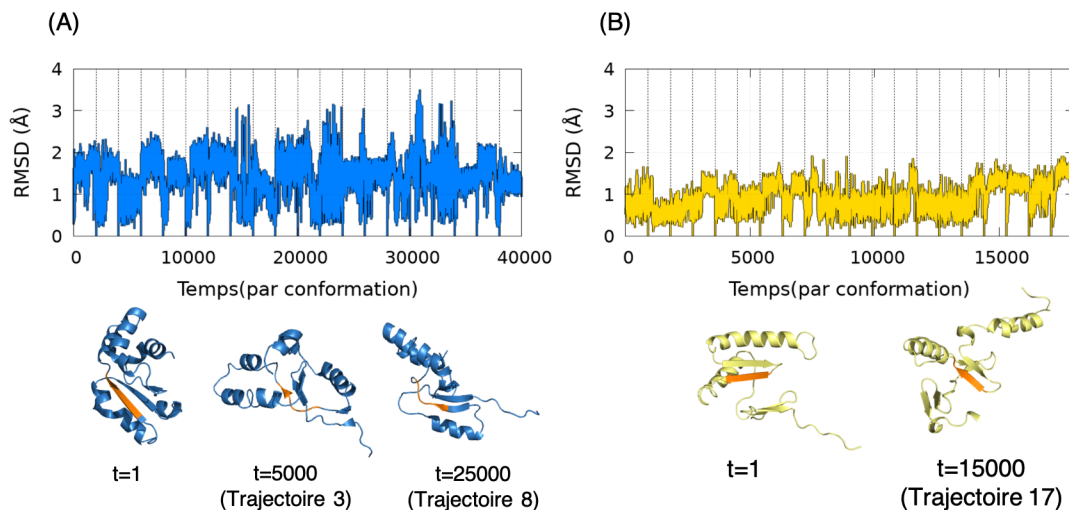


FIGURE 5.12 – RMSD du segment possédant le motif VHGVA ou LYVGS dans des simulations de dynamiques moléculaires (NVT) réalisées avec UNRES en partant de deux protéines cristallisées 2jfq (VHGVA) (bleue, panneau (A)) et 3c6u (LYVGS) (jaune, panneau (B)). Les barres pointillées donnent les débuts et fins des différentes trajectoires réalisées. Les structures 3D représentées donnent en orange la portion de la séquence étudiée dans le RMSD.

## Stabilité de la tresse avec un champ de force tous atomes

Deux analyses ont été réalisées afin de déterminer si la tresse est une structure possible dans une représentation tous atomes. Deux conformations possédant une tresse reconstruites en tous atomes grâce au programme MMTSB [\[90\]](#) ont été sélectionnées aléatoirement comme étant le point de départ d'une trajectoire de dynamique moléculaire tous atomes. Deux champs de force différents ont été utilisés dans cette étude. Le premier est CHARMM36m (C36m) construit pour l'analyse des protéines intrinsèquement désordonnées, ce champ de force est associé à un modèle d'eau TIP3P [\[97\]](#). Le second champ de force est AMBER [\[98, 99\]](#). L'étude précédente a montré que l'entièreté de la protéine était nécessaire pour la formation d'une tresse, les simulations tous atomes ont donc été réalisées en utilisant la protéine complète.

Les RMSD du segment 38-53 relatifs à la structure initiale à  $t=0$  calculés avec ces deux champs de force sont représentés à la [figure 5.13](#). La première simulation ([figure 5.13](#) panneau (A)) montre que la forme globale de la tresse s'est stabilisée après une nanoseconde de simulation, sans conserver parfaitement sa géométrie avec les deux champs de force. Le RMSD a fait un saut d'environ 3 Å au début puis s'est stabilisé après 1 nanoseconde. Les trois boucles

initiales se sont dépliées en deux boucles puis une seule dans la dernière structure générée par la simulation à  $t=140$ ns.

Dans la seconde simulation (figure 5.13 panneau (B)), la tresse s'est très vite dépliée avec le champ de force C36m pour former une boucle simple qu'elle maintiendra tout au long de la simulation comme le montre la stabilité du RMSD. Avec AMBER, la structure s'est rapidement dépliée en deux boucles. Le RMSD montre plusieurs grandes fluctuations à partir de 30 ns. Le segment va complètement se déplier et prendre plusieurs configurations aléatoires désordonnées illustrées par la dernière structure à  $t=140$ ns.

Les contraintes stériques de la structure initiale montre que ce type de structure est possible mais instable en tous atomes avec les champs de force C36m et AMBER.

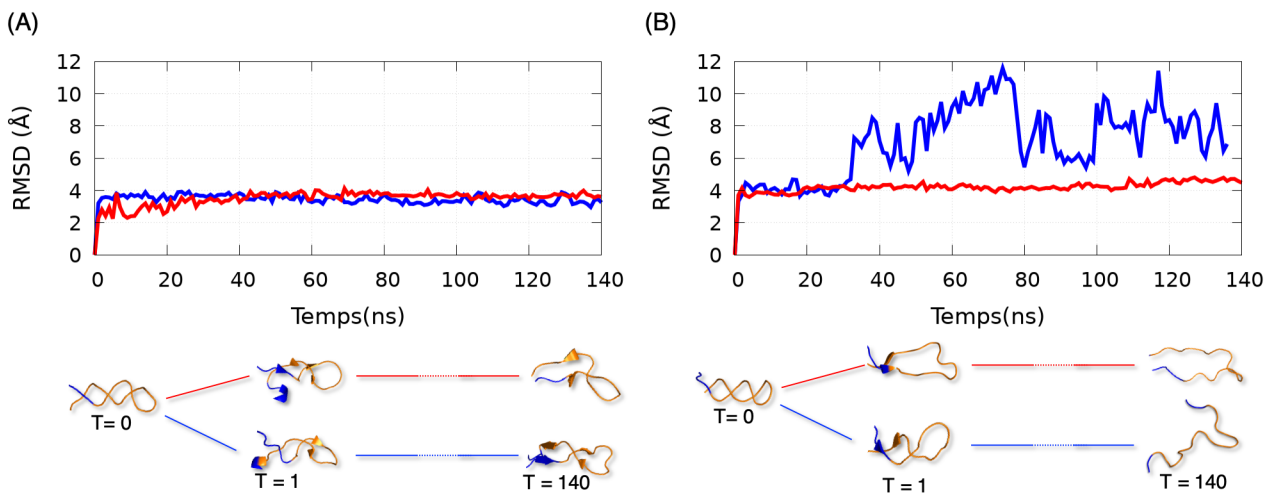


FIGURE 5.13 – RMSD du segment 38-53 pour les simulations tous atomes réalisées à partir d'une structure initiale d' $\alpha$ -syn complète possédant une tresse. Avec en bleu les simulations réalisées avec AMBER et en rouge celles réalisées avec C36m. En dessous des courbes, des représentations 3D du segment 38-53 sont données à différents temps. Le chemin bleu correspond aux structures pour AMBER et le chemin rouge à celles pour C36m. Les panneaux (A) et (B) montrent les deux trajectoires simulées avec des structures initiales différentes.

## Formation de tresse chez $A\beta$

La protéine  $A\beta$  est réputée pour son implication dans la maladie d'Alzheimer. [7, 8, 9]. Son comportement est très semblable à l' $\alpha$ -syn car elle est également capable de former des fibrilles riches en feuillets  $\beta$  générant une neurotoxicité élevée. Cette protéine est cependant plus courte, il en existe 2 types  $A\beta_{1-40}$  et  $A\beta_{1-42}$ . L'ajout de deux résidus (une isoleucine et une alanine) en fin de chaîne engendre une toxicité plus élevée de la protéine [100]. La séquence d' $A\beta$  est DAEFRHDSGYEVVHQQKLVFFAEDVGSNKGAIIGLMVGGVV(IA). On peut identifier dans cette séquence une forte présence de valines et de glycines en fin de chaîne VGGVV probablement propice à la formation de tresse. 10 trajectoires de dynamique moléculaire (NVT) avec UNRES ont été réalisées dans les mêmes conditions et pour la même durée que le segment 38-53 seul pour chacune des protéines  $A\beta_{1-40}$  et  $A\beta_{1-42}$ . Ces simulations sont motivées par une publication précédente réalisée avec UNRES en 2010 et dont les résultats sont montrés à la figure 5.14 pan-

neau (A) [101]. 3 types de structures différentes prises par  $A\beta_{1-40}$  ont été identifiées en fonction de leur énergie. La conformation la plus probable présente une hélice (conformation (1)). Il est possible d'identifier sur la représentation 3D de la conformation (1) une structure semblable à une tresse dans le C-terminal. Cette remarque s'applique également à la conformation la moins probable présentant des feuillets  $\beta$  (conformation (2)). Les structures 3D représentatives de nos simulations pour  $A\beta_{1-40}$  sont données figure 5.14 panneau (B) et montrent les 3 types de conformations retrouvées. Les conformations (1) et (2) sont semblables à celles retrouvées dans la ref [101] et la troisième (3) est une autre conformation où une partie de l'hélice a été tronquée pour former une tresse.

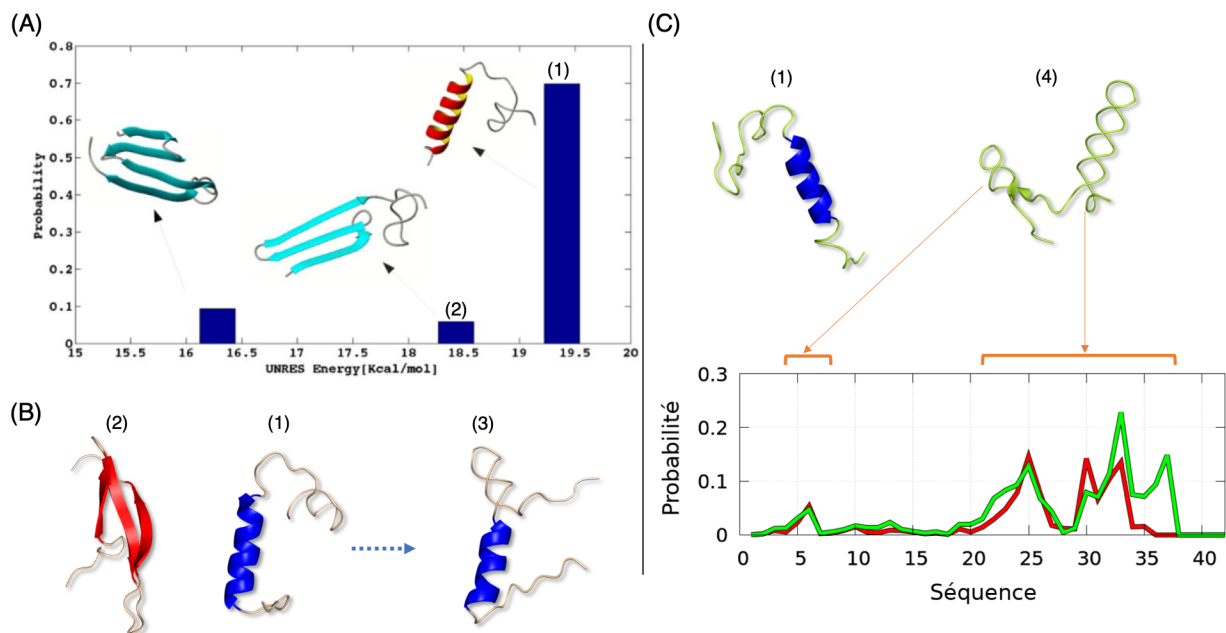


FIGURE 5.14 – Panneau A : Proportion des populations retrouvées dans des simulations réalisées avec UNRES sur  $A\beta_{1-40}$  et structures 3D correspondantes [101].

Panneau B : Représentations 3D de structures d' $A\beta_{1-40}$  retrouvées dans nos simulations.

Panneau C : Probabilité de formation d'angle favorable au repliement d'une tresse pour  $A\beta_{1-40}$  (en rouge) et pour  $A\beta_{1-42}$  (en vert) et représentations 3D des structures d' $A\beta_{1-42}$  retrouvées dans nos simulations.

Le graphique donné à la figure 5.14 panneau (C) représente la probabilité qu'un résidu possède les angles propices à la formation de la tresse calculée pour les trajectoires d' $A\beta_{1-40}$  (courbe rouge) et  $A\beta_{1-42}$  (courbe verte). On remarque pour  $A\beta_{1-40}$  une probabilité assez élevée à partir du résidu 20 jusqu'au résidu 33. On retrouve 3 pics majeurs pour G25, A30 et G33. On remarque que G25 est précédé d'une valine V24 ainsi que A30 précédé de V29. Le résidu G33 est également précédé de deux isoleucines et suivi par une alanine qui sont deux types de résidus très hydrophobes.

Pour  $A\beta_{1-42}$ , la probabilité de former une tresse est similaire à  $A\beta_{1-40}$  mais reste élevée jusqu'au résidu 37 soit 4 résidus de plus qu' $A\beta_{1-40}$ . Le résidu 37 est une glycine précédée une nouvelle fois d'une alanine A36. Le pic du résidu G33 est également plus élevé chez  $A\beta_{1-42}$ . Deux exemples de conformations retrouvées dans nos simulations pour  $A\beta_{1-42}$  sont donnés au

dessus du graphique. On retrouve majoritairement une conformation avec une hélice semblable à la conformation (1) de  $A\beta_{1-40}$  mais également des conformations très rares statistiquement présentant des très grandes tresses (conformation (4)).

## 5.4 Généralisation de l'étude des boucles dans les monomères d'alpha-synucléine

On peut étendre l'analyse des hairpins à la totalité de la séquence de la protéine pour détecter si d'autres régions de la protéine sont également propices à leurs formations. La courbe de formation de boucles à deux résidus détectables par CUTABI au sein des hairpins par rapport à la séquence est donnée [figure 5.15](#) (en rouge) accompagnée de la courbe de probabilité de formation de feuillet  $\beta$  (en vert) et d'hélice (en bleu) pour les monomères du WT de l' $\alpha$ -syn. Les courbes des hélices et des feuillets  $\beta$  ont déjà été discutées dans notre publication sur les monomères [\[1\]](#) ([chapitre 3](#)). La courbe rouge vient s'intercaler dans plusieurs minima de la courbe verte montrant ainsi qu'une partie des feuillets  $\beta$  formés sont des hairpins. Les pics rouges n'étant pas de la même hauteur que les pics verts on peut en déduire que les hairpins ne sont pas la conformation principale dans laquelle les résidus s'arrangent pour former des feuillets  $\beta$ . Ainsi, des feuillets  $\beta$  à moyenne et longue distance mais également des hairpins constituées de plus de deux résidus dans leur boucle, se retrouvent en majorité dans notre ensemble de monomères du WT.

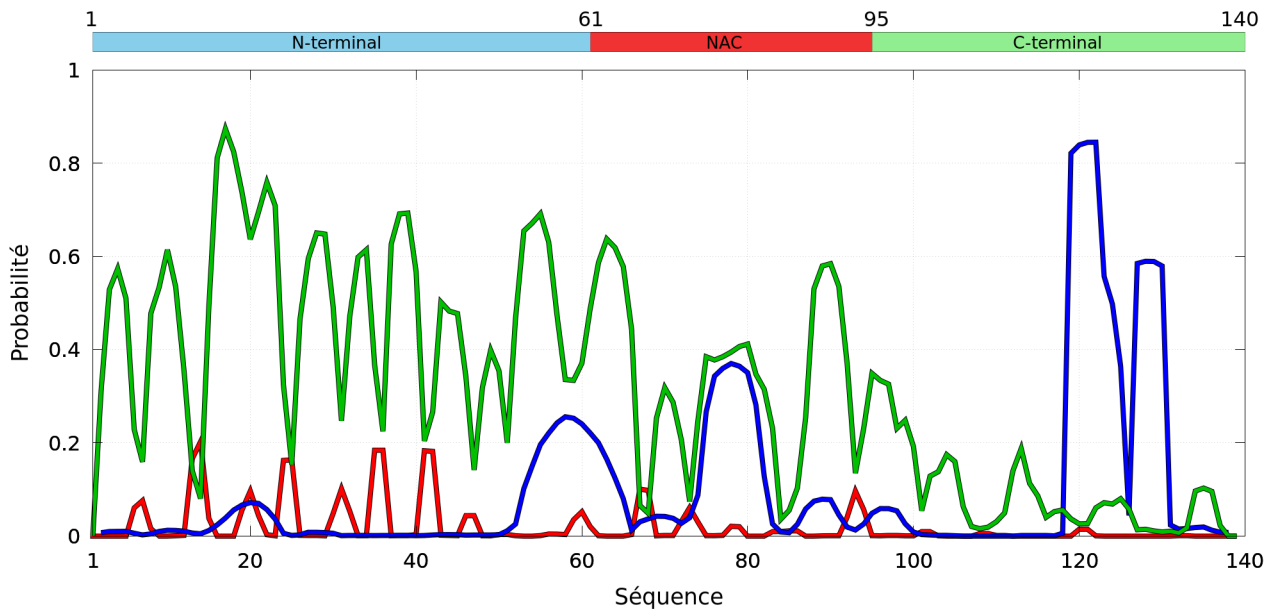


FIGURE 5.15 – Probabilité par résidu de formation d'hélice (bleu), de feuillet  $\beta$  (vert) et de boucle dans une hairpin (rouge) pour les monomères du WT d' $\alpha$ -syn.

On retrouve un grand nombre d'hairpins différentes possibles dans la protéine majoritairement localisées dans la région du N-term dans laquelle on compte 4 pics majeurs correspondant à 4 couples de résidus. On retrouve  $H_{13-14}$  ( $p=0.21$ ),  $H_{24-25}$  ( $p=0.19$ ), ainsi que les deux pics

des  $H_{35-36}$  et  $H_{41-42}$ . Le hairpin  $H_{46-47}$  ne se présente pas comme un pic majeur dans cette courbe. Dans le NAC on retrouve deux régions majeures : un pic correspondant à la double glycine caractéristique d'une grande flexibilité  $H_{G67-G68}$  ( $p=0.19$ ), et un pic un peu plus large intervenant dans 2 couples  $H_{T92-G93}$  et  $H_{G93-F94}$  avec  $p_{T92}=0.05$ ,  $p_{G93}=0.10$  et  $p_{F94}=0.06$ . Le pic en K60 est proche du résidu le plus probable en hélice K58. La probabilité en K60 de formation de boucle est très faible pourtant la probabilité de former un feuillet  $\beta$  présente deux pics bien distincts dans cette région dont les maxima sont formés à V55 et V63. La probabilité de formation de feuillet  $\beta$  des résidus K58, T59 et K60 est inférieure à 0.4, ce qui signifie qu'on peut avoir de manière assez probable une hairpin possédant une boucle à 3 résidus indétectable par CUTABI. Ces résidus ne contribuent donc pas à la probabilité de formation des boucles représentée.

L'autre possibilité est que cette région est en mesure de former des feuillet  $\beta$  à plus longue distance. On peut s'intéresser aux deux résidus dont la probabilité est maximale dans cette région V55 et V63. En utilisant CUTABI, on peut sortir la probabilité pour chaque résidu d'être en interaction sous forme de feuillet  $\beta$  avec le résidu étudié dans l'ensemble des monomères. La courbe correspondante est donnée [figure 5.16](#) pour V55 et V63. Les pics majoritaires sont l'interaction entre ces deux résidus sous forme d'une hairpin. On retrouve à probabilité égale une interaction de V63 avec T72 probablement sous forme d'une autre hairpin. V55 réalise plusieurs feuillet  $\beta$  à longue distance avec le Nterm mais également avec le reste du NAC et même avec certains résidus du Cterm. V63 favorise également un feuillet  $\beta$  avec le résidu A19 avec un probabilité de près de 0.1 et interagit également avec le reste du NAC mais très peu avec le Cterm. Les résidus au voisinage du résidu 60 de la protéine sont donc capables de former une hélices ou un feuillet  $\beta$  sous forme d'une hairpin ou un feuillet  $\beta$  à longue distance. Ils sont donc probablement très importants dans la protéine.

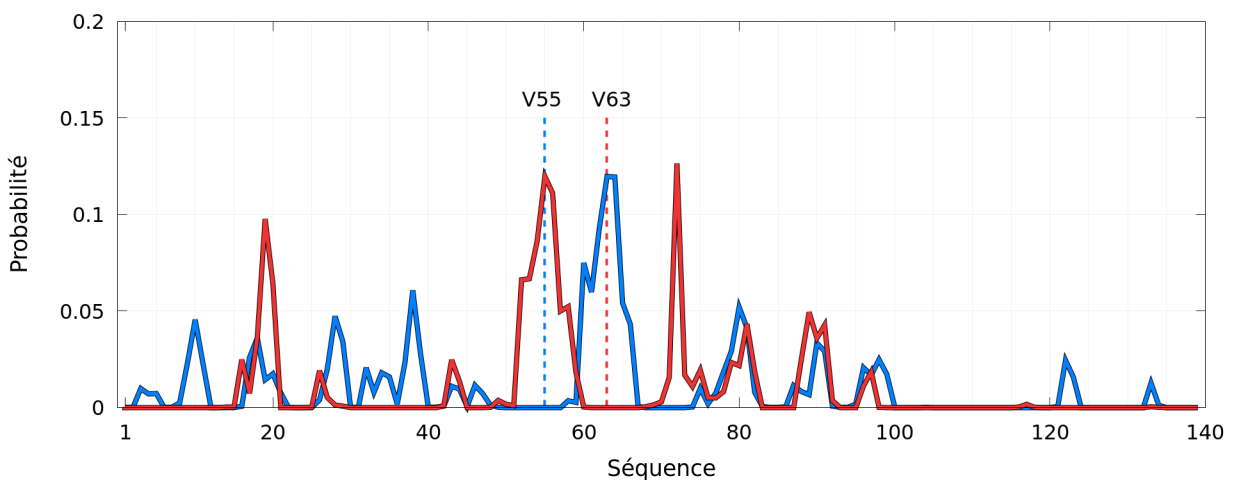


FIGURE 5.16 – Probabilité de formation de feuillet  $\beta$  entre un résidu et le reste de la séquence dans l'ensemble des monomères isolés de la forme WT d' $\alpha$ -syn. Les deux résidus étudiés V55 et V63 sont identifiés sur la séquence par une barre pointillée de la même couleur que leur courbe correspondante (bleue et rouge respectivement)

## 5.5 Analyses des trois types de structures dans les dimères

L'étude du segment 35-53 a été étendue au dimère, pour déterminer si ces structures persistent malgré la dimérisation. Les 3 types de structures secondaires  $H_{41-42}$ ,  $H_{46-47}$  et la tresse du segment 38-53 ainsi que les  $H_{35-36}$  et  $H_{35-36/41-42}$  ont été recherchées pour chaque monomère des dimères. On réalise ainsi un histogramme de probabilité d'apparition de chaque type de repliement pour l'ensemble des monomères dans les dimères donné [figure 5.17](#).

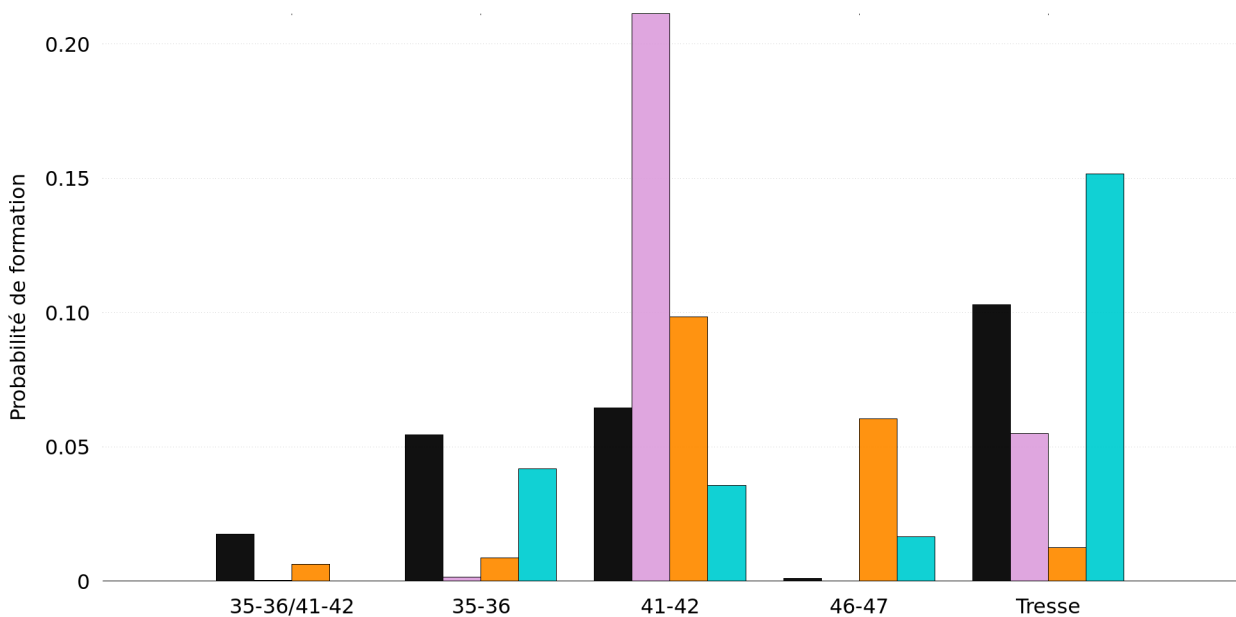


FIGURE 5.17 – Histogramme de probabilité de formation de hairpins suivant la position de la boucle constituée de deux résidus dans l'ensemble des monomères dans les dimères. Le label des boucles est donnée avec  $X-(X+1)$  sur l'axe des abscisses. Le cas 35-36/41-42 donne la probabilité de formation des deux boucles simultanément.

En comparant la [figure 5.17](#) des monomères dans les dimères avec la [figure 5.4](#) des monomères isolés, on remarque que les probabilités sont complètement différentes.

L'apparition de contacts inter-chaines réduit à une probabilité de moins de 0.02 la proportion de la tresse dans A53T alors qu'elle était de pratiquement 0.3 dans le cas des monomères isolés. On observe une plus faible diminution de cette probabilité pour les autres mutants et le WT. La mutation A53T semble avoir déstabilisé la tresse pour favoriser plutôt des  $H_{46-47}$  qui ne se formaient pratiquement pas dans les monomères isolés. Dans les dimères, la  $H_{46-47}$  est inexistante pour le WT et A30P ( $p < 0.01$ ) et très peu présente chez E46K ( $p = 0.01$ ). Le changement majeur intervient pour la  $H_{41-42}$  dont la probabilité d'apparition a été fortement réduite pour le WT passant d'environ 0.2 dans les monomères isolés à seulement 0.07 dans les monomères des dimères. A30P connaît à l'inverse une forte augmentation de la présence de la  $H_{41-42}$  avec une probabilité de formation supérieure à 0.2 alors qu'elle était inférieure à 0.1 dans les monomères isolés. Les probabilités concernant E46K et A53T ont légèrement augmenté pour cette hairpin. Le WT continue de former cependant plus de  $H_{35-36}$  que les mutants mais avec une probabilité divisée par 3 par rapport aux monomères isolés.

Le comportement intrinsèque du segment 38-53 d'un monomère est donc fondamentalement différent lorsque celui-ci forme des liaisons inter-chaines avec un autre monomère.

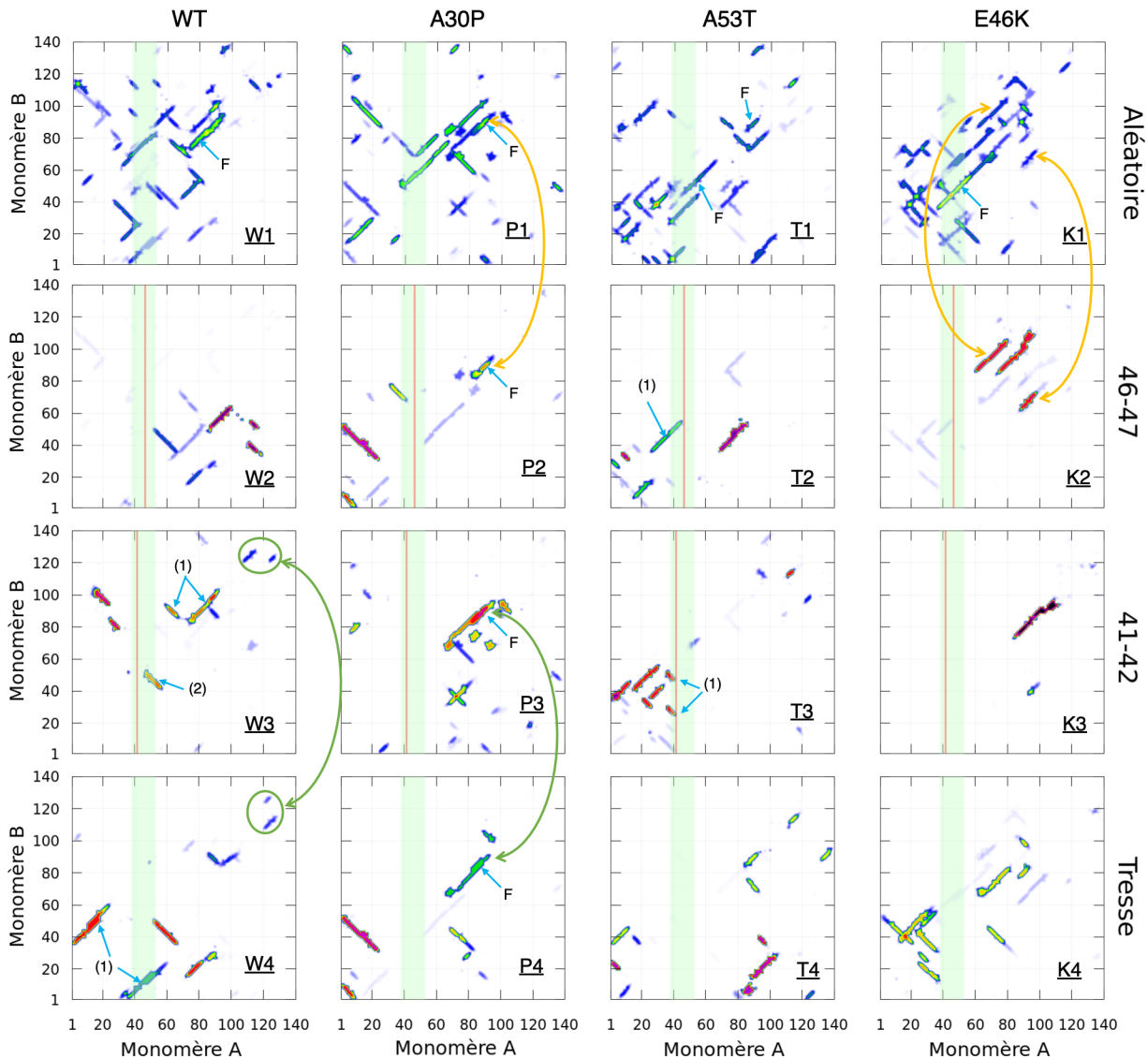


FIGURE 5.18 – Cartes de contact inter-chaines des dimères dont un des monomères possède sur son segment 38-53 soit une  $H_{41-42}$ , une  $H_{46-47}$ , une tresse ou une configuration aléatoire. Chaque ligne identifie un type de structure sur le segment 38-53 et chaque colonne une protéine différente. L'abscisse des cartes est la séquence du monomère possédant la structure attribuée à la ligne (Monomère A). L'ordonnée est la séquence de l'autre monomère formant le dimère (Monomère B). La région de contact inter-chaines du segment 38-53 du monomère A est mise en évidence par un rectangle vert. Une ligne rouge complète cette région pour pointer la position de la boucle dans le cas des hairpins.

Les dimères dont au moins un monomère possède un type de repliement sur son segment 38-53, soit  $H_{41-42}$ ,  $H_{46-47}$  ou la tresse ont été regroupés dans des classes associées au type de repliement. Il est possible qu'une conformation de dimère se retrouve dans deux classes différentes dans le cas où les deux monomères possèdent chacun le segment 38-53 replié de manière différente. Si les deux monomères ont le même type de repliement pour le segment 38-53 le dimère va apparaître deux fois dans la même classe correspondant à ce repliement. La classe



"aléatoire" ne regroupe que les dimères ne possédant aucune des structures particulières sur son segment 38-53 pour les deux monomères le composant. On réalise des CM inter-chaines de chaque classe pour chaque protéine (figure 5.18). La séquence du monomère donnée en abscisse à la figure 5.18 correspond à celle possédant le repliement associé à la classe. Le segment 38-53 est identifié par un rectangle vert pour permettre de visualiser les interactions de ce segment avec l'autre chaîne. Les positions des boucles sont également données par un trait vertical rouge.

La première ligne de la figure 5.18 (CM W1,P1,T1 et K1) présente le cas où un des monomères dans le dimère ne possède pas de structure particulière pour son segment 38-53. Dans ce cas de figure, pour les 4 protéines, une grande variété de contacts est possible, y compris sur le segment 38-53. On remarque la présence de traces sur la diagonale correspondant à des CTF (identifiées par la lettre F figure 5.18). La couleur de ces traces indique une probabilité plus élevée que les autres traces de la carte pour le WT, A53T et E46K (CM W1,T1 et K1 respectivement). Cette probabilité est plus faible pour A30P (CM P1). A30P réalise un plus grand nombre de CTFs lorsqu'il possède des hairpins  $H_{41-42}$  ou  $H_{46-47}$  (CM P2 et P3). On remarque que pour les CM des dimères dont un monomère possède une hairpin (CM W2, P2, T2, K2, W3,P3,T3 et K3), le segment 38-53 de ce monomère ne possède pas une grande affinité avec l'autre monomère comparé au cas aléatoire (CM W1,P1,T1 et K1) Dans le cas de la  $H_{46-47}$ , on observe uniquement une interaction pour A53T où un brin du  $H_{46-47}$  interagit avec l'autre monomère sous forme d'un feuillet  $\beta$  parallèle (Trace (1) sur la CM T2). On observe également une possibilité de feuillets  $\beta$  composés de très petits brins sur la  $H_{41-42}$  (Trace (1) sur la CM T3). Pour le WT, des contacts sont possibles entre le segment 38-53 en présence d'une  $H_{41-42}$  (trace (2) sur la CM W3). Les traces (1) et (2) de la carte W3 sont illustrées dans une structure de dimères à la figure 5.19 (A). La conformation présentée montre la présence de la  $H_{41-42}$  combinée à la  $H_{35-36}$ . Le monomère possédant ces deux structures va venir réaliser un feuillet  $\beta$  anti-parallèle avec l'autre monomère donnant la trace (2) sur la CM W3. Ce feuillet  $\beta$  va former un plan  $\beta$  avec 3 autres brins dont le dernier se forme de manière parallèle dans le NAC des deux monomères proche de la diagonale (trace (1) sur la CM W3). La présence d'une tresse dans un dimère empêche la formation de contact du segment 38-53 avec l'autre monomère chez les mutants (CM P4,T4 et K4). Seul le WT parvient étonnamment à former des contacts sur ce segment. L'analyse 3D d'une conformation de dimère contenant une tresse est donnée figure 5.19 (B). Les traces (1) données sur la CM W4 concernent la formation d'une "tresse inter-chaines". Les chaînes des deux monomères vont venir s'enrouler à la manière d'une tresse. Les angles diédraux des résidus concernés sont les mêmes que celui d'une "tresse intra-chaîne, CUTABI détecte donc la présence d'une tresse. La trace (1) sur la CM W4 possède une symétrie puisque l'autre monomère possède également une tresse facilement identifiable sur la représentation 3D (figure 5.19 (B)). Ces symétries se retrouvent également entre les différentes cartes notamment entre W4 et W3 où des contacts dans le Cterm se retrouvent sur les deux cartes. Cette observation signifie que le monomère B associé à la CM W4 possède parfois une  $H_{41-42}$ . On le retrouve donc comme monomère A sur la CM W3 inversant les tâches dans le Cterm par symétrie par rapport à la diagonale. Ce phénomène se retrouve également entre les

CM P3 et P4 pour A30P. Un dernier phénomène intéressant est la présence de tâches similaires entre les cartes signifiant que le monomère A a modifié sa structure au sein du dimère. On retrouve ce phénomène entre la CM P2 et P1 pour A30P mais aussi entre K1 et K2 pour E46K avec deux tâches majeurs identifiables lors de la présence d'une  $H_{46-47}$ .

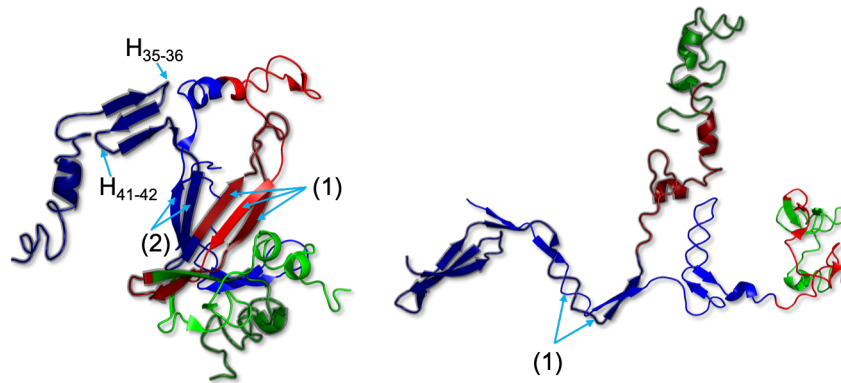


FIGURE 5.19 – Représentations 3D de dimères du WT. Les régions Nterm, NAC et Cterm sont colorés en bleu, rouge et vert respectivement. Des couleurs plus sombres ou plus claires permettent de différencier les deux monomères dans le dimère.

(A) : Dimère possédant une  $H_{41-42}$  associé à la carte W3 [figure 5.18](#).

(B) : Dimère possédant une tresse inter-chaînes sur une des deux chaînes et intra-chaîne sur l'autre chaîne.

## 5.6 Conclusion : Études du segment 38-53 de l' $\alpha$ -synucléine

L'extension de CUTABI à la détection de hairpin constituée d'une boucle à deux résidus a permis une analyse approfondie du segment 38-53. Nous avons pu identifier 3 types de repliement différents dans cette région, la hairpin, la tresse et le repliement aléatoire. Les hairpins retrouvées ne forment pas leur boucle constamment sur les mêmes résidus mais en deux positions dont une majoritaire 41-42 et une minoritaire 46-47. On a également identifié une autre hairpin proche de cette région constituée d'une boucle 35-36 dont le couplage est possible avec la hairpin  $H_{41-42}$ .

La comparaison entre les mutants et le WT a montré une grande différence de probabilité de formation de  $H_{41-42}$ ,  $H_{35-36}$  et du couplage des deux qui se forment bien plus fréquemment pour le WT. Les impacts de ces hairpins sont très variables d'une protéine à l'autre et leur longueur est également changeante. L'impact de la tresse sur le reste de la protéine est difficilement identifiable. A l'échelle des dimères, les présences de ces structures dans les monomères sont totalement modifiées pour le WT et ses mutants.

L'étude approfondie de la tresse valide sa possible apparition comme un état métastable dans la protéine  $\alpha$ -syn et a permis de faire un lien possible avec la protéine  $A\beta$  impliquée dans la maladie d'Alzheimer.

L'étude généralisée des hairpins sur le reste de la séquence a mis en évidence la région 53-64 capable de former plusieurs types de structures secondaires tels que des hélices, des feuillets  $\beta$  à longue distance et des hairpins.

# Chapitre 6

## Resultats : Mutations répétées sur le motif KTKEGV

### 6.1 Les mutants

L' $\alpha$ -syn possède le motif KTKEGV répété environ 7 à 9 fois dans sa séquence. Il a été suggéré que ce motif pouvait jouer un rôle dans la formation des tétramères et sur la toxicité de la protéine [58] (1 section 1.4 page 14).

Parmi les mutants étudiés dans la référence [58], nous avons sélectionné 5 mutants, 4 considérés comme très toxiques et ne formant pas de tétramères et 1 considéré comme peu toxique formant peu d'inclusions dans les cellules comme le WT, nous les décrivons ci-dessous :

#### Les mutants très toxiques :

- le mutant KTKKGV : les acides glutamiques (E) en 4<sup>ème</sup> position du motif initial sont remplacés par des lysines (K). Les acides glutamiques étant chargés négativement et les lysines positivement la charge globale de la protéine est donc considérablement modifiée . Ce mutant possède donc la mutation E46K. Nous noterons par la suite ce mutant "Kmut".

- le mutant KLKEGV : les thréonines (T) en 2<sup>ème</sup> position du motif initial sont remplacées par des leucines (L). Les leucines sont des résidus très hydrophobes contrairement aux thréonines. L'hydrophobicité de la protéine déjà très élevée par son grand nombre de valines et d'alanines et elle est donc renforcée dans ce mutant. Nous noterons par la suite ce mutant "Lmut".

- le mutant KTKEGW : les valines (V) en 6<sup>ème</sup> position du motif initial sont remplacées par des tryptophanes (W). La chaîne latérale de la valine n'est constituée que de 3 groupements carbones-hydrogènes tandis que le tryptophane possède une chaîne latérale cyclique constituée de 9 carbones et 1 azote. L'encombrement stérique est donc bien plus important dans le cas de ce mutant. Nous noterons par la suite ce mutant "Wmut".

- le mutant KTKEIV : les glycines (G) en 5<sup>ème</sup> position dans le motif initial sont remplacées par des isoleucines(I). La glycine est un casseur d'hélice tandis que les isoleucines sont très

favorables à la formation d'hélice. On retrouve donc une forte expression des hélices dans nos simulations. Des hélices se retrouvent présentes sur la globalité de la protéine excepté dans le Cterm. L'analyse de ce mutant n'a donc pas été poursuivie pour les dimères.

Le mutant peu toxique :

- le mutant KTEEGV : les lysines (K) en 3<sup>ème</sup> position du motif initial sont remplacées par des acides glutamiques (E). Contrairement à Kmut, on ajoute des charges négatives en enlevant des charges positives. Malgré cette modification considérable, ce mutant est considéré comme peu toxique (semblable au WT) et parvient à former des tétramères. Nous noterons par la suite ce mutant "Emut".

Les séquences de ces différents mutants sont données [figure 6.1](#). Ces 4 protéines ont été simulées en utilisant UNRES puis triées suivant les conditions présentées dans la partie [section 2.1](#) page [23](#).

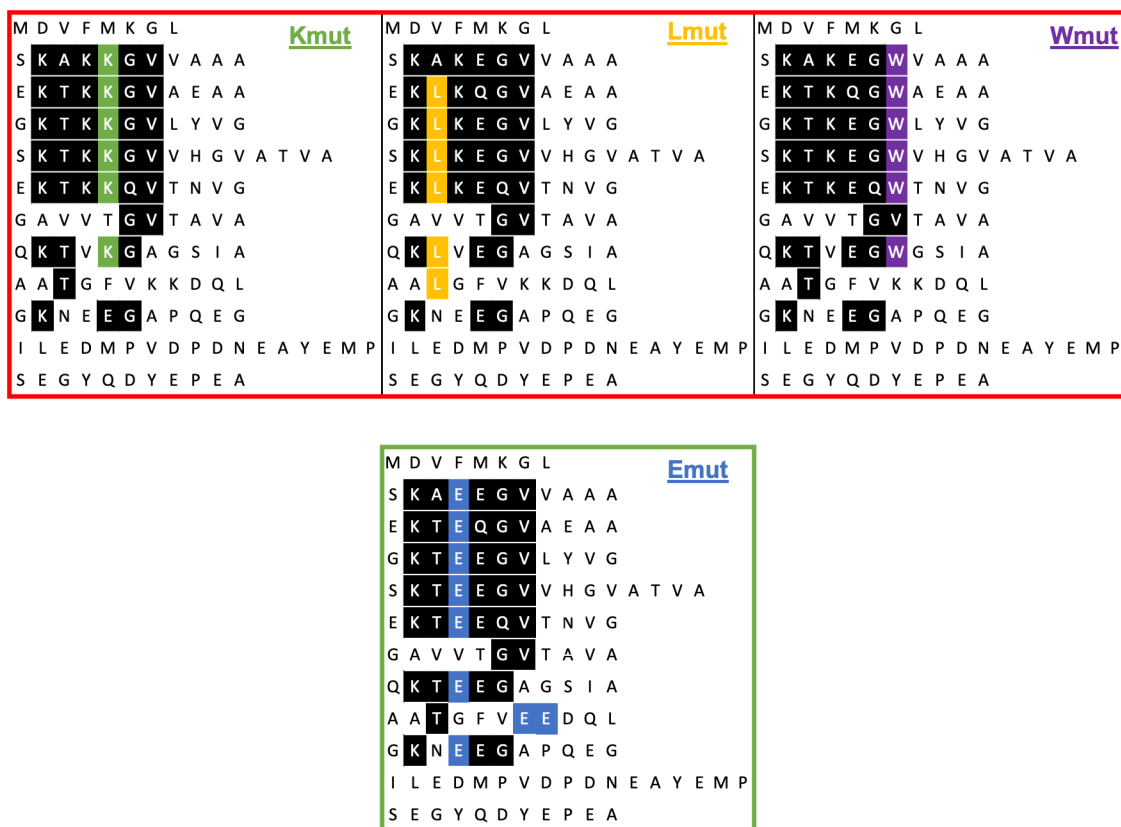


FIGURE 6.1 – Séquence des mutants étudiés. Les mutants dans le cadre rouge sont considérés comme très toxiques dans l'étude ref [\[102\]](#) et le mutant dans le cadre vert est considéré comme peu toxique. Les mutations sont mises en évidence en couleur, le motif répété est représenté en noir. Le code couleur de chaque mutation sera utilisé pour distinguer les mutants dans les autres figures.

## 6.2 Etude des monomères

### Probabilité de formation de structures secondaires le long de la séquence

Le programme de détection d'hélices et de feuillets  $\beta$  CUTABI a été appliqué sur l'ensemble des données de simulation des monomères de ces quatre mutants permettant de définir la probabilité de formation des hélices et des feuillets  $\beta$  le long de la séquence donnée à la [figure 6.2](#).

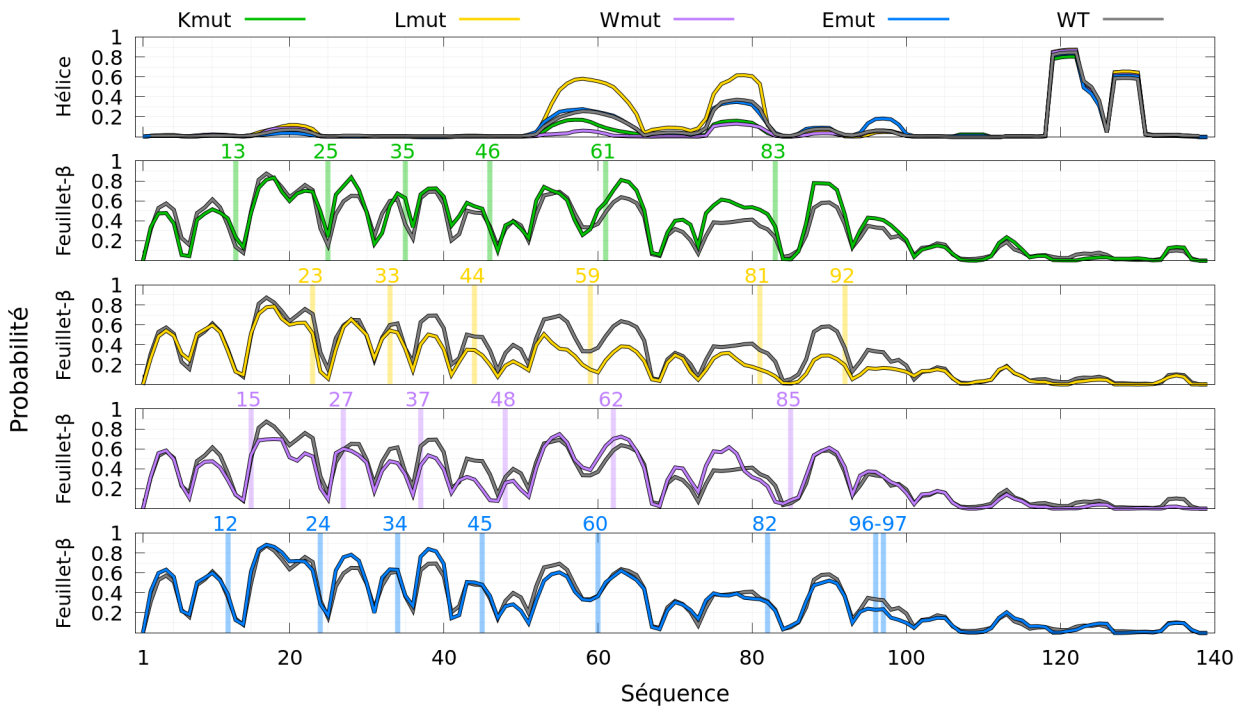


FIGURE 6.2 – Probabilité des structures secondaires le long de la séquence pour les 4 mutants Kmut (vert), Lmut (jaune), Wmut (violet) et Emut (bleu) comparée à celle du WT (gris). La première courbe donne la probabilité pour les hélices pour les 5 protéines. Chacune des 4 autres courbes donnent la probabilité pour les feuillets  $\beta$  comparé au WT pour un mutant. Les positions des mutations sont mises en évidence par un trait vertical de la couleur attribuée au mutant surmonté de l'index du résidu.

Nous discutons dans un premier temps la formation d'hélices le long de la séquence ([figure 6.2](#) 1<sup>ère</sup> courbe). Wmut et Kmut défavorisent l'hélice 53-64 par rapport au WT avec une probabilité maximum  $P_{Wmut}(58)=0.06$  et  $P_{Kmut}(57)=0.17$  comparés à  $P_{WT}(58)=0.25$ . La forme de la courbe pour Kmut est asymétrique avec une décroissance plus grande pour le résidu K61 comparé au WT ( $pK(61)=0.09$  et  $pWT(61)=0.22$ ). Lmut possède une mutation au résidu 59 qui semble doubler la proportion de cette hélice  $pL(59)=0.56$ . Des écarts similaires se retrouvent pour l'hélice 75-81 avec une nouvelle fois pour Lmut une mutation présente dans cette hélice qui fait croître sa probabilité  $pL(81)=0.52$  (maximum atteint pour  $pL(79)=0.6$ ) contre  $pWT(78)=0.37$ . Kmut et Wmut ont respectivement une mutation plus loin dans la séquence en 83 et 85, probablement responsable de la forte réduction de la probabilité du résidu

78 comparée au WT ( $pK(78)=0.16$  et  $pW(78)=0.13$ ). Le mutant Emut présente une courbe pratiquement confondue avec celle du WT, excepté pour une très petite hélice 95-99 incluant la double mutation en 96-97 avec  $pE(97)=0.18$  contre  $pWT(97)=0.06$ . La probabilité de formation des hélices dans le Cterm reste identique au WT pour les 4 mutants.

Des différences sont également notables pour la formation des feuillets  $\beta$  pour Kmut, Lmut et Wmut mais pas pour Emut qui une nouvelle fois possède une courbe très similaire à celle du WT. Pour Emut, seule la région possédant la double mutation 96-97 montre également des écarts par rapport au WT avec une diminution de la formation de feuillets  $\beta$ , probablement corrélée avec l'augmentation de la formation d'hélices. Les feuillets  $\beta$  du Nterm de Kmut se forment majoritairement comme le WT, les positions des mutations ne semblent pas avoir d'importants effets sur les probabilités, ce comportement rejoint celui de E46K. Cependant à partir du résidu 61, la probabilité de former des feuillets  $\beta$  des résidus de Kmut devient supérieure à celle du WT pour pratiquement tous les résidus du NAC, puis rejoint celle des résidus du WT dans le Cterm. Ce comportement est cette fois différent de celui du mutant E46K dont l'impact de la mutation est moindre dans le NAC. Les deux mutations E61K et E83K semblent donc jouer un rôle majeur pour Kmut autant en terme de la formation de feuillet  $\beta$  que celles des hélices. Pour le mutant Lmut, les modifications T23L et T33L ne semblent pas impacter la probabilité locale. A partir du résidu 37, la probabilité en feuillet  $\beta$  chute pour Lmut et notamment dans la région du hairpin (38-53) présentant une mutation au résidu 44 mais également dans la région de l'hélice 53-64 avec la mutation T59L, et de l'hélice 87-93 avec la mutation T92L. Pour le mutant Wmut, la probabilité de formation de feuillets  $\beta$  est réduite dans le Nterm comparée au WT. En effet, le mutant Wmut par son encombrement est logiquement moins à même de former des feuillets  $\beta$ . La formation des hairpins retrouvées majoritairement dans le Nterm pour le WT est donc probablement fortement impactée par les mutations de Wmut. Les probabilités dans le NAC sont similaires excepté dans la région 74-78 avec une augmentation de 0.2. Au delà du résidu 100, les 4 mutants possèdent des probabilités faibles semblables au WT.

## Densité de probabilité des conformations en fonction du nombre de résidus en feuillet $\beta$ et en hélice

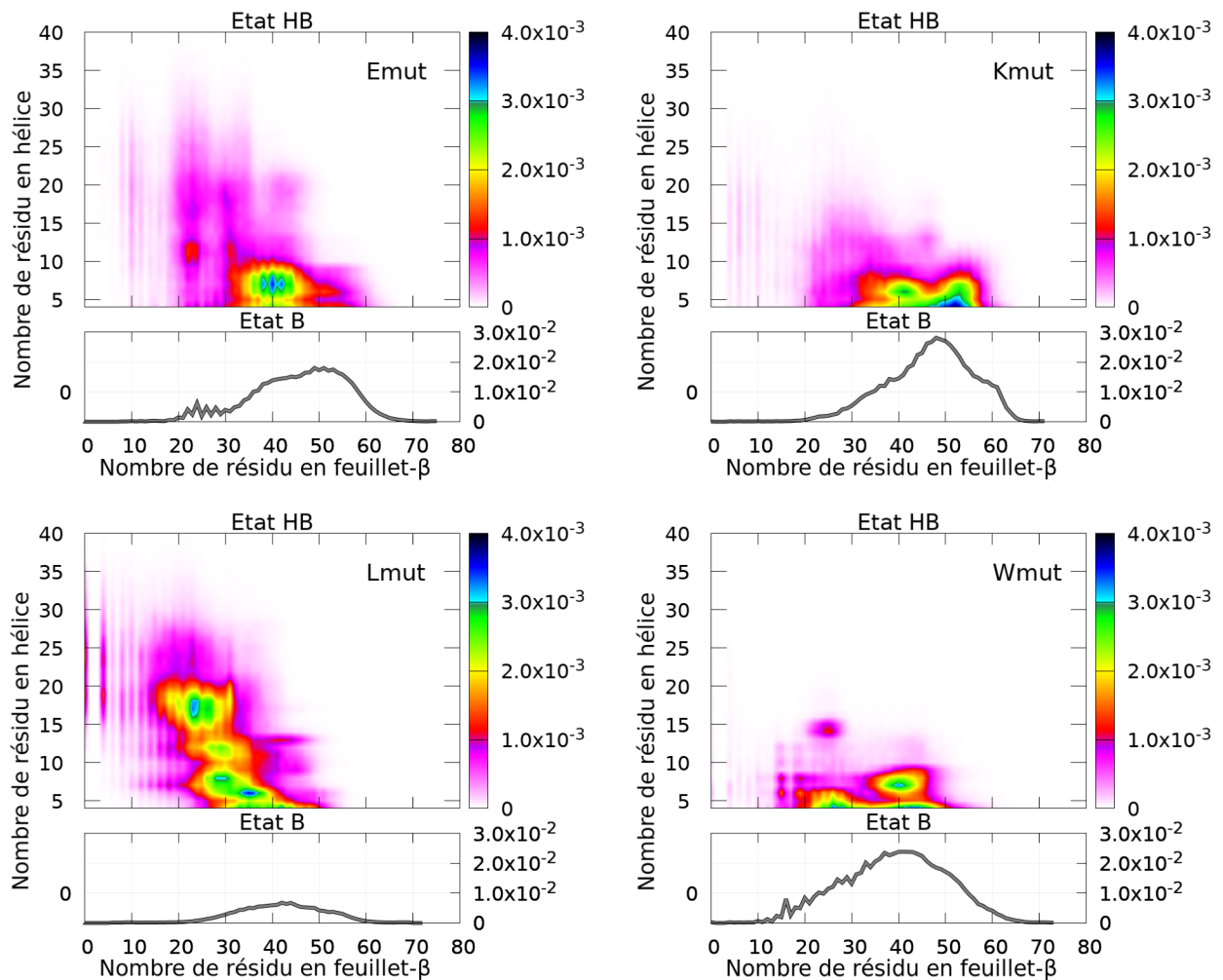


FIGURE 6.3 – Densité de probabilité du couple du nombre de résidus en hélice et du nombre de résidus en feuillet  $\beta$  de l'ensemble des monomères. L'état HB (pour Helix and Beta-sheet) est donné en deux dimensions et se réfère à l'échelle de couleur sur la droite, l'état B est donné par une courbe à une dimension se référant à l'échelle de droite. Attention dans cette figure les échelles sont différentes de celle de la figure 4 dans notre publication (3) pour le WT, A30P, A53T et E46K)

Les densités de probabilité du couple du nombre de résidus en feuillet  $\beta$  et en hélice calculées pour les mutants Emut, Kmut, Lmut et Wmut sont représentées sur la figure 6.3. Cette analyse a été réalisée pour le WT et les mutants A30P, A53T et E46K dans notre publication (voir chapitre 3). Elle avait montré la présence de deux états distincts. Un premier état HB considérant les structures possédant à la fois des hélices et des feuillet  $\beta$  et un second état B contenant les conformations ne possédant que des feuillet  $\beta$  bien plus majoritaires dans les 4 protéines étudiées. La répartition de l'état B du mutant Emut (figure 6.3) est semblable à celle du WT avec un pic pour 50 résidus en feuillet  $\beta$  d'une probabilité légèrement inférieure à 0.02. Le bassin principal de l'état HB est centré sur 40 résidus en feuillet  $\beta$  et 7 résidus en hélice, il est décalé de quelques résidus supplémentaires en feuillet pour le WT mais possède le même



nombre de résidus en hélice. L'exploration du reste de la carte est également proche de celle du WT. Étonnamment, Kmut ne possède pas une carte semblable à celle de E46K mais plutôt à celle de A30P pour l'état B qui forme un pic aux alentours de 50 résidus en feuillet  $\beta$  et un second pic deux fois moins probable situé à 60 résidus en feuillet  $\beta$ . A30P possède en effet une allure de courbe similaire légèrement décalée vers des valeurs plus élevées (52 et 65 respectivement) avec une probabilité d'apparition plus faible de 0.02 contre 0.03 pour l'état principal chez Kmut. L'état HB ne possède que peu d'états avec un état principal (50,4) correspondant à la limite de détection des hélices pour CUTABI. Les états au dessus de 10 résidus en hélice ne sont que très faiblement explorés. Lmut se rapproche d'A53T par sa formation prononcée d'hélices (voir [figure 6.2](#) et [chapitre 3](#) figure 4). L'étude de la carte ([figure 6.3](#)) le confirme, puisque Lmut visite une grande variété d'états possédant des hélices et un état B très peu peuplé. La carte de Wmut possède des légères similitudes avec celle de E46K, notamment un état principal d'environ 40 résidus en feuillet  $\beta$  pour l'état B mais également un état principal (28,4) assez proche de l'état (32,7) de E46K dans l'état HB. De façon remarquable, un autre état moins probable (28,14) pour Wmut est proche de l'état (20,17) pour E46K. L'état B de Wmut se caractérise par la formation significative de structures possédant peu de feuillet  $\beta$  (entre 10 et 30 résidus) très peu probables chez toutes les autres protéines.

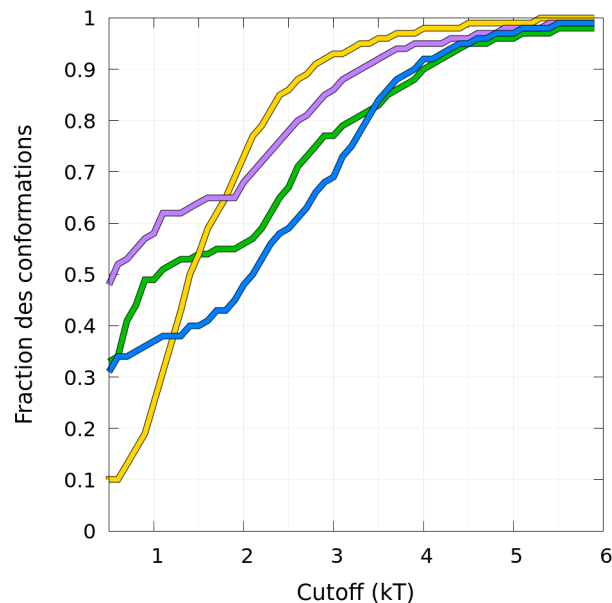


FIGURE 6.4 – Courbe cumulative du nombre total de conformations simulées de l'ensemble des monomères en fonction du cutoff d'énergie libre (en unité kT) à partir du minimum global de l'état B représenté [figure 6.3](#). Les courbes donnent Kmut(vert), Lmut(jaune), Wmut(violet) et Emut(bleu).

Les résultats présentés sur la [figure 6.4](#) mettent en évidence les différents états possibles visités par les différents mutants. Par définition, la dérivée de ces courbes représente la densité des conformations (DOS pour Density of micro-States). Similairement au WT et E46K, le comportement linéaire de la courbe d'Emut jusqu'à environ 2kT donne une DOS constant correspondant à l'état B. Cet état se peuple à hauteur de 42% (environ 41% pour WT et

E46K) du nombre global de conformations avant un changement brutal du coefficient directeur à partir de 2kT montrant l'apparition de l'état HB avec une DOS plus grand. Après 3.5kT, la DOS diminue jusqu'à atteindre un plateau. Avant 3kT, les fluctuations de ces courbes sont très diverses chez les mutants. Lmut, dont une grande partie des caractéristiques se rapproche de A53T, ne possède que très peu de conformations dans l'état B (15%). L'apparition de l'état HB se fait très tôt à environ 1kT avec une DOS très élevée, bien supérieure à tous les mutants étudiés et au WT. Kmut forme un premier état jusqu'à 1kT avec une DOS très élevée correspondant à 50% des conformations, il s'agit du pic principale retrouvé dans l'état B. De 1kT à 2kT un plateau avec une DOS plus faible se forme correspondant au peuplement du reste de l'état B jusqu'à 55% des conformations. Après 2kT, l'état HB commence à se peupler et la courbe repart donc avec une DOS plus grande. Wmut se comporte d'une façon similaire à Kmut excepté qu'au lieu d'avoir deux paliers distincts jusqu'à 2kT, la courbe a une forme arrondie correspondant à l'état B. Cet état B représente 65% des structures. Après 2kT, l'état HB apparait. A titre de comparaison, Wmut et Kmut se rapprochent du comportement de A30P.

## Classification global : Hairpin, Tresse, Aléatoire

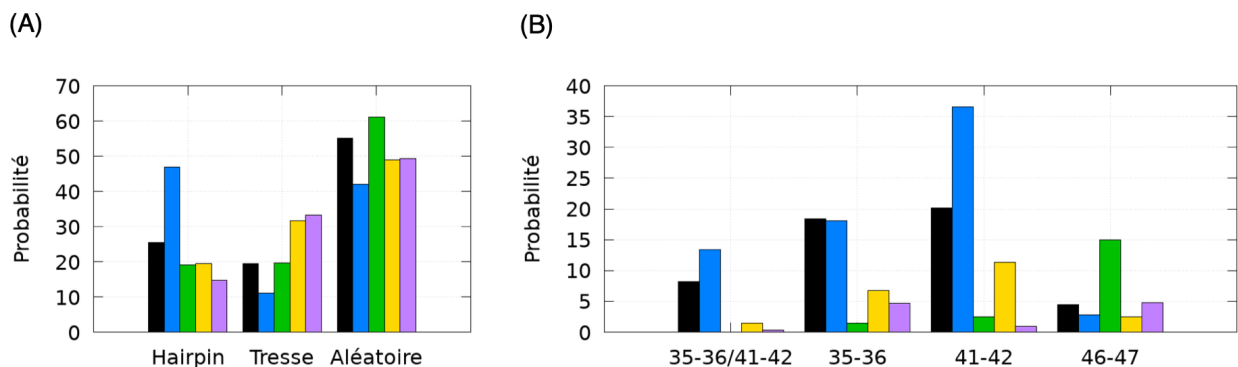


FIGURE 6.5 – panneau (A) : Proportion des différentes conformations prises par la région 35-53 dans les monomères des mutants. panneau (B) : Proportion des différents hairpins retrouvés dans la région 35-53 dans les monomères des mutants. Les valeurs données en abscisses sont les résidus composant la boucle du hairpin. La boucle 35-36/41-42 signifie que les deux boucles sont présentes en même temps. Le code couleur est Kmut(vert), Lmut(jaune), Wmut(violet) et Emut(bleu)

L'étude sur la région 35-53 menée dans le chapitre précédent sur les protéines WT, A30P, A53T et E46K a été reproduite pour ces mutants. On donne pour le segment 35-53 la proportion des 3 types de conformations hairpin, tresse et aléatoire sur la figure [figure 6.5](#) panneau (A). Il est clairement identifiable que Emut forme bien plus de hairpin dans cette région que les autres mutants très toxiques. Il en forme d'ailleurs même plus que le WT, les autres mutants en forment également moins que le WT mais plus que les 3 mutations ponctuelles (voir [figure 5.3](#) page [92](#)). En détaillant cette classe formant des hairpins ([figure 6.5](#) panneau (B)) on trouve une proportion très grande de  $H_{41-42}$  pour Emut et une très faible pour les autres mutants. Kmut favorise plutôt la  $H_{46-47}$  tandis que Lmut formera majoritairement  $H_{41-42}$  et  $H_{35-36}$ . La

combinaison  $H_{35-36}$  et  $H_{41-42}$  est pratiquement inexistante chez les mutants très toxiques alors quelle représente presque 15% pour Emut. Wmut ne forme que très peu  $H_{46-47}$  et  $H_{35-36}$  et pratiquement pas  $H_{41-42}$  en raison de son fort encombrement. Ce résultat pourrait expliquer la décroissance du nombre de feuillets  $\beta$  dans le Nterm comparé au WT.

### Lien avec la forme globale de la protéine

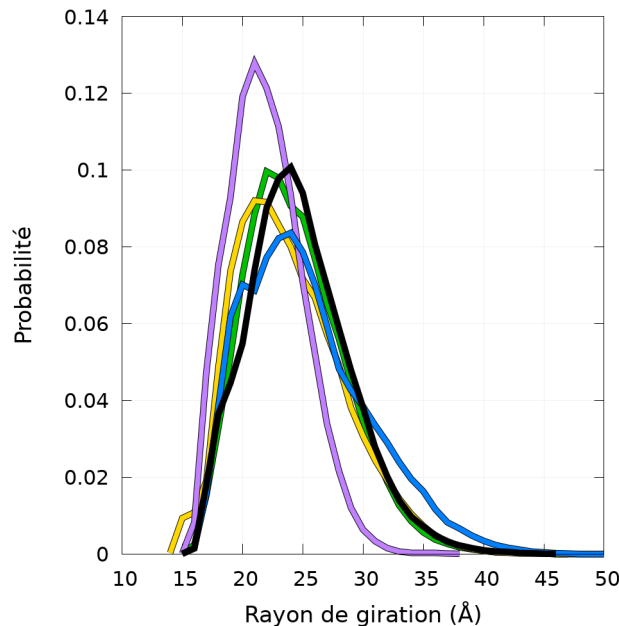


FIGURE 6.6 – Distribution du rayon de giration pour les monomères des 4 mutants et du WT avec Kmut(vert), Lmut(jaune), Wmut(violet) , Emut(bleu) et WT(noir).

Les données précédemment explicitées peuvent être reliées au rayon de giration (figure 6.6). Les maxima du WT et de Emut sont de 24 Å. Malgré les nombreuses similitudes de ces deux protéines dans les analyses précédentes, les courbes possèdent des différences avec un léger coude à 20 Å pour Emut et une proportion plus grande de structure au delà de 30 Å. L'origine de ces différences pourrait provenir de la formation plus élevée de la  $H_{41-42}$  chez Emut. La courbe de Kmut est légèrement décalée avec un maximum à 22 Å mais suit ensuite la courbe du WT à partir de 25 Å. Lmut et Wmut présentent un maximum à 21 Å. La courbe de Wmut se démarque des autres avec un pic beaucoup plus fin. Les conformations possédant un rayon de giration au delà de 30 AA sont très rares. Le nombre de feuillets  $\beta$  est pourtant bien moindre dans le Nterm de Wmut que pour les autres protéines (figure 6.1 page 114). La figure 6.7 montre en effet que même la structure possédant 33 résidus en hélice reste très compacte chez Wmut. Un enroulement sous forme de tresse entre le NAC et le Nterm, impliquant une mutation W (flèche 1) couplé à une interaction entre le début de Nterm et la fin du Cterm (flèche 2), induit un rayon de giration proche de 24 Å tandis que pour les 3 autres mutants le rayon de giration avoisine les 30 Å. Wmut se comporte très souvent comme une pelote de laine en entremêlant facilement sa chaîne avec elle-même. La figure 6.7 nous montre l'immense variété de conformations que peut prendre la protéine. La quantification de chaque état par l'intermédiaire des

outils précédemment utilisés dans ce chapitre permet un tri de la grande densité d'informations contenue dans notre statistique, l'étude au cas par cas de chaque conformation étant bien évidemment impossible.

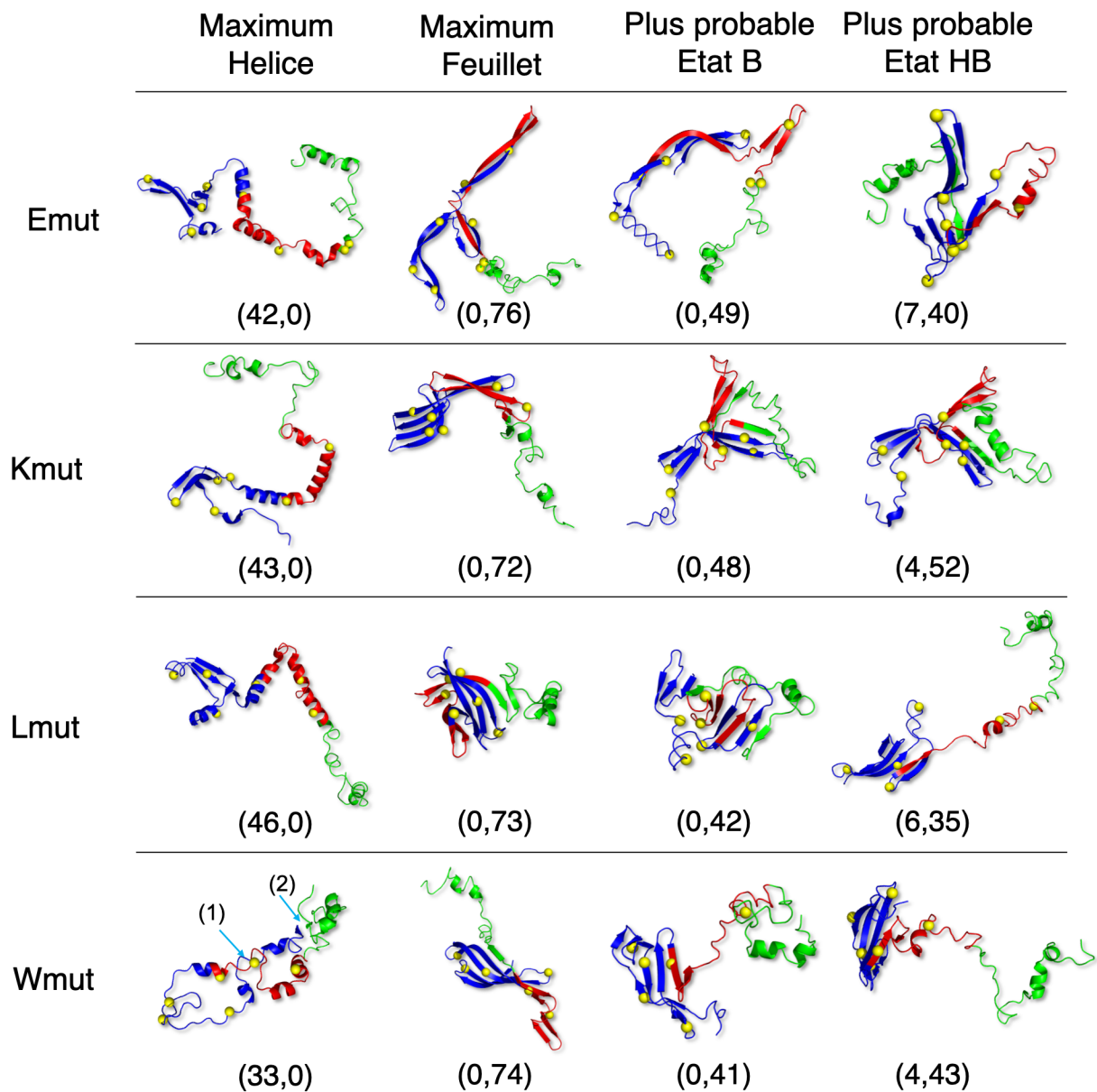


FIGURE 6.7 – Représentation en trois dimensions d'une conformation retrouvée dans 4 classes différentes pour les 4 protéines étudiées : conformation comportant le maximum de résidus en hélice (1ère colonne), maximum en feuillet  $\beta$  (2ème colonne), le plus probable dans l'état B (3ème colonne) et le plus probable pour l'état HB (4ème colonne). Les structures sont identifiées par les coordonnées des cartes de la [figure 6.3](#) (nombre de résidu en hélice, nombre de résidus en feuillet  $\beta$  ). Les sphères jaunes représentent la position des différentes mutations.

## 6.3 Etude des dimères

### Régions d'agrégation pour la formation de dimères

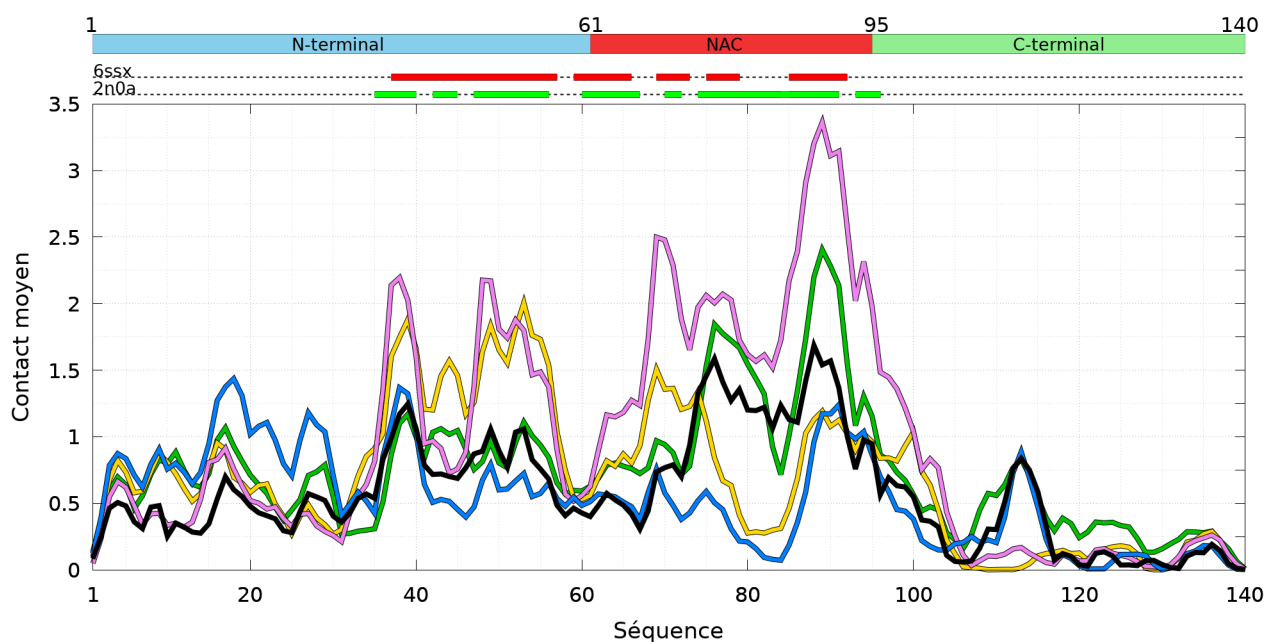


FIGURE 6.8 – Contact moyen inter-chaînes (chaîne A et B confondues) pour chaque résidu pour un rayon de coupure de 6 Å pour les WT (noir), Kmut (vert), Lmut (jaune), Wmut (violet) et Emut (bleu). Les 3 régions de la protéine sont mises en évidence par des rectangles de couleurs au dessus du graphique (bleu : Nterm, rouge : NAC, vert : Cterm). Les rectangles de couleur sur les barres pointillées donnent la position des feuillets  $\beta$  inter-chaînes des structures expérimentales de code pdb 6ssx (en rouge) et 2n0a (en vert) calculées par CUTABI.

L'impact des mutations répétées dans le motif peut également avoir un effet sur la formation des dimères. La première étude du nombre moyen de contact illustrée à la [figure 6.8](#) montre une modification de l'agrégation apportée par les différentes mutations du motif KTKGEV au sein de la protéine, y compris pour Emut dont la toxicité est semblable à celle du WT.

Le mutant Emut comporte 3 mutations sur les résidus 12, 24 et 34 qui vont avoir une tendance à faire croître la probabilité d'agrégation de la région 15-35 avec notamment A18 qui possède un contact moyen de 1.43. Les autres mutants possèdent également des modifications plus modérées de l'agrégation comparé au WT dans cette région.

Le premier pic majeur des courbes se synchronise avec l'apparition des feuillets  $\beta$  de 6ssx. Les résidus L38 et Y39 présentent des valeurs très proches, le maximum n'est pas le même pour ces deux résidus pour toutes les protéines : dans l'ordre décroissant, L38(Wmut)=2.18, Y39(Lmut)=1.88, L38(Emut)=1.36, Y39(Kmut)=1.17 et Y39(WT)=1.24. Sur la suite de la séquence Emut et Kmut vont suivre le comportement du WT jusqu'au résidu V70 où la courbe va commencer à décroître pour Emut avec un minimum atteint pour G84 (0.07) proche de la mutation E82 (0.09) tandis que Kmut continuera de suivre la courbe du WT avant de former un contact moyen supérieur à celui du WT pour la région 88-91 ( $>2$ ). Emut formera également ce pic de manière plus proche de celle du WT. Lmut et Wmut vont quant à eux avoir un

comportement très différent de celui du WT. Lmut maintient un contact moyen supérieur à 1.5 pour la majorité de ces résidus entre 37 et 57. De même pour Wmut, qui à la différence de Lmut, montre une décroissance du contact moyen entre les résidus 41 et 46. La probabilité de contact chute pour ces deux mutants dans la région de l'hélice rejoignant le WT, Emut et Kmut avant de remonter très fortement pour les résidus 68-70 avec A69 à 1.5 pour Lmut et 2.5 pour Wmut. La probabilité de contact de Lmut diminue ensuite comme Emut avec un minimum pour K80(0.23) proche de la mutation T81L. Lmut va voir sa courbe augmenter de manière similaire à celle de Emut pour le trio d'alanines 89-91. Wmut n'a pas cette décroissance dans cette région de la séquence et au contraire atteint son pic maximal pour A89 avec un contact moyen de 3.36 doublant ainsi la valeur du WT.

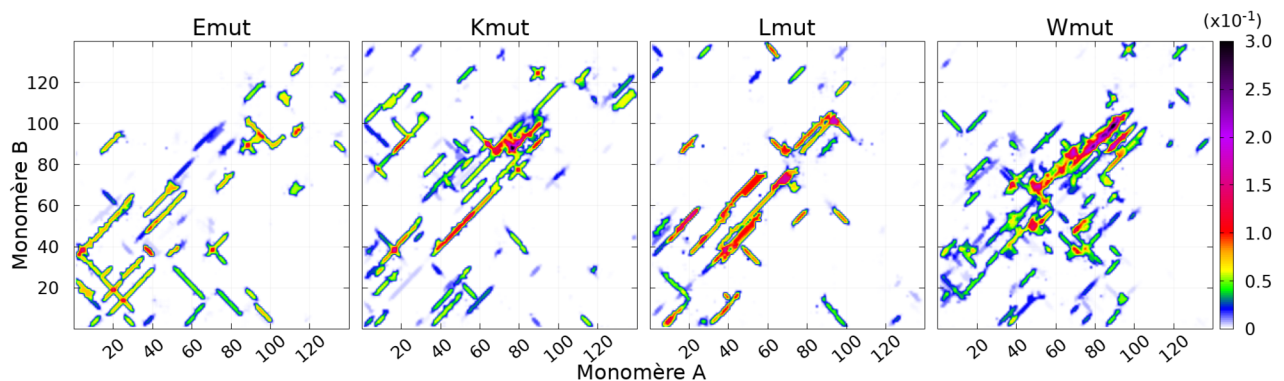


FIGURE 6.9 – Carte de contact inter-chaînes pour les dimères des mutants Emut, Kmut, Lmut et Wmut.

Le caractère intrinsèquement désordonnée de la protéine  $\alpha$ -syn peut amener à penser que certaines caractéristiques proches entre deux protéines étudiées selon une seule coordonnée, telle que la courbe donnée [figure 6.8](#), induisent des comportements similaires. L'agrégation n'est cependant pas un phénomène homogène. Un résidu d'une protéine peut posséder un contact moyen identique à celui d'une autre protéine sans pour autant réaliser ces contacts avec les mêmes résidus. La [figure 6.9](#) illustre ce propos. En effet, malgré des comportements assez similaires entre Lmut et Wmut dans la région 38-60 dans la [figure 6.8](#), les cartes de contact inter-chaînes sont en réalité très différentes avec une formation plutôt rectiligne de traces majoritairement centrées sur la diagonale dans cette région pour Lmut tandis que les contacts sont beaucoup plus hétérogènes et ponctuels chez Wmut interagissant avec toutes les régions excepté le Cterm. Wmut a une tendance globale à former des traces moins rectilignes que les autres mutants et possède une carte beaucoup plus complexe avec énormément de possibilités de contacts différents. L'agrégation du Nterm mise en évidence dans Emut entre 15-35 se fera de plusieurs manières avec d'autres régions du Nterm et le NAC. Les cartes permettent également de voir des traces rectilignes sur leur diagonale, très visibles chez Kmut et Lmut, laissant penser à la formation de contact de type fibrille (CTF). L'étude de ces cas particuliers est réalisée dans la partie suivante. On comprend donc ici l'intérêt du couplage de la [figure 6.8](#) donnant une image globale de l'agrégation avec la [figure 6.9](#) donnant le détail des régions impliquées mais obligeant une analyse au cas par cas.

## Conformations particulières

### - Les dimères type fibrille

De nombreuses études corrélient les maladies neurodégénératives à l'apparition des fibrilles d' $\alpha$ -syn dans le cerveau. Ces fibrilles ont la particularité de former des contacts reliant deux résidus identiques de chaque chaîne sur une grande partie de la protéine, ces contacts pour rappel sont appelés contacts type fibrille (CTF). Nous avons donc cherché dans les dimères ceux qui possèdent des caractéristiques similaires, ces dimères seront appelés dimères type fibrille (DTF). Un dimère sera considéré comme tel s'il possède plus de 5 CTF consécutifs. La proportion des classes des DTF est donnée [figure 6.10](#). On remarque que tous les mutants très toxiques forment plus de DTF que le WT (8%) et bien plus que Emut (6%) avec environ 13% pour Lmut, 9% pour Wmut et 10% Kmut.

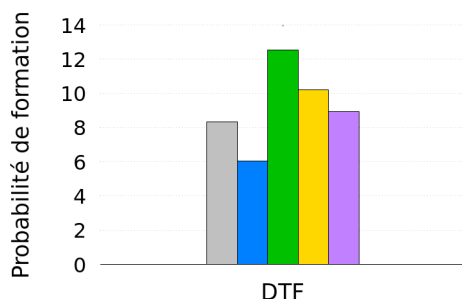


FIGURE 6.10 – Proportion de dimères possédant plus de 5 CTF consécutifs appelés dimères type fibrille (DTF). Le code couleur donne le WT en gris, Emut en bleu, Kmut en vert, Lmut en jaune et Wmut en violet.

On peut s'intéresser ensuite à la position des CTF sur la séquence et au comportement du reste des deux chaînes lorsque ce type de contact intervient. La [figure 6.11](#) donne plusieurs propriétés en fonction de la séquence des DTF : leur formation d'hélice, leur formation de feuillet  $\beta$  intra-chaîne et inter-chaînes et également de feuillet  $\beta$  inter-chaînes type fibrille (FTF) impliquant des CTF. Pour rappel (publication soumise [chapitre 4](#)), le WT favorise majoritairement des CTF dans la région 80-93, les feuillets  $\beta$  inter-chaînes sont concentrés dans la région de formations des feuillets  $\beta$  semblables à 6ssx excepté au niveau de l'hélice centrale qui est présente dans toutes les conformations des DTF. Les feuillets  $\beta$  intra-chaîne du WT ne se forment que dans le Nterm jusqu'au début de l'hélice.

Dans la faible proportion de DTF présents dans l'ensemble des dimères de Emut, les FTF ne se forment que dans le Nterm avant le résidu 34, c'est à dire hors de la région retrouvée dans les fibrilles expérimentaux commençant au résidu 38. On remarquera que Emut forme un grand nombre de feuillet  $\beta$  intra-chaîne et notamment la  $H_{41-42}$  facilement discernable par la probabilité proche de 0 pour ces deux résidus entourés de deux pics avoisinant la probabilité de 1 dans la [figure 6.11](#). Pour Kmut, bien que la proportion de feuillets  $\beta$  intra-chaîne soit élevée et répartie sur la totalité du Nterm et du NAC, ce mutant forme ces CTF majoritairement dans la région 37-65 coïncidant avec les régions de 6ssx. Lmut de manière diamétralement opposée

à Kmut va avoir la capacité de former un grand nombre d'hélices et notamment dans la région proche de sa mutation T81L. La formation de FTF se concentre donc la région (27)33-41 et 52-54. Ce type de mutant ne forme que très peu de feuillet  $\beta$  intra-chaine dans ce cas et ceux-ci se concentrent en quasi majorité avant le résidu 35. Le dernier mutant toxique Wmut affiche deux régions pour la formation de FTF : une première coïncidant avec la présence de sa mutation au résidu 48 jusqu'au résidu 55. Il est remarquable que les deux pics de FTF se forment de la même façon pour des feuillet  $\beta$  intra-chaine impliquant possiblement un double repliement à la fois inter-chaines et intra-chaine. La seconde région identifiée débute après la mutation sur le résidu 87 avec une probabilité de 1. Les FTF s'étendent jusqu'au résidu 100 mais reste plus probable jusqu'au résidu 91. Cette succession de FTF correspond à l'enchaînement de la triple alanine. D'autres types de feuillet  $\beta$  inter-chaines se repèrent dans les régions 60-65 et 70-85, régions fortement agrégées chez Wmut (voir [figure 6.8](#)).

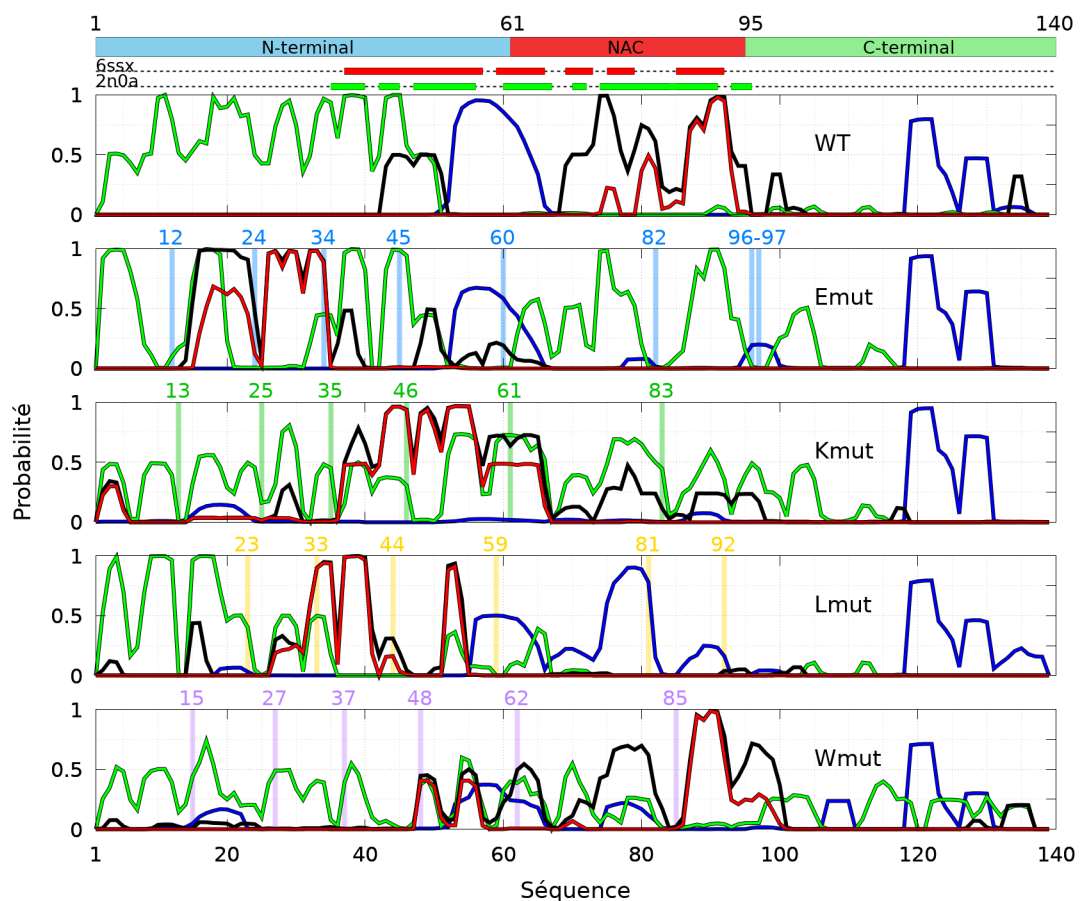


FIGURE 6.11 – Structures secondaires des conformations des DTF. Chaque courbe présente pour chaque protéine les feuillets  $\beta$  intra-chaine (chaîne A et B mélangés) en vert, les hélices (chaîne A et B mélangées) en bleu, les feuillets  $\beta$  inter-chaines en noir et les feuillets  $\beta$  type fibrille (FTF) en rouge. Les positions des différentes mutations sont données par une barre de couleur pour chaque mutant. Les rectangles de couleur sur les barres pointillées donnent la position des feuillets  $\beta$  inter-chaines des structures expérimentales de code pdb 6ssx (en rouge) et 2n0a (en vert).



La [figure 6.11](#) nous a permis d'identifier les caractéristiques globales le long de la séquence de la classe des DTF pour chaque type de protéine. Nous n'avons cependant pas l'information concernant la taille des FTF. Dans une fibrille, on retrouve une continuité de CTF du résidu 37 au résidu 95 entre les différents monomères. La courbure de certaines régions ne permet pas d'affecter ces contacts obligatoirement à des feuillets  $\beta$  (voir les structures de 2n0a et 6ssx présentées [figure 1.1](#) page [9](#)). On va donc par la suite s'intéresser à la notion de CTF et non à la notion de FTF bien que les deux propriétés soient fortement liées dans la plupart des cas.

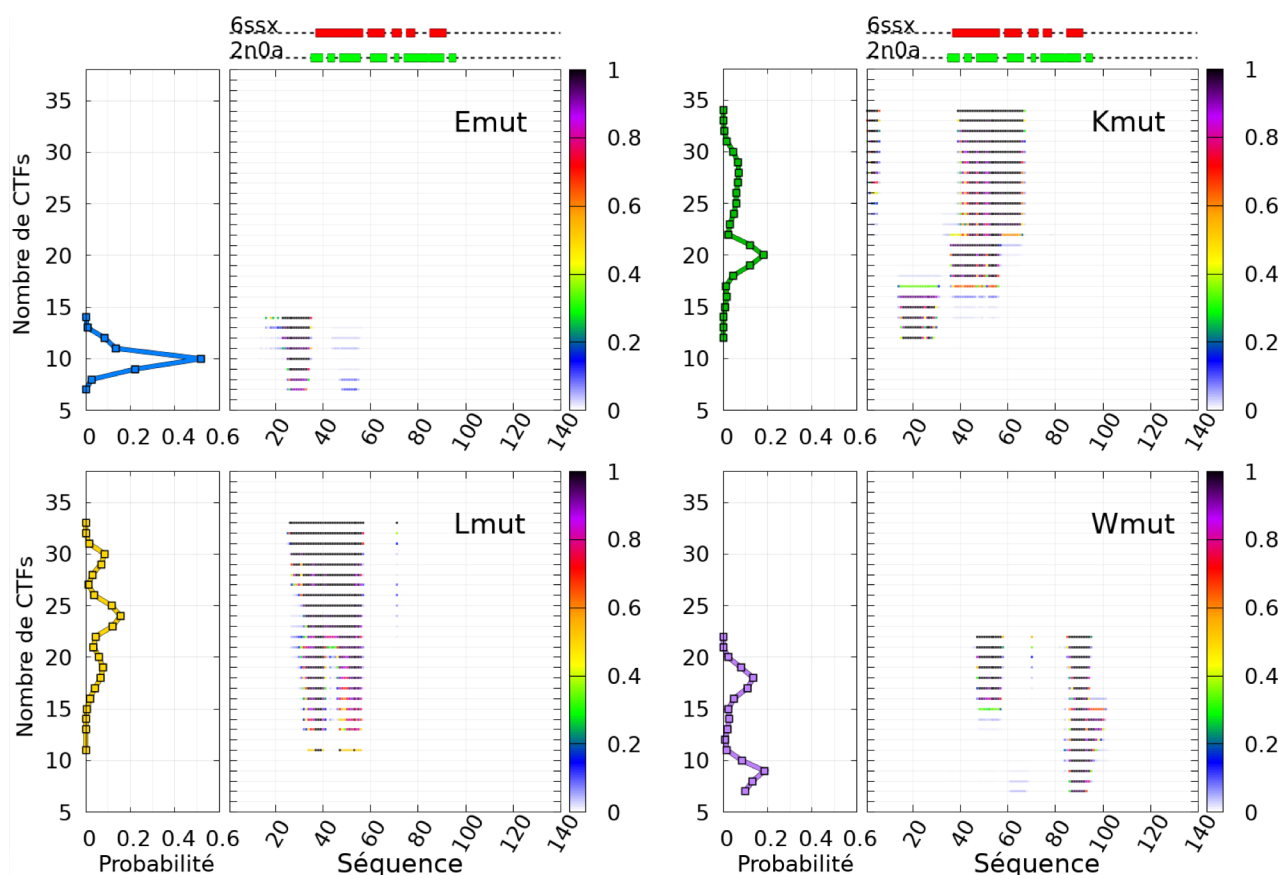


FIGURE 6.12 – Courbe de gauche : Densité de probabilité du nombre de CTF dans les conformations de la classe des DTF. Un CTF inclut une paire de résidus alignés des chaînes A et B. Courbe de droite : CTF le long de la séquence formés par les conformations présentant le nombre de CTF dans la courbe de gauche. La variation de couleur représente la probabilité de formation de ce contact au sein de la classe étudiée.

La [figure 6.12](#) permet d'avoir l'information sur la quantité de conformations comportant un certain nombre de CTF (courbe de gauche) en fonction de la position le long de la séquence (courbe de droite). Le nombre de CTF est très variable d'une protéine à l'autre, Emut ne forme que de très petits feuillets  $\beta$  de maximum 13 contacts. Le nombre de CTF le plus probable ( $p=0.3$ ) est de 9 résidus. On peut identifier, pour un nombre de CTF égale à 12, que certaines conformations sont parvenues à faire des CTF dans la région 40-60. Au vu des probabilités, cet événement reste très rare. Les contacts vont se localiser pour ce mutant sur le résidu 35 avant de s'étendre en direction du Nterm pour les conformations ayant plus de CTF. La représentation 3D de l'un de ces dimères donnée à la [figure 6.13](#) permet d'identifier une formation assez

élevée de feuillets  $\beta$  dans le Nterm induisant une distance assez grande entre les NAC des deux monomères.

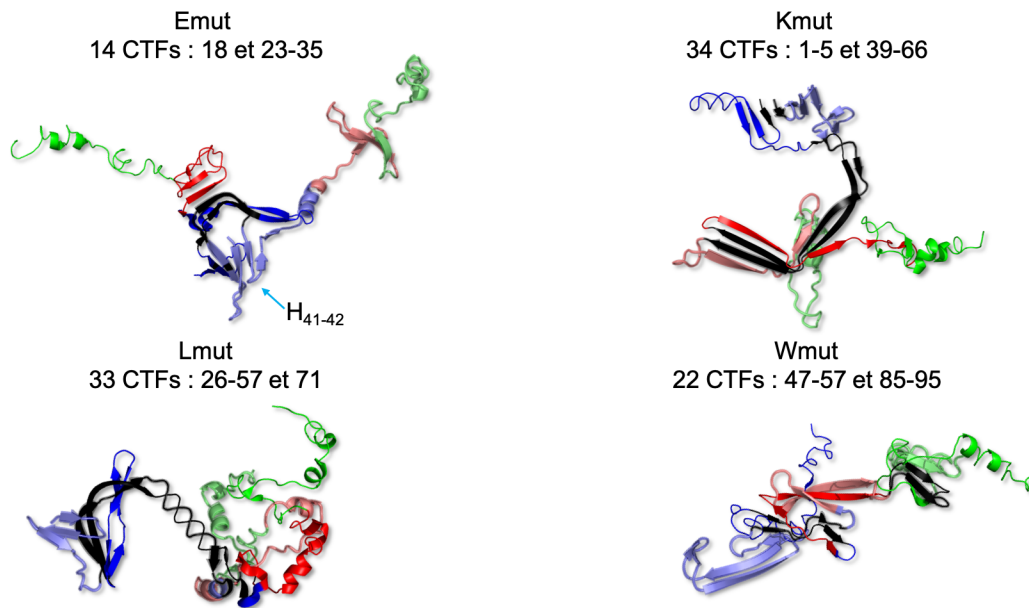


FIGURE 6.13 – Représentations 3D de DTF. Les trois régions Nterm, NAC et Cterm sont représentées par le code couleur bleu, rouge et vert, respectivement. La région noire correspond à la région possédant les CTF, et les résidus impliqués sont données au dessus de chaque représentation. Les deux chaînes du dimère se différencient par des couleurs plus claires pour l'une des deux.

Kmut est la protéine ayant des DTF avec le plus grand nombre de CTF avec un maximum de 34 (seulement 2 conformations) et un minimum de 13. On dénombre deux maxima à 20 et 28, dont les probabilités sont proches de 0.18 et 0.07, respectivement. La région 14-30 de la séquence n'était que très peu discernable sur la [figure 6.11](#) et n'apparaît que pour un faible nombre de contacts. Ce cas reste rare et instable puisqu'il n'apparaît pas dans les conformations avec un nombre de contacts plus élevé. Le pic à 20 (courbe de gauche) possède ses contacts entre les résidus 35 et 57, le second pic à 28 CTF ne possède plus les résidus 35,36 et 37. Les CTF se forment des résidus 38 à 57. Enfin pour le maximum de CTF, les 34 CTF se forment dans les régions 1-5 et 39-66. La représentation 3D de ce type de dimère est donné [figure 6.13](#).

Lmut est capable de faire des DTF ayant des régions en contact de différentes tailles allant de 11 contacts à 33 contacts. Les conformations contenant entre 11 et 20 contacts possèdent un "trou" allant des résidus 42 à 53 visible sur la [figure 6.11](#). Pour les structures avec un nombre de CTF plus élevé, les résidus 42-53 vont venir former des CTF accompagnés d'une extension jusqu'au résidu 26. On voit également apparaître quelques contacts dans la région 70. Lmut va former une sorte de tresse inter-chaînes dans la région 42-53 (voir [figure 6.13](#)) qui ne sera donc jamais détectable comme un feuillet  $\beta$  dans la [figure 6.11](#).

Enfin Wmut forme de 6 à 22 CTF avec deux maxima 9(0.09) et 18(0.14). Le premier maximum correspond à des contacts du résidu 84 au résidu 94. Pour les conformations avec le plus de contacts, la région 47-57 forme aussi des CTF. On remarque également des contacts qui

s'étendent dans une partie du Cterm jusqu'au résidu 102, mais qui semblent être peu probables et peu stables puisque les conformations possédant un plus grand nombre de CTF ne possèdent pas ce type de contacts.

Une remarque générale peut être faite à partir des cartes données [figure 6.12](#), la formation de CTF semble se stopper très souvent aux alentours des résidus 56-59. Ce résultat est aussi observé sur la [figure 6.8](#) avec une décroissance de toutes les courbes vers le résidu 60. Cette région correspond d'après la [figure 6.2](#) à une région favorable à la formation d'hélice.

### - Les dimères très agrégés

Les DTF ne représentent qu'une minorité des conformations trouvées pour les dimères de ces mutants. Comme le montre la [figure 6.10](#), les proportions de DTF sont de l'ordre de quelques pourcents à une dizaine de pourcents pour les mutants très toxiques. Une seconde méthode de classification a donc été établie afin de chercher au sein de l'ensemble des structures simulées si d'autres types de conformations pouvaient être présentes chez les mutants très toxiques et pas pour Emut. Au vu des précédents résultats, les notions de contact et d'agrégation semblent être des critères intéressants pour l'étude des dimères. Dans les études précédentes, nous ne nous sommes intéressés qu'aux contacts inter-chaines, cependant les monomères dans les dimères reste capable de former des contacts intra-chaine, comme on peut le voir sur les DTF de toutes les protéines à la [figure 6.11](#) où les probabilités de formation de feuillets  $\beta$  intra-chaine sont parfois très élevées pour certains résidus. Des cartes croisant à la fois les nombres de contacts inter-chaines et intra-chaine ont donc été réalisées pour l'ensemble des dimères et sont données [figure 6.14](#). La construction de ces cartes est expliquée dans le [chapitre 2 section 2.3 figure 2.3](#).

Le mutant Emut ne possède que des puits avec des nombres de contacts inter-chaines assez faibles, majoritairement inférieurs à 100 contacts. Le nombre de contacts intra-chaine est entre 250 et 350 pour l'ensemble des dimères ([figure 6.14](#) (A)). Les états de Emut sont donc très regroupés et proches statistiquement. Ce comportement se rapproche de celui du WT. L'état le plus probable est le couple de valeurs 81/268 (nombre de contacts inter-chaines/nombre de intra-chaine) marqué (E1) assez proche de celui retrouvé pour le WT (91/295). 3 états sont ensuite très rapprochés statistiquement avec les minimas 37/314, 55/314 et 67/327 (E2) dont les différences d'énergie libre sont de 0.34, 0.29 et 0.31 respectivement. Deux états bien moins probables forment un plus grand nombre de contacts inter-chaines 101/299(0.65) et 115/264(1.25) (E3). La classe des DTF ne possède qu'un seul état proche d'un des 3 états précédemment cités avec pour couple 70/306 ([figure 6.14](#) (B)).

Les mutants très toxiques ont pour ces critères d'étude des différences très marquées avec le mutant peu toxique par la formation de conformations possédant plus de contacts inter-chaines et de ce fait des contacts intra-chaine bien moins présents. En effet, une grande proportion des conformations de Kmut se situe entre 100 et 200 contacts inter-chaines avec notamment son état le plus probable à 114-256(0.77) (K1) ([figure 6.14](#) (A)). On définit aussi trois états

dissociés par le nombre de contacts intra-chaine (K2) proches statistiquement : 71/283 (0.84) , 80/350(0.79) et 79/320(0.88). L'état situé au milieu de ce trio correspond au couple de valeurs retrouvées chez les DTF, le second état natif identifiable dans le panneau du dessous se situe dans l'état K1 (figure [figure 6.14](#) (B)). Un autre état K3 43/307 (0.97) est très proche de l'état E2 37/314 retrouvé chez Emut. Enfin on identifie un état avec un nombre maximum de contact inter-chaines K4 168/232 (1.43).

Lmut possède un état (L1) le plus fortement probable 67/300 (0.0), trois états notés (L2) 100/282 (0.21) , 133/249 (1.29) et 35/347 (0.16) ([figure 6.14](#) (A)). Ce dernier état présentant une grosse proportion des conformations de Lmut ayant très peu de contacts inter-chaines. Un état L3 très détaché du reste des autres états est formé d'un très grand nombre de contacts inter-chaines. Son couple est 250/148 (1.34). La classe de DTFs se situent dans les puits 100/282 et 133/249 qui sont les deux classes qui possèdent le plus de contacts inter-chaines sans compter l'état L3.

Wmut a seulement un état principal présent nettement sous la barre des 100 contacts inter-chaines : 65/344 (0.93) (W2). Le reste des états majoritaires possèdent plus de 100 contacts inter-chaines et se répartissent de manière très étendue. La variété des états ne dépend principalement que des contacts inter-chaines. On identifie facilement une proportionnalité où la décroissance des contacts intra-chaine laisse place à une croissance des contacts inter-chaines. L'état le plus probable W1 est le couple 161/259 (0.67) accompagné de deux sous états (W3) 96/309 (0.68) et 139/254 (1.08). De la même façon que Lmut deux états W4 se distinguent par leur très grand nombre de contact inter-chaines 215/204 (1.75) et 256/172(1.59) ([figure 6.14](#) (A)). Les DTFs de Wmut se positionnent au niveau de l'unique groupe sous la barre des 100 contacts inter-chaines (W2) et également dans le cluster le plus probable (W1). On retrouve également des traces de DTFs pour un nombre de contacts inter-chaines supérieur à 200 ([figure 6.14](#) (B)).

Une remarque générale peut être faite sur la classe des DTF : les trois mutants considérés comme toxiques possèdent deux états bien dissociés contrairement à Emut qui n'en possède qu'un (figure [figure 6.14](#) (B)).

Les états avec un maximum de contacts inter-chaines, très atypiques, notamment pour Lmut et Wmut ont été étudiés plus en détail dans la [figure 6.15](#). Cette figure donne les CM d'une classe réalisée autour des minima de ces puits à  $\pm 3$  contacts, associé à une représentation 3D correspondant à une des conformations appartenant à cette classe. On remarque la formation d'une multitude de contacts consécutifs parallèles à la diagonale chez les mutants très toxiques, bien moindre chez Emut. La majorité de ces contacts s'apparentent à des feuillets  $\beta$  identifiables sur les représentations 3D. Les régions d'interaction entre les deux monomères s'étendent de la région 35-40 à 95 pour Wmut et Lmut, 50-100 pour Kmut et 38-59 pour Emut. Les mutants très toxiques forment donc des feuillets  $\beta$  dans la totalité de la zone retrouvée dans 6ssx contrairement à Emut. Le décalage par rapport à un CTF s'étend à plus de 20 résidus pour Kmut avec une interaction du résidu 60 du monomère A avec le résidu 80 du monomère B. Pour Lmut, Emut et Wmut le décalage est inférieur à 10 résidus. Ces conformations proches d'un DTF dans

l'allure générale sont facilement identifiables sur les cartes de contact inter-chaînes (figure 6.9) et semblent être beaucoup plus présentes chez les mutants très toxiques. Elles pourraient donc jouer un rôle dans la toxicité de ces mutants.

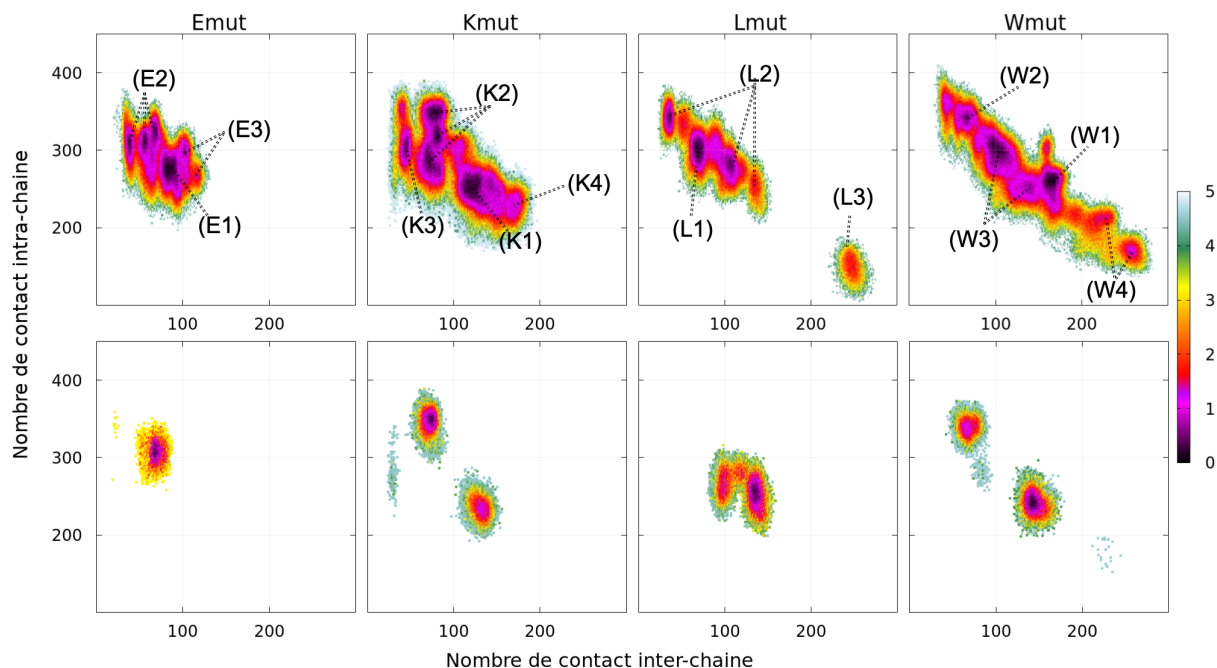


FIGURE 6.14 – Cartes 2D de  $-\ln\left(\frac{P(n_{inter}, n_{intra})}{P_{max}}\right)$  calculées à partir de la densité de probabilité  $P$  du couple du nombre de contact inter-chaînes ( $n_{inter}$ ) et intra-chaîne ( $n_{intra}$ ) entre les résidus de l'ensemble des dimères des mutants d' $\alpha$ -syn . Dans chaque carte,  $P_{max}$  est la valeur maximal de la probabilité de la carte

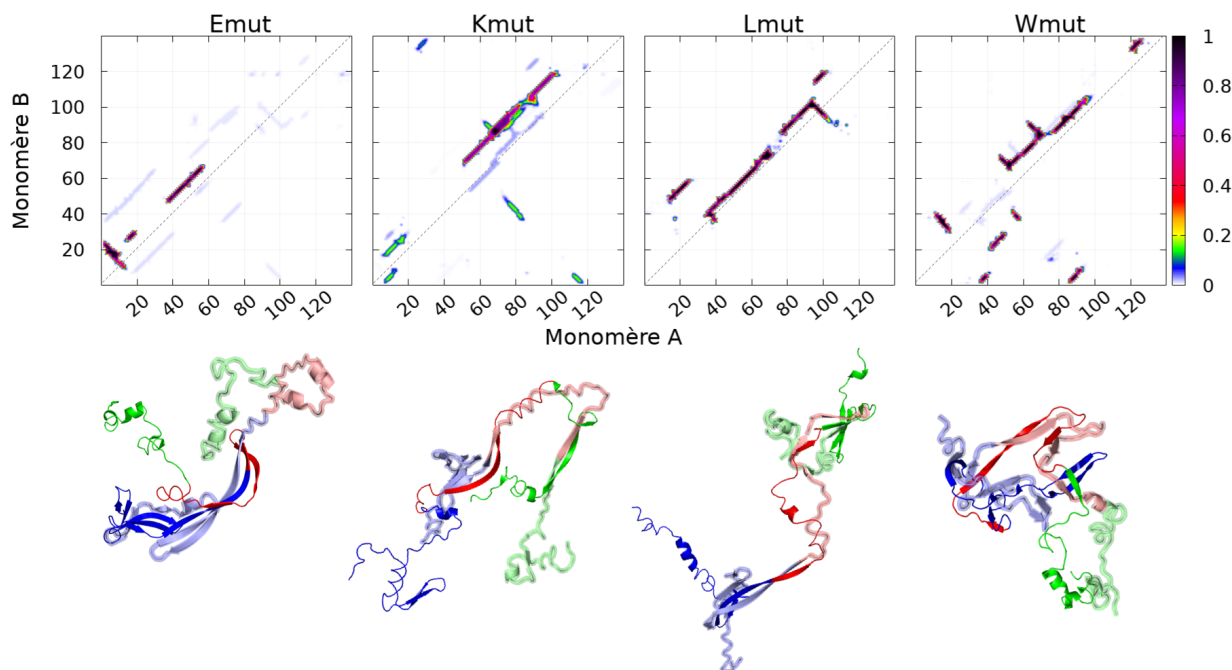


FIGURE 6.15 – Cartes de contacts des conformations appartenant aux états possédant le maximum de contacts inter-chaînes sur la figure 6.14. Une structure 3D prise aléatoirement dans la classe est donnée sous chaque carte.

## 6.4 Conclusion : Mutations répétées sur le motif KTKEGV

L'étude des mutants considérés comme très toxiques, issus de la publication donnée à la ref [\[102\]](#), a montré que leurs mutations modifient la formation des structures secondaires à l'échelle monomérique et influent considérablement sur les propriétés des dimères. La formation des hélices est notamment fortement diminuée dans la région du résidu 58, notamment pour Wmut et Kmut comparés au WT et Emut. La formation des tétramères riches en hélices pourrait donc être plus complexe pour ces deux mutants.

Pour les monomères, les mutations répétées ne déstabilisent que peu les feuillets  $\beta$  (comme A53T et E46K) mais plus fortement la formation d'hélices (comme A30P et A53T). Les deux phases retrouvées dans l'analyse des résidus en feuillets  $\beta$  et en hélices sont également retrouvées chez tous les mutants. La présence des  $H_{41-42}$  et  $H_{35-36}$  est bien plus grande pour Emut, considéré comme peu toxique, et plus faibles pour les mutants très toxiques. La structure des dimères est très fortement impactée par les mutations des mutants très toxiques. Deux types de dimères ont pu être mis en évidence : i) les DTF plus abondants pour les mutants très toxiques que pour Emut et le WT avec des régions d'agrégation différentes, ii) des dimères très fortement agrégés avec des contacts sous forme de feuillets  $\beta$  parallèles.



# Chapitre 7

## Discussion

A l'aide de la dynamique moléculaire à gros grains, nous avons étudié le comportement de l' $\alpha$ -syn sauvage (WT), des mutants liés à la maladie de Parkinson (A30P, A53T et E46K) et des mutants artificiels étudiés *in vito* et *in vivo* (Emut, Kmut, Lmut et Wmut)[\[102\]](#). L'étude des mutants a pour objectif d'identifier leurs différences avec le WT et ainsi de déterminer de possibles caractéristiques de la protéine pouvant favoriser sa toxicité et ainsi l'apparition de la maladie de Parkinson. Ces caractéristiques recherchées sont par exemple des résidus spécifiques ou un groupement de résidus jouant un rôle important dans la structure de la protéine.

Les résultats principaux de notre thèse seront dans un premier temps comparés aux données expérimentales disponibles puis aux simulations de dynamique moléculaire de l' $\alpha$ -syn réalisés par d'autres auteurs. Enfin nous expliquerons les causes des différences structurelles entre le WT et les mutants, notamment chez les dimères, afin de pouvoir émettre des hypothèses sur les propriétés favorisant l'agrégation et par hypothèse sa toxicité.

### 7.1 Hétérogénéité des structures de l' $\alpha$ -synucléine

La caractéristique majeure de l' $\alpha$ -syn est sa capacité à changer constamment de forme ce qui rend cette protéine très complexe à étudier. Reproduire l'ensemble de l'espace conformationnel de l' $\alpha$ -syn au travers de simulations de dynamique moléculaire est donc très difficile. Par ailleurs, tous les différents champs de force utilisés pour tenter de reproduire la dynamique de l' $\alpha$ -syn possèdent des biais.

En comparant la formation d'hélices et de feuillets  $\beta$  pour le WT d' $\alpha$ -syn simulé avec différents champs de force ([\[75\]](#)[\[68\]](#)[\[74\]](#)[\[73\]](#)) ainsi qu'avec UNRES (cette thèse), on s'aperçoit facilement qu'il existe des écarts marqués entre les différents résultats ([chapitre 3](#) figure 10 panneau du haut). Comparé à notre étude, certaines simulations, notamment les réfs [\[75\]](#)[\[68\]](#)[\[73\]](#) surestiment la formation d'hélices contrairement à celles de la réf [\[74\]](#) qui ont plutôt tendance à la sous estimer. On retrouve cependant des maxima de probabilité des hélices similaires entre certaines études, notamment au voisinage du résidu 60 dans les travaux des réfs [\[75\]](#) et [\[73\]](#). Nos résultats présentent le plus grand accord pour la formation des hélices avec la réf [\[75\]](#), excepté pour les résidus autour de A78, bien plus favorables à la formation d'hélice dans notre



cas. Expérimentalement, les régions 20-23 et 48-90 ont été identifiées comme ayant une forte tendance à former des hélices dans la forme tétramérique hélicoïdale de l' $\alpha$ -syn [103] ainsi que pour le monomère en interaction avec une membrane [61], incluant donc la région du résidu A78.

Pour les feuillets  $\beta$ , les différences entre les différentes simulations sont bien plus complexes à identifier (chapitre 3 figure 10 panneau du bas). Les travaux des refs [75] et [74] montrent une tendance à la formation de feuillets  $\beta$  sur l'entièreté de la séquence, tandis que ceux des refs [54] et [73], prédisent des régions en feuillets  $\beta$  plus localisées avec des probabilités bien plus faibles.

Comparer les probabilités de formation des structures secondaires de la séquence est difficilement réalisable expérimentalement. Un paramètre intéressant plus global est la fraction moyenne de résidus en hélice et en feuillet  $\beta$  au sein des conformations du WT en solution. Des données de dichroïsme circulaire (DC) donnent environ  $2\pm 3\%$  et  $11\pm 7\%$  pour les fractions en hélice et en feuillet- $\beta$  respectivement [104]. D'autres données, extraites par la même méthode, donnent dans une première étude  $< 2\%$  pour les hélices et  $30\%$  pour les feuillets  $\beta$  [105], et dans une seconde  $3\pm 1\%$  pour les hélices et  $23\pm 8\%$  pour les feuillets  $\beta$  [106]. Des simulations de dynamique moléculaire prédisent dans une étude  $2\%$  en hélices et  $11\%$  en feuillet  $\beta$  [75] et dans une autre,  $20\pm 4\%$  et  $26.8\pm 6.8\%$ , respectivement [74]. Dans notre cas, nous obtenons  $10\%$  de résidus en hélice et  $31\%$  en feuillet  $\beta$ . Ce ratio de 3 entre les deux types de structures secondaires se rapproche des résultats de [104]. On notera que dans notre étude, l'hélice la plus présente se situe dans le Cterm. Si on ignore le Cterm, la fraction en hélice tombe à  $5\%$  et se rapproche ainsi des autres études [104] [106] [75]. La nature des feuillets  $\beta$  est également un critère de comparaison. En effet, les fibrilles formées par l' $\alpha$ -syn et caractérisées par FTIR présentent majoritairement des feuillets  $\beta$  parallèles, tandis que les oligomères sont majoritairement constitués de feuillets  $\beta$  anti-parallèles [47]. On retrouve en effet dans notre ensemble plus de feuillets  $\beta$  parallèles dans nos DTFs et plus de feuillets  $\beta$  anti-parallèles dans les autres dimères désordonnés. A partir des données FTIR, la proportion de feuillets  $\beta$  est estimée à  $35\pm 5\%$  dans des oligomères constitués de 10 à 40 molécules et de  $65\pm 10\%$  dans les fibrilles. Bien que notre ensemble conformationnel d' $\alpha$ -syn ne se résume qu'à des dimères et que l'échelle de temps de nos simulations est beaucoup plus réduite que la phase de nucléation des fibrilles observée expérimentalement, on retrouve les mêmes ordres de grandeurs avec  $44.4\%$  pour les dimères désordonnés et  $57.8\%$  pour les DTFs.

Le rayon de giration ( $R_g$ ) est une métrique très utilisée expérimentalement et donne une idée globale de la taille de la protéine. Un  $R_g$  moyen de  $27.2\pm 0.44 \text{ \AA}$  est mesuré par diffusion des rayons X pour le WT en solution [107], proche de la valeur calculée dans notre ensemble pour les monomères ( $24.7 \text{ \AA}$ ). Les rayons de giration d' $\alpha$ -syn WT et des mutants est de l'ordre de  $10 \text{ \AA}$  plus grand (chapitre 4 table 3). A notre connaissance, malheureusement, il n'existe pas de données SAXS ou SANS pour les dimères. On retrouve cependant dans le cas de solutions constituées de monomères du WT et dans certaines conditions particulières, des  $R_g$  allant jusqu'à  $42.7 \text{ \AA}$  [107]. Il est intéressant de noter que non seulement les paramètres physico-chimiques

de la solution peuvent altérer le Rg de la protéine [107] mais que nos résultats indiquent que la présence de dimères en solution peut conduire à un Rg apparent des monomères sur-estimé. Nos calculs de la distribution de Rg (chapitre 4 figure 10), montrent que ce paramètre ne permet pas de classer de manière pertinente les conformations d' $\alpha$ -syn même si plusieurs sous-population de Rg sont discernables.

L'hétérogénéité de l' $\alpha$ -syn et les différences entre les mutants et le WT se traduisent également par la présence des DTFs en plus grand nombre chez A30P et E46K avec 14.04% et 15.73% respectivement au sein des dimères. A53T en forme plus faiblement avec 10.65%, suivi du WT avec 8.33%. On observe cependant seulement 23% de dimères dans notre ensemble complet de conformations pour A30P, contre 29 %,31% et 35% pour A53T, WT et E46K, respectivement. On peut imaginer ces pourcentages de la façon suivante :

- Pour 123 monomères en solution A30P forme 20 dimères désordonnés, 3 DTFs et 77 monomères libres
- Pour 121 monomères en solution WT forme 28 dimères désordonnés, 3 DTFs et 69 monomères libres
- Pour 129 monomères en solution A53T forme 26 dimères désordonnés, 3 DTFs et 71 monomères libres
- Pour 135 monomères en solution E46K forme 30 dimères désordonnés, 5-6 DTFs et 65 monomères libres

Hormis E46K, la formation de DTFs est équivalente pour ces protéines. Les différences importantes entre les DTF WT et les mutants sont discernables dans la localisation et la taille des régions de la séquence agrégées.

## 7.2 Les points d'accroche des DTFs

Nos résultats ont montré la présence de deux régions d'accroche principales entre les deux monomères, une première dans le Nterm les résidus 36-56 et une seconde dans le NAC les résidus 69-95. La probabilité de formation de contacts entre les monomères de ces régions dépend de la présence d'une mutation. A30P formera moins de contacts dans le Nterm tandis que A53T en formera moins dans le NAC. On retrouve sensiblement les mêmes régions d'accroche dans les DTFs. Ces résultats peuvent être corrélés à une étude réalisée par AFM sur des dimères en solution [55]. Un des monomères est accroché à la pointe de l'AFM et l'autre sur une surface. Les courbes force-distance ainsi réalisées permettent d'extraire une longueur de contour de la protéine. Cette donnée exprime la taille de la partie de la protéine qui n'intervient pas dans la dimérisation. Elle est donnée de l'extrémité du Cterm jusqu'au dernier résidu en contact entre les deux monomères. Dans l'étude [55], la distribution de cette donnée présente plusieurs pics pour E46K et A53T contrairement à A30P et au WT. L'existence de plusieurs points d'accroche dans nos simulations pour les DTFs de A53T (Nterm + NAC) et E46K (Nterm + prolongement

du Nterm) et d'un seul point d'accroche pour A30P et le WT (NAC) pourraient expliquer ce résultat expérimental [55]. En considérant la distance entre deux  $C^\alpha$  d'environ 3.8 Å, on peut évaluer une longueur de contour comme étant deux fois la distance entre le dernier résidu du Cterm et le premier résidu formant un CTF. Pour les DTF du WT et de A30P de notre étude, on retrouve ainsi une longueur de contour de 357 Å des résidus 94 à 140. Pour A53T et E46K, on trouve deux valeurs 380 Å (91-140) et 585 Å (64-140). Dans l'expérience, [55], les maxima de probabilité majeurs de la longueur de contour ont des valeurs de 340 Å et 440 Å, soit du même ordre de grandeur.

Le segment 38-53 présente un très grand intérêt dans l'étude de l' $\alpha$ -syn, en effet sa région de départ 36-40 est à l'origine de bon nombre d'observations expérimentales : i) le résidu L38 est le premier résidu de la boucle dans la forme hélicoïdale de l' $\alpha$ -syn liée à une membrane, ii) cette région contient les premiers résidus formant le début des feuillets  $\beta$  natifs retrouvés chez 2n0a et 6ssx (premier résidu 35 et 37 respectivement), iii) cette région est le point de départ de la formation d'une hairpin [68]. On trouve au début de cette région une des quatre tyrosines présentes dans la protéine. Cette tyrosine Y39 est l'unique présente dans le Nterm, les trois autres se situent dans le Cterm (aux résidus 125, 133 et 136). Le segment 38-53 a aussi la particularité d'être le siège des mutations connues favorisant la formation de fibrilles amyloïdes, notamment E46K et A53T [108] [109]. Seule A30P n'y appartient pas et c'est justement la seule mutation qui n'a pas d'impact sur l'oligomérisation de la protéine. [27]

La comparaison de la formation d'une boucle dans le segment 35-53 du WT et des mutants A30P, A53T et E46K a mis en évidence une bien plus grande probabilité de formation pour le WT des hairpins  $H_{41/42}$  et  $H_{35/36}$  ainsi que la formation d'un plan  $\beta$  constitué de ces deux boucles. Le mutant Emut, également peu toxique [102] réalise une proportion encore plus grande de ces structures. Tous les mutants étudiés, excepté Emut, possèdent une proportion de ces structures bien inférieure à celle du WT (voir figure 5.4 page 93 chapitre 5 et figure 6.5 page 119 chapitre 6).

Ces résultats peuvent être reliés à la formation plus favorables des DTF pour tous les mutants. La formation de CTF se fait sur des segments contenant en partie ou complètement le segment 38-53 pour A53T, E46K, Kmut, Lmut et Wmut alors que les CTF se forment uniquement sur la fin du NAC pour A30P. Le WT possède un comportement semblable à A30P pour la formation des CTF. Emut ne forme que très peu de CTF qui sont positionnés avant le segment 38-53. L'hypothèse qui résulte de ces observations est que la formation des hairpins  $H_{41/42}$  et  $H_{35/36}$  pourrait protéger la formation de CTF dans la région 38-53 et notamment au voisinage du résidu L38. La figure 7.1 illustre cette hypothèse en présentant les 2 cas de figures principaux retrouvés de manière très probable pour le WT et Emut (Structures A). Ces deux structurations possibles de ce segment le rendent très compact et complique la formation de CTF entre les résidus impliqués dans les feuillets  $\beta$ . Dans le cas de figure où deux monomères constitués de ces structures se rencontrent, la probabilité de former des CTF est très faible. L'encombrement stérique des différentes structures proposées pourrait ainsi influencer sur la probabilité de formation de CTF. Ainsi il serait plus difficile pour les monomères de Emut et du

WT de former des DTF dans cette région.

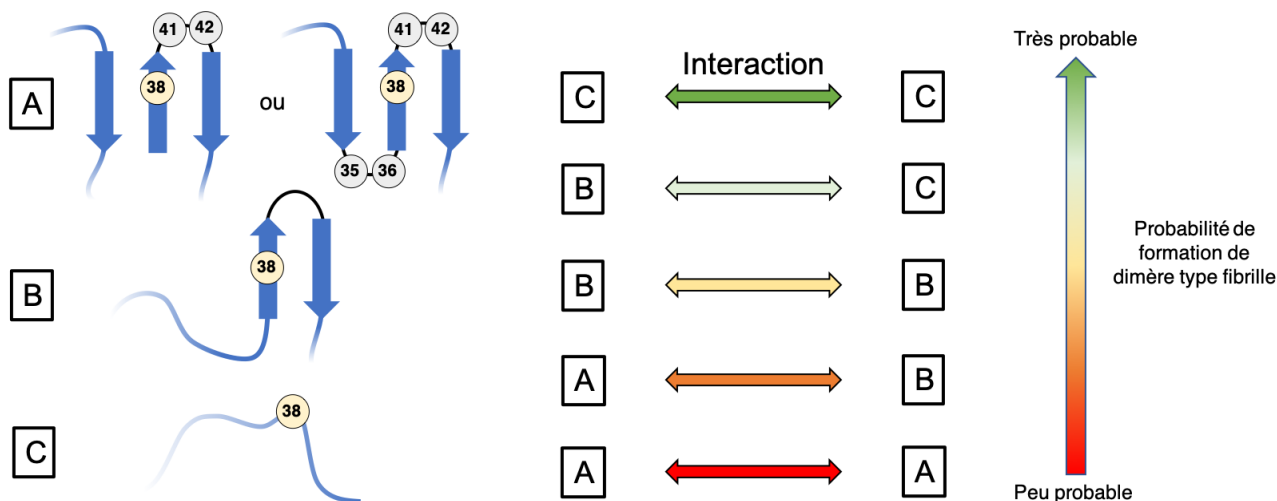


FIGURE 7.1 – Schéma des possibilités d'agrégation de la région 38-53 en fonction de la présence des hairpins  $H_{35-36}$  et  $H_{41-42}$ . L'échelle de couleur illustre la probabilité de formation de CTF entre les différentes structurations A, B et C de cette région.

Cette hypothèse est illustrée par exemple sur la structure d'Emut donnée [figure 6.13](#) page [127](#) qui montre la présence de la  $H_{41-42}$  associée à un troisième brin  $\beta$  formé par le début du Nterm. Le plan  $\beta$  ainsi formé protège l'expansion des CTF au delà du résidu 35.

Les écarts de formation de feuillets  $\beta$  sont moindres dans cette région entre les mutants et le WT probablement car les mutations n'impactent que peu la formation de feuillets  $\beta$  mais impactent grandement la structuration de ces feuillets  $\beta$  au sein de la protéine.

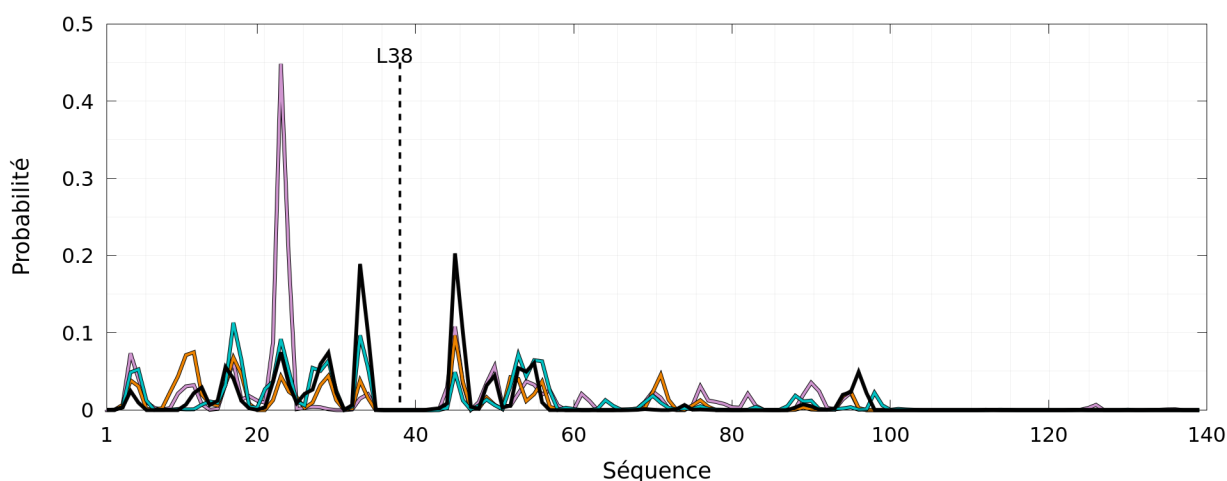


FIGURE 7.2 – Probabilité par résidu de formation de feuillets  $\beta$  avec le résidu L38 dans l'ensemble de nos simulations des monomères pour le WT (noir), A30P(rose), A53T(orange) et E46K(cyan). La position du résidu L38 est identifiée par une barre pointillée.

A30P est un cas particulier puisqu'il ne possède que peu de hairpins  $H_{41-42}$  et  $H_{35-36}$  mais est dans l'incapacité de former des CTF dans cette région. La proline forme, sur la quasi totalité de l'ensemble des conformations, une boucle induisant une forte augmentation de la

proportion de feuillets  $\beta$  dans la région (figure 8 [chapitre 3](#)). Ce phénomène totalement absent chez les autres protéines étudiées pourrait également permettre de protéger les résidus voisins du résidu 38. Une analyse complémentaire donnée [figure 7.2](#) donne la probabilité de formation d'un feuillet  $\beta$  entre le résidu 38 et le reste de la séquence. Le résidu 38 a dans le cas de A30P (courbe rose) une très forte affinité avec le résidu 23 à hauteur de près de 50%. Ce phénomène pourrait défavoriser la formation des CTF du résidu 38 avec son homologue dans un autre monomère. La courbe noire du WT montre la forte affinité du résidu 38 avec les résidus 33 et 45 dû à la présence des  $H_{35-36}$  et  $H_{41-42}$ .

Les DTF pourraient être en principe détectés en utilisant la technique FRET. En effet, nous avons comparé dans la publication soumise ([chapitre 4](#)) nos simulations aux données FRET mesurées sur des dimères et oligomères. Cette technique consiste à fixer une paire de molécules fluorescentes donneur/accepteur d'énergie sur deux entités dont on veut mesurer l'interaction, ici deux  $\alpha$ -syn identiques. Dans l'expérience à laquelle nous avons comparé nos résultats, les fluorophores sont attachés de manière covalente au résidu 90 de mutants A90 [\[110\]](#). La comparaison de nos résultats avec ces données nécessiterait de simuler les fluorophores ce qui est très complexe et au-delà des objectifs de notre travail. Une comparaison qualitative est possible en simulant l'efficacité FRET  $E$  par la formule  $E = 1/[1 + (d/R_0)^6]$  où  $R_0$  est la distance de Förster  $R_0 = 60\text{\AA}$  et  $d$  est la distance entre les  $C^\alpha$  des résidus A90 de notre simulation augmentée de  $20\text{\AA}$  pour tenir compte de la taille des fluorophores. Dans nos dimères désordonnés, on retrouve un pic à faible efficacité ( $<0.15$ ) et à haute efficacité ( $>0.75$ ). Dans les DTF, un seul pic est visible avec une efficacité  $>0.9$ . Un grand nombre de CTF induit donc une augmentation de l'efficacité FRET mesurable. Dans la ref [\[110\]](#), il a été démontré que les oligomères avec une haute efficacité FRET sont toxiques. Les dimères montraient un mélange de signaux de faible et haute efficacité difficiles à séparer. Cela suggère que les DTF pourraient correspondre par comparaison à ces données et la proximité structurale avec les fibrilles à des formes toxiques d'agrégats.

### 7.3 Elongation des CTFs dans les DTFs

L'analyse de l'ensemble des DTF trouvés dans nos simulations montrent que les DTF faibles en CTF s'agrègent plutôt dans la région du résidu 85 (observable pour A53T, A30P, WT figure 9 [chapitre 4](#) et Wmut [figure 6.12](#) [chapitre 6](#) page [126](#)). Ces DTF forment difficilement des CTF en direction du Nterm. Les DTF ayant une région de CTF la plus longue sont plutôt dans la région du résidu 38 (observable pour E46K, A53T figure 9 [chapitre 4](#) et Lmut, Kmut [figure 6.12](#) [chapitre 6](#) page [126](#)). De plus les CTF formés en dehors des régions retrouvées pour les fibrilles expérimentales (PDB ID : 6ssx et 2n0a) sont absents sur les DTF avec le plus grand nombre de CTF (observables sur E46K, A53T figure 9 [chapitre 4](#) et Emut, Kmut, Wmut ([figure 6.12](#) [chapitre 6](#))).

Un autre phénomène intéressant est la faible présence de CTF dans la région 57-70 de la séquence et l'absence totale de CTF dans la région 70-80 pour le WT et les mutants. L'analyse

de la formation de CTF dans nos DTF met en évidence que la moitié des DTF trouvés chez A53T, E46K et Kmut ne forment pas de CTF au delà du résidu 57 et que l'autre moitié ne va pas au delà du résidu 70. Pour Wmut et Lmut la totalité des DTFs ne va pas au delà du résidu 57. En allant du Nterm au Cterm, à partir du résidu 57, le nombre de contact chute fortement pour toutes les protéines et ne remontent que pour certaines à partir du résidu 70 (figure 2 [chapitre 4](#) et [figure 6.8](#) page [122](#)).

Une nouvelle fois, le comportement à l'échelle du monomère pourrait expliquer ces phénomènes. Le segment 53-68 a été observé dans des simulations de dynamique moléculaire sous forme d'hélice [\[28, 68, 111\]](#). Dans nos simulations, on retrouve également cette forte proportion d'hélices pour le WT et les mutants avec un maximum atteint pour le résidu 58. On observe également une forte proportion de feuillets  $\beta$  avec deux maxima pour les résidus 55 et 63. Les analyses effectuées ont montré que ces feuillets  $\beta$  étaient un mélange d'hairpins mais également de feuillets  $\beta$  à longue distance. De même, la région 75-82 a la possibilité de former à la fois des hélices et des feuillets  $\beta$ . Ces régions ont donc une très grande diversité de conformations possibles à l'échelle des monomères, ce qui pourrait expliquer leur difficulté à se dimériser notamment sous forme de DTF.

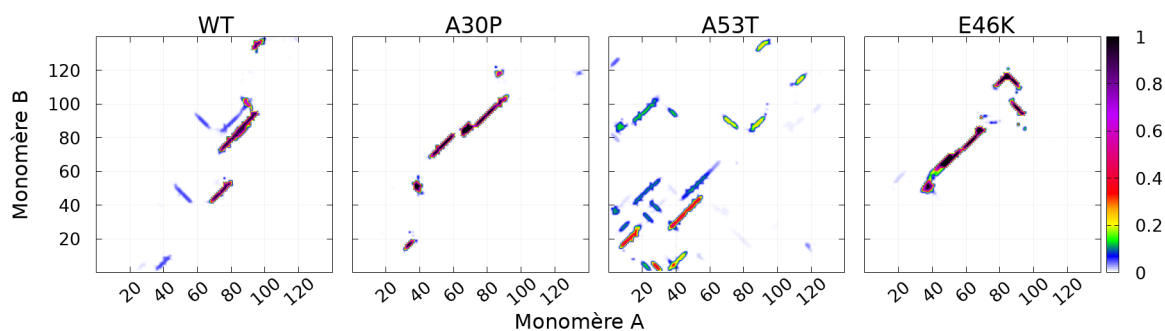


FIGURE 7.3 – Cartes de contacts des conformations appartenant aux états possédant le maximum de contacts inter-chaînes sur la figure 1 du [chapitre 4](#).

Dans les DTF possédant déjà des CTF formés des résidus 38 à 57, les présences d'hélices et de feuillets  $\beta$  intra-chaîne se maintiennent dans la majorité des protéines et peuvent expliquer la difficulté d'élongation de la région des CTF. De plus, au sein de la portion de séquence 38-95 dans le cas du WT (segment favorable à l'agrégation des fibrilles expérimentaux), la portion de séquence 57EKTKEQVTNVGG68 ne possède que 2 résidus hydrophobes V63 et V66 soit environ 12% de la portion de séquence. Ce segment diffère grandement de la portion de séquence positionnée avant, 38-56 qui possède 43% de résidus hydrophobes avec une portion de 7 résidus non hydrophobes 41GSKTKEG47 dont le nombre de contact moyen présente une légère décroissance pour la plupart des protéines (figure 2 [chapitre 4](#) et [figure 6.8](#) page [122](#).) Le segment 37-68 diffère également de la portion de séquence suivante 69-95 qui est composée de 53% de résidus hydrophobes. Toutes les mutations étudiées n'affectent que très peu la quantité de résidus hydrophobes hormis Wmut qui ajoute un résidu plus hydrophobe que la valine présente initialement dans tous les motifs KTKEGV. Dans les dimères les plus agrégés retrouvés pour les mutants Kmut, Lmut, Wmut, A30P et E46K ([figure 6.15](#) [chapitre 6](#) et

figure 7.3), le segment 57-66 d'une des chaînes sera cependant capable de former des feuillets  $\beta$  parallèles avec un segment différent de l'autre chaîne.

## 7.4 Identification de résidus clés dans l'agrégation de l' $\alpha$ -syn

De nombreuses analyses expérimentales ont permis de définir des régions jouant un rôle dans l'agrégation de la protéine  $\alpha$ -syn. Pour le WT, on retient le résidu I88 qui est le résidu qui s'agrège le plus dans les dimères. Cette observation est en accord avec le fait que la suppression de la portion de séquence A85-E94 entraîne une réduction de la polymérisation de l' $\alpha$ -syn [112]. L' $\alpha$ -syn dont les résidus V71 à V82 sont tronqués ne parvient pas à se polymériser sous forme de fibrilles [112]. Cela est en accord avec le fait que les résidus K80-V82 forment des CTF chez le WT. Les résidus A76 et V77 ont été identifiés comme n'étant pas essentiels à la polymérisation individuellement mais leur manque simultané entraîne l'absence de formation de fibrilles [112]. Dans notre étude, T75 et A76 ont en effet une probabilité élevée de formation de contacts dans les dimères et notamment sous forme de CTF.

Le NAC est souvent décrit comme étant la région principale d'agrégation, cependant notre étude montre que le Nterm ne doit pas être oublié. En effet, pour A53T et E46K c'est cette région qui forme principalement des CTF. Son rôle dans la structuration des monomères est important comme démontré dans les sections précédentes avec une forte affinité des résidus du Nterm pour la formation de feuillets  $\beta$  et notamment de hairpins. Il a récemment été découvert que les segments G36-S42 (appelé P1) et K45-E57 (appelé P2) régulent la polymérisation de l' $\alpha$ -syn [113][9]. Le retrait de P1 à un pH de 7.5 empêche l'agrégation de l' $\alpha$ -syn mais pas à un pH de 4.5. Le retrait de P1 et P2 aux deux pH différents empêche l'agrégation. Dans nos simulations, le champ de force a été calibré uniquement pour le pH de 7. Les mutations E46K et A53T ont eu un fort impact dans la formation de CTF dans cette région. De plus, dans les DTF du WT, on observe une forte proportion de feuillets  $\beta$  intra-chaîne dans la région G36-S42 précédée d'une région formant des feuillets  $\beta$  inter-chaînes 43-51 (figure 8 chapitre 4). Il est donc possible que la formation de feuillets  $\beta$  intra-chaîne soit couplée à la formation de fibrilles (ce qui conforte également l'hypothèse émise sur les hairpins). La région P2 possède des résidus en commun avec la région formant des hélices pour les monomères du WT (figure 8 chapitre 3) et les dimères (figure 6 et 8 chapitre 4). Ces observations sont également valables pour A30P. On peut donc clairement penser que ce segment favorable à la formation d'hélice joue un rôle important dans la multimérisation.

## 7.5 Les différences monomères-dimères

Il est intéressant de remarquer que les propriétés structurales des dimères du WT et des mutants sont très différentes. Toutes les analyses ont montré des écarts considérables, que ce soit dans la quantité d'agrégation, les régions d'agrégation, les différentes structures trouvées

le long de la séquence , etc. Cependant à l'échelle monomérique, les profils de probabilité de former des feuillets  $\beta$  et des hélices le long de la séquence, les densités de probabilités des couples du nombre de résidus en feuillets  $\beta$  et en hélices, etc, ont montré des différences moindres ne permettant pas de tirer des conclusions. Notamment, E46K, est très similaire au WT à l'échelle monomérique. Kmut se différencie par une légère hausse de feuillets  $\beta$  et une diminution forte d'hélice par rapport au WT. La même diminution d'hélice est observée pour Wmut qui cependant possède pour le reste une structure très proche du WT, cependant plus compacte. A53T et Lmut ont une plus grande quantité d'hélices. La seule grande différence qui relie tous ces mutants plus toxiques que le WT est la forte diminution des  $H_{41-42}$  et  $H_{35-36}$  . Enfin Emut, le mutant peu toxique, est très proche du WT dans tous les domaines à l'échelle monomérique. Avec sa mutation forte amenant une proline, A30P voit cependant sa structure proche de la mutation très affectée et par conséquent ne possède que peu de  $H_{41-42}$  et  $H_{35-36}$  . Le reste des observations sur A30P reste proche du WT. Deux choses peuvent ressortir de ces observations. Premièrement, les conformations du WT et Emut semblent être le résultat d'un équilibre délicat entre des conformations en hélices et en feuillets  $\beta$  . Une surexpression ou sous expression, même légère des hélices ou des feuillets  $\beta$  chez les mutants les plus toxiques, semble pouvoir entraîner de lourde différence dans la structuration des dimères et sur leur mode d'agrégation. Pour A30P, les conséquences de la mutation sur les dimères sont également fortes mais ce mutant conserve une grande similitude avec le WT pour la classe des DTF. Deuxièmement, les  $H_{41-42}$  et  $H_{35-36}$  , qui sont pratiquement invisibles lorsqu'on regarde les probabilités de formation de feuillets  $\beta$  par rapport à la séquence pour les monomères, ont un rôle majeur dans la dimérisation révélé par notre analyse des conformations des dimères.

## 7.6 Nouvelle structure secondaire : la tresse

Nos analyses ont mis en évidence la présence d'un nouveau type de structure secondaire que nous avons nommé tresse. Ce type de structure, mélangeant à la fois un comportement d'hélice et de feuillets  $\beta$  , est abondant pour l' $\alpha$ -syn à la fois dans sa forme sauvage et pour les mutants. Il a été démontré que cet état est un état métastable lorsque que l' $\alpha$ -syn est simulée avec d'autres champs de force. Les différentes études menées ont permis de démontrer que ce phénomène n'est probablement pas un simple artefact du champ de force UNRES. Il est très intéressant de voir qu'on retrouve ce type de structure pour  $A\beta$ , la protéine liée à la maladie d'Alzheimer, dans les simulations d'UNRES.  $A\beta_{1-42}$  forme plus de tresses que  $A\beta_{1-40}$  considérée comme moins toxique. . Il est cependant difficile de corréliser la présence de cette structure et les différentes hypothèses d'augmentation de la toxicité des protéines établies précédemment. Des analyses complémentaires en calculant le déplacement chimique de ces structures pourraient être réalisées afin de vérifier expérimentalement si on retrouve la tresse chez l' $\alpha$ -syn et  $A\beta$ .



## 7.7 Origine de la toxicité

Les raisons de la toxicité des protéines est un sujet encore très débattu. La présence abondante de fibrilles amyloïdes d' $\alpha$ -syn dans le cerveau laisse à penser que ce type de structure est la cause principale de la toxicité et du déclenchement des maladies dégénératives comme la maladie de Parkinson. Plusieurs études ont montré cependant qu'en réalité certains oligomères pourraient être encore plus toxiques que ces fibrilles y compris des oligomères de très petites tailles comme des dimères pour le WT. Au travers notre étude sur les dimères, nous avons essayé d'identifier les causes qui pourrait justifier à cette échelle une toxicité plus grandes pour certains mutants.

Le cas idéal serait de trouver un critère commun à tous les mutants les plus toxiques artificiels (Kmut, Lmut et Wmut) [58] et les mutants ponctuels qui favorisent l'apparition de la maladie de Parkinson A53T et E46K non représenté ou faiblement pour le WT, Emut et A30P et ainsi de pouvoir facilement émettre une hypothèse sur ce qui augmente la toxicité à l'échelle moléculaire.

La quantité de DTF aurait pu être la raison la plus évidente pour justifier la toxicité. Cependant nous avons vu précédemment que leur quantité au sein de notre ensemble de dimère était quasiment équivalente pour toutes les protéines de l' $\alpha$ -syn . De plus, une étude suggère que ce n'est pas la quantité de dimères qui serait le paramètre le plus pertinent mais les propriétés structurales de ces dimères pendant la phase d'auto-assemblage dans le processus de neuro-dégénérescence [42]. En effet, dans nos simulations les DTF se différencient par les points d'accroche entre les deux monomères chez le WT et les mutants. Le WT ne forme ses CTF qu'à la fin du NAC avec une taille allant jusqu'à 18 résidus. Emut n'en forme que des petits au début du Nterm dans une région hors des régions d'agrégation des fibrilles expérimentales. Les mutants A53T et Wmut sont capables d'en former de la même façon que le WT ou dans la région 35-57. Lmut, Kmut et E46K en forment uniquement de très longs dans la région 38-57(70). A30P en forme dans la même région que le WT mais en quantité moindre. Les mutants considérés comme les plus toxiques, par leur formation plus favorables d'inclusions, ont donc des DTF plus longs faisant intervenir le Nterm. Cette caractéristique peut donc potentiellement accélérer le processus de formation des fibrilles qui sera bien plus lent pour le WT puisque ses dimères ne parviennent que difficilement à former des grandes quantités de CTF. Le processus sera encore plus long pour A30P qui forme encore moins de CTF que le WT en accord avec le ralentissement de la formation des fibrilles causé par cette mutation [27]. Cette hypothèse est donc probablement la plus robuste puisqu'elle relie plusieurs observations expérimentales [55] [42].

Une seconde observation importante est la grande population de dimères très agrégés qui se trouvent chez les mutants les plus toxiques Ces dimères sont constitués de feuillets  $\beta$  parallèles dont l'interaction se fait entre des résidus différents de chaque chaîne ce qui les différencie des DTF. Ces structures font intervenir la région 57-90 d'un des monomères très facilement alors que cette région ne forme que très rarement des CTF dans toutes les protéines étudiées. L'autre monomère fera intervenir une des régions plus hydrophobes ce qui facilite l'agrégation. Le WT ne forme quasiment pas ces dimères et son état le plus agrégé est en réalité l'état des

DTF. Emut en forme des moins agrégés. A l'inverse, A30P forme ces dimères très agrégés. Seul un mutant considéré comme favorisant la maladie de Parkinson fait exception, A53T, qui ne forme que des dimères peu agrégés. Cette observation ne doit pas être négligée car ces dimères pourraient également jouer un rôle important dans l'équilibre de la protéine.



# Chapitre 8

## Conclusion et perspectives

Trouver des traitements pour les maladies neurodégénératives est un des plus grands challenges du 21<sup>ème</sup> siècle. La première étape de ce défi est de comprendre les causes de ces maladies. La maladie de Parkinson ressort comme étant liée à la protéine  $\alpha$ -syn et son comportement dans le cerveau humain. Le caractère intrinsèquement désordonné de l' $\alpha$ -syn implique une multitude de conformations possibles et complique grandement son étude expérimentale. La puissance informatique et les avancées scientifiques dans la description des interactions entre les atomes permettent aujourd'hui d'étudier des protéines de manière numérique et d'obtenir une grande quantité d'informations sur leurs caractéristiques structurales. La méthode de dynamique moléculaire à gros grains UNRES utilisée dans cette thèse a permis notamment de réaliser des simulations à des échelles de temps effectives importantes pour se rapprocher au mieux des temps pertinents pour l'agrégation de l' $\alpha$ -syn . Elle a permis également d'étudier cette protéine sous forme de dimère, cette performance n'a que très rarement été réalisée dans la littérature scientifique. Plusieurs millions de conformations induisant plusieurs centaines de gigas de données ont ainsi pu être traités et analysés pour apporter de nouvelles informations sur la protéine  $\alpha$ -syn . Ainsi nous avons pu contribuer à répondre aux différentes problématiques autour de l' $\alpha$ -syn données dans l'introduction de cette thèse :

Q1 : Quelles sont les différentes conformations du monomère d' $\alpha$ -syn ?

Nous avons identifié dans une premier temps deux phases dans les monomères. Une première phase constituée de conformations contenant uniquement des feuillets  $\beta$  et une seconde phase contenant un mélange d'hélices et de feuillets  $\beta$  . Les hélices moins nombreuses se retrouvent dans toutes les régions de la protéine et de différentes tailles. Les feuillets  $\beta$  peuvent se former à longue distance entre les différentes régions de la protéine comme sous forme de hairpin avec notamment les  $H_{41-42}$  et  $H_{35-36}$  identifiées comme étant bien plus présentes chez le WT que chez les mutants plus toxiques de l' $\alpha$ -syn . L'impact des mutations est facilement identifiable pour A30P avec une augmentation forte de la formation de feuillets  $\beta$  dans la région de la mutation tandis que la mutation A53T fait croître la formation d'hélices. E46K reste quant à lui proche du comportement du WT.

Q2 : Quelles structures monomériques de l' $\alpha$ -syn seraient favorables à l'agrégation ?

La formation des  $H_{41-42}$  et  $H_{35-36}$  dans le Nterm semblent être un frein dans la formation de CTF. La structuration des monomères au moment de leur interaction influencerait sur l'arrangement des dimères et pourrait donc être importante pour former des fibrilles. La forte hydrophobicité du NAC en fait une région majeure de contact dans la dimérisation d' $\alpha$ -syn.

L'impact des hélices et d'autres types de feuillets  $\beta$  à longue distance est difficile à établir. Le rôle du Cterm semble être également complexe à évaluer puisqu'il ne forme que très peu de feuillets  $\beta$  mais très souvent de petites hélices. Peu de contacts se forme dans cette région entre les monomères dans les dimères.

Q3 : Quelles sont les régions favorables à la formation de dimères de type fibrille chez l' $\alpha$ -syn ?

L'étude des CTF a permis de répondre à cette question. Ces contacts formés entre deux résidus de même indice entre les deux monomères sont la signature de la formation des fibrilles chez  $\alpha$ -syn . Dans les dimères simulés ici, on observe la formation de tel contacts. Les DFT se forment préférentiellement dans une petite portion de la séquence à la fin du NAC pour le WT et A30P tandis que les mutations E46K et A53T favorisent la formation de DTF dans une très grande région allant du Nterm à partir du résidu 36 jusqu'au début du NAC. , A53T est également capable de faire un petit nombre de CTF dans la même région que le WT et A30P.

Q4 : Existe-t-il d'autres types de dimères chez l' $\alpha$ -syn ?

Les dimères type fibrille ne sont effectivement pas les seuls dimères présents dans notre ensemble. Ils sont même minoritaires puisque la plus grande partie des dimères sont désordonnés. On retrouve notamment un grand nombre de dimères capables de former des feuillets  $\beta$  parallèles avec des indices de résidu décalés entre les deux monomères. Ce type de dimères se retrouve en grand nombre chez A30P et E46K.

Nos analyses ont confirmé le caractère très désordonné de la protéine et la possibilité de formation de dimères pré-fibrillaires dans l'étape d'auto-assemblage de fibrilles amyloïdes. Nos simulations restent cependant une prédiction théorique du comportement de la protéine. L'utilisation d'une méthode à gros grains induit une perte de résolution sur la structure de la protéine par les différents facteurs moyennés lors du calcul. Par comparaison avec d'autres études de dynamique moléculaire, le champ de force utilisé UNRES semble surexprimer les feuillets  $\beta$  . Le point de départ de nos simulations utilisé peut également être un biais, en effet nous avons fait le choix de mettre face à face les mêmes régions de la protéine pour les deux monomères. Ce choix était naturel au vu des différents résultats expérimentaux et notamment l'allure des fibrilles amyloïdes qui font interagir les mêmes régions des monomères entre elles. Nos simulations ont donc probablement plus facilement favorisé la formation de contacts de type fibrille. Des calculs partant avec deux monomères en sens opposé ont été réalisés et montrent que la formation de CTF est plus difficile à réaliser dans ce cas. La structure des deux monomères va

devoir subir de nombreux réarrangements pour trouver la conformation favorable à la formation de CTF. Un temps de calcul bien plus long serait donc nécessaire pour observer la formation de DTF en utilisant ces conditions initiales défavorables. Néanmoins, il serait intéressant dans le futur, bien que coûteux en temps de calcul, de poursuivre de telles simulations et d'examiner en détail les dimères désordonnés qui en résulteraient.

Cette thèse et les simulations réalisées ont pour but de donner des pistes de recherche aux expérimentateurs sur la protéine  $\alpha$ -syn. De nombreuses analyses complémentaires peuvent être réalisées pour notamment comprendre l'importance du C-terminal très discuté dans la littérature scientifique. L'impact de cette région n'est pas ressorti dans nos résultats. Différentes trajectoires d'une  $\alpha$ -syn<sub>1-95</sub> ne possédant pas de C-terminal pourraient permettre de comprendre son rôle. Plusieurs simulations supprimant d'autres portions de la protéine pourraient également mettre en avant leur importance notamment sur des portions possédant le motif KTKEGV. Le rôle des hélices peut être également approfondi, le motif 58KTKEQV63 présente un acide glutamique très favorable à la formation d'hélice contrairement à la glycine plus communément retrouvée à cette position dans le motif KTKEGV. Ce résidu est très défavorable à la formation d'hélice. La mutation Q62G pourrait donc considérablement déstabiliser l'hélice et ainsi mettre en avant son importance. Enfin l' $\alpha$ -syn ne se limite pas à la formation de dimères, grâce aux modèles à gros grains il serait possible de réaliser des analyses de trimères jusqu'au tétramère. Il pourrait être intéressant d'étudier la croissance d'une fibrille en partant de la structure expérimentale de code pdb 2n0a et ainsi d'étudier les résidus clés pour la fixation. La simulation prenant comme point de départ une structure de DTF de notre ensemble pourrait également être intéressante afin de voir si la fixation d'un troisième monomère viendrait favoriser la formation de CTF comme pour les structures expérimentales.



# Bibliographie

- [1] Adrien Guzzo, Patrice Delarue, Ana Rojas, Adrien Nicolai, Gia G. Maisuradze, and Patrick Senet. Missense mutations modify the conformational ensemble of the alpha-synuclein monomer which exhibits a two-phase characteristic. *Frontiers in Molecular Biosciences*, 8, nov 2021.
- [2] Claudio Soto. Unfolding the role of protein misfolding in neurodegenerative diseases. *Nature Reviews Neuroscience*, 4(1) :49–60, January 2003. Number : 1 Publisher : Nature Publishing Group.
- [3] Fabrizio Chiti and Christopher M. Dobson. Protein Misfolding, Functional Amyloid, and Human Disease. *Annual Review of Biochemistry*, 75(1) :333–366, June 2006. Publisher : Annual Reviews.
- [4] Leonidas Stefanis. alpha-Synuclein in Parkinson’s disease. *Cold Spring Harbor Perspectives in Medicine*, 2(2) :a009399, February 2012.
- [5] Fabrizio Chiti and Christopher M. Dobson. Protein Misfolding, Amyloid Formation, and Human Disease : A Summary of Progress Over the Last Decade. *Annual Review of Biochemistry*, 86 :27–68, June 2017.
- [6] Benedict Tanudjojo, Samiha S. Shaikh, Alexis Fenyi, Luc Bousset, Devika Agarwal, Jade Marsh, Christos Zois, Sabrina Heman-Ackah, Roman Fischer, David Sims, Ronald Melki, and George K. Tofaris. Phenotypic manifestation of alpha-synuclein strains derived from Parkinson’s disease and multiple system atrophy in human dopaminergic neurons. *Nature Communications*, 12(1) :3817, December 2021.
- [7] Phuong H. Nguyen, Ayyalusamy Ramamoorthy, Bikash R. Sahoo, Jie Zheng, Peter Faller, John E. Straub, Laura Dominguez, Joan-Emma Shea, Nikolay V. Dokholyan, Alfonso De Simone, Buyong Ma, Ruth Nussinov, Saeed Najafi, Son Tung Ngo, Antoine Loquet, Mara Chiricotto, Pritam Ganguly, James McCarty, Mai Suan Li, Carol Hall, Yiming Wang, Yifat Miller, Simone Melchionna, Birgit Habenstein, Stepan Timr, Jiaying Chen, Brianna Hnath, Birgit Strodel, Rakez Kayed, Sylvain Lesné, Guanghong Wei, Fabio Sterpone, Andrew J. Doig, and Philippe Derreumaux. Amyloid Oligomers : A Joint Experimental/Computational Perspective on Alzheimer’s Disease, Parkinson’s Disease, Type II Diabetes, and Amyotrophic Lateral Sclerosis. *Chemical Reviews*, 121(4) :2545–2647, February 2021. Publisher : American Chemical Society.
- [8] N. Cremades, S. W. Chen, and C. M. Dobson. Chapter Three - Structural Characteristics of alpha-Synuclein Oligomers. In Massimo Sandal, editor, *International Review of*



- Cell and Molecular Biology*, volume 329 of *Early Stage Protein Misfolding and Amyloid Aggregation*, pages 79–143. Academic Press, January 2017.
- [9] Matthew G. Iadanza, Matthew P. Jackson, Eric W. Hewitt, Neil A. Ranson, and Sheena E. Radford. A new era for understanding amyloid structures and disease. *Nature Reviews Molecular Cell Biology*, 19(12) :755–773, December 2018.
- [10] Peter E Wright and H. Jane Dyson. Intrinsically unstructured proteins : re-assessing the protein structure-function paradigm. *Journal of Molecular Biology*, 293(2) :321–331, October 1999.
- [11] Vladimir N. Uversky. Intrinsically Disordered Proteins and Their “Mysterious” (Meta)Physics. *Frontiers in Physics*, 7, 2019. Publisher : Frontiers.
- [12] Antonio Deiana, Sergio Forcelloni, Alessandro Porrello, and Andrea Giansanti. Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell. *PLOS ONE*, 14(8) :e0217889, August 2019. Publisher : Public Library of Science.
- [13] R. Jakes, M. G. Spillantini, and M. Goedert. Identification of two distinct synucleins from human brain. *FEBS letters*, 345(1) :27–32, May 1994.
- [14] Maria Grazia Spillantini, Marie Luise Schmidt, Virginia M.-Y. Lee, John Q. Trojanowski, Ross Jakes, and Michel Goedert. alpha-Synuclein in Lewy bodies. *Nature*, 388(6645) :839–840, August 1997. Number : 6645 Publisher : Nature Publishing Group.
- [15] Leonid Breydo, Jessica W. Wu, and Vladimir N. Uversky. alpha-synuclein misfolding and Parkinson’s disease. *Biochimica Et Biophysica Acta*, 1822(2) :261–285, February 2012.
- [16] Hilal A. Lashuel. Do Lewy bodies contain alpha-synuclein fibrils ? and Does it matter ? A brief history and critical analysis of recent reports. *Neurobiology of Disease*, 141 :104876, July 2020.
- [17] Victoria A. Trinkaus, Irene Riera-Tur, Antonio Martínez-Sánchez, Felix J. B. Bäuerlein, Qiang Guo, Thomas Arzberger, Wolfgang Baumeister, Irina Dudanova, Mark S. Hipp, F. Ulrich Hartl, and Rubén Fernández-Busnadiego. In situ architecture of neuronal alpha-Synuclein inclusions. *Nature Communications*, 12(1) :2110, April 2021. Number : 1 Publisher : Nature Publishing Group.
- [18] Mihael H. Polymeropoulos, Christian Lavedan, Elisabeth Leroy, Susan E. Ide, Anindya Dehejia, Amalia Dutra, Brian Pike, Holly Root, Jeffrey Rubenstein, Rebecca Boyer, Edward S. Stenroos, Settara Chandrasekharappa, Aglaia Athanassiadou, Theodore Papapetropoulos, William G. Johnson, Alice M. Lazzarini, Roger C. Duvoisin, Giuseppe Di Iorio, Lawrence I. Golbe, and Robert L. Nussbaum. Mutation in the alpha-Synuclein Gene Identified in Families with Parkinson’s Disease. *Science*, 276(5321) :2045–2047, June 1997. Publisher : American Association for the Advancement of Science Section : Report.
- [19] Rejko Krüger, Wilfried Kuhn, Thomas Müller, Dirk Voitalla, Manuel Graeber, Sigfried Kösel, Horst Przuntek, Jörg T. Epplen, Ludger Schols, and Olaf Riess. AlaSOPro mu-

- tation in the gene encoding alpha-synuclein in Parkinson's disease. *Nature Genetics*, 18(2) :106–108, February 1998. Number : 2 Publisher : Nature Publishing Group.
- [20] Juan J. Zarranz, Javier Alegre, Juan C. Gómez-Esteban, Elena Lezcano, Raquel Ros, Israel Ampuero, Lídice Vidal, Janet Hoenicka, Olga Rodriguez, Begoña Atarés, Verónica Llorens, Estrella Gomez Tortosa, Teodoro del Ser, David G. Muñoz, and Justo G. de Yebenes. The new mutation, E46K, of alpha-synuclein causes parkinson and Lewy body dementia. *Annals of Neurology*, 55(2) :164–173, 2004. \_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1002/ana.10795>.
- [21] Julia Fuchs, Ales Tichopad, Yulia Golub, Marita Munz, Katherine J. Schweitzer, Björn Wolf, Daniela Berg, Jakob C. Mueller, and Thomas Gasser. Genetic variability in the SNCA gene influences alpha-synuclein levels in the blood and brain. *The FASEB Journal*, 22(5) :1327–1334, 2008. \_eprint : <https://faseb.onlinelibrary.wiley.com/doi/pdf/10.1096/fj.07-9348com>.
- [22] Silke Appel-Cresswell, Carles Vilarino-Guell, Mary Encarnacion, Holly Sherman, Irene Yu, Brinda Shah, David Weir, Christina Thompson, Chelsea Szu-Tu, Joanne Trinh, Jan O. Aasly, Alex Rajput, Ali H. Rajput, A. Jon Stoessl, and Matthew J. Farrer. Alpha-synuclein p.H50Q, a novel pathogenic mutation for Parkinson's disease. *Movement Disorders*, 28(6) :811–813, 2013. \_eprint : <https://movementdisorders.onlinelibrary.wiley.com/doi/pdf/10.1002/mds.25421>.
- [23] Petra Pasanen, Liisa Myllykangas, Maija Siitonen, Anna Raunio, Seppo Kaakkola, Jukka Lyytinen, Pentti J. Tienari, Minna Pöyhönen, and Anders Paetau. A novel alpha-synuclein mutation A53E associated with atypical multiple system atrophy and Parkinson's disease-type pathology. *Neurobiology of Aging*, 35(9) :2180.e1–2180.e5, September 2014.
- [24] Simona Petrucci, Monia Ginevrino, and Enza Maria Valente. Phenotypic spectrum of alpha-synuclein mutations : New insights from patients and cellular models. *Parkinsonism & Related Disorders*, 22 :S16–S20, January 2016. Publisher : Elsevier.
- [25] Diana F. Lázaro, Eva F. Rodrigues, Ramona Langohr, Hedieh Shahpasandzadeh, Thales Ribeiro, Patrícia Guerreiro, Ellen Gerhardt, Katharina Kröhnert, Jochen Klucken, Marcos D. Pereira, Blagovesta Popova, Niels Kruse, Brit Mollenhauer, Silvio O. Rizzoli, Gerhard H. Braus, Karin M. Danzer, and Tiago F. Outeiro. Systematic Comparison of the Effects of Alpha-synuclein Mutations on Its Oligomerization and Aggregation. *PLOS Genetics*, 10(11) :e1004741, November 2014. Publisher : Public Library of Science.
- [26] Michele Perni, Annemieke van der Goot, Ryan Limbocker, Tjakko J. van Ham, Francesco A. Aprile, Catherine K. Xu, Patrick Flagmeier, Karen Thijssen, Pietro Sormanni, Giuliana Fusco, Serene W. Chen, Pavan K. Challa, Julius B. Kirkegaard, Romain F. Laine, Kai Yu Ma, Martin B. D. Müller, Tessa Sinnige, Janet R. Kumita, Samuel I. A. Cohen, Renée Seinstra, Gabriele S. Kaminski Schierle, Clemens F. Kaminski, Denise Barbut, Alfonso De Simone, Tuomas P. J. Knowles, Michael Zasloff, Ellen A. A. Nollen, Michele

- Vendruscolo, and Christopher M. Dobson. Comparative studies in the a30p and a53t-synuclein c. elegans strains to investigate the molecular origins of parkinson's disease. *Frontiers in Cell and Developmental Biology*, 9, mar 2021.
- [27] K. A. Conway, S. J. Lee, J. C. Rochet, T. T. Ding, R. E. Williamson, and P. T. Lansbury. Acceleration of oligomerization, not fibrillization, is a shared property of both alpha-synuclein mutations linked to early-onset Parkinson's disease : implications for pathogenesis and therapy. *Proceedings of the National Academy of Sciences of the United States of America*, 97(2) :571–576, January 2000.
- [28] Orly Ullman, Charles K. Fisher, and Collin M. Stultz. Explaining the Structural Plasticity of alpha-Synuclein. *Journal of the American Chemical Society*, 133(48) :19536–19546, December 2011. Publisher : American Chemical Society.
- [29] Danielle E. Mor, Scott E. Ugras, Malcolm J. Daniels, and Harry Ischiropoulos. Dynamic structural flexibility of alpha-synuclein. *Neurobiology of Disease*, 88 :66–74, April 2016.
- [30] Tim Bartels, Joanna G. Choi, and Dennis J. Selkoe. alpha-Synuclein occurs physiologically as a helically folded tetramer that resists aggregation. *Nature*, 477(7362) :107–110, September 2011. Number : 7362 Publisher : Nature Publishing Group.
- [31] Wei Wang, Iva Perovic, Johnathan Chittuluru, Alice Kaganovich, Linh T. T. Nguyen, Jingling Liao, Jared R. Auclair, Derrick Johnson, Anuradha Landaru, Alana K. Simorellis, Shulin Ju, Mark R. Cookson, Francisco J. Asturias, Jeffrey N. Agar, Brian N. Webb, ChulHee Kang, Dagmar Ringe, Gregory A. Petsko, Thomas C. Pochapsky, and Quyen Q. Hoang. A soluble alpha-synuclein construct forms a dynamic tetramer. *Proceedings of the National Academy of Sciences*, 108(43) :17797–17802, October 2011. Publisher : National Academy of Sciences Section : Biological Sciences.
- [32] Giuliana Fusco, Maximo Sanz-Hernandez, and Alfonso De Simone. Order and disorder in the physiological membrane binding of alpha-synuclein. *Current Opinion in Structural Biology*, 48 :49–57, February 2018.
- [33] Marcus D. Tuttle, Gemma Comellas, Andrew J. Nieuwkoop, Dustin J. Covell, Deborah A. Berthold, Kathryn D. Kloepper, Joseph M. Courtney, Jae K. Kim, Alexander M. Barclay, Amy Kendall, William Wan, Gerald Stubbs, Charles D. Schwieters, Virginia M. Y. Lee, Julia M. George, and Chad M. Rienstra. Solid-state NMR structure of a pathogenic fibril of full-length human alpha-synuclein. *Nature Structural & Molecular Biology*, 23(5) :409–415, May 2016. Number : 5 Publisher : Nature Publishing Group.
- [34] Ricardo Guerrero-Ferreira, Nicholas MI Taylor, Ana-Andreea Arteni, Pratibha Kumari, Daniel Mona, Philippe Ringler, Markus Britschgi, Matthias E Lauer, Ali Makky, Joeri Verasdonck, Roland Riek, Ronald Melki, Beat H Meier, Anja Böckmann, Luc Bousset, and Henning Stahlberg. Two new polymorphic structures of human full-length alpha-synuclein fibrils solved by cryo-electron microscopy. *eLife*, 8 :e48907, December 2019. Publisher : eLife Sciences Publications, Ltd.

- [35] Ricardo Guerrero-Ferreira, Lubomir Kovacic, Dongchun Ni, and Henning Stahlberg. New insights on the structure of alpha-synuclein fibrils using cryo-electron microscopy. *Current Opinion in Neurobiology*, 61 :89–95, April 2020.
- [36] Katsuya Araki, Naoto Yagi, Koki Aoyama, Chi-Jing Choong, Hideki Hayakawa, Harutoshi Fujimura, Yoshitaka Nagai, Yuji Goto, and Hideki Mochizuki. Parkinson’s disease is a type of amyloidosis featuring accumulation of amyloid fibrils of alpha-synuclein. *Proceedings of the National Academy of Sciences*, 116(36) :17963–17969, September 2019. ISBN : 9781906124113 Publisher : National Academy of Sciences Section : Biological Sciences.
- [37] Manuel Schweighauser, Yang Shi, Airi Tarutani, Fuyuki Kametani, Alexey G. Murzin, Bernardino Ghetti, Tomoyasu Matsubara, Taisuke Tomita, Takashi Ando, Kazuko Hasegawa, Shigeo Murayama, Mari Yoshida, Masato Hasegawa, Sjors H. W. Scheres, and Michel Goedert. Structures of alpha-synuclein filaments from multiple system atrophy. *Nature*, 585(7825) :464–469, September 2020. Bandiera\_abtest : a Cg\_type : Nature Research Journals Number : 7825 Primary\_atype : Research Publisher : Nature Publishing Group Subject\_term : Cryoelectron microscopy;Neurodegeneration Subject\_term\_id : cryoelectron-microscopy ;neurodegeneration.
- [38] Luc Bousset, Laura Pieri, Gemma Ruiz-Arlandis, Julia Gath, Poul Henning Jensen, Birgit Habenstein, Karine Madona, Vincent Olieric, Anja Böckmann, Beat H. Meier, and Ronald Melki. Structural and functional characterization of two alpha-synuclein strains. *Nature Communications*, 4(1) :2575, October 2013. Bandiera\_abtest : a Cg\_type : Nature Research Journals Number : 1 Primary\_atype : Research Publisher : Nature Publishing Group Subject\_term : Diseases of the nervous system;Molecular biophysics Subject\_term\_id : diseases-of-the-nervous-system ;molecular-biophysics.
- [39] Leif Antonschmidt, Rıza Dervişoğlu, Vrinda Sant, Kumar Tekwani Movellan, Ingo Mey, Dietmar Riedel, Claudia Steinem, Stefan Becker, Loren B. Andreas, and Christian Griesinger. Insights into the molecular mechanism of amyloid filament formation : Segmental folding of alpha-synuclein on lipid membranes. *Science Advances*, 7(20) :eabg2174, May 2021. Publisher : American Association for the Advancement of Science Section : Research Article.
- [40] Dhiman Ghosh, Pradeep K. Singh, Shruti Sahay, Narendra Nath Jha, Reeba S. Jacob, Shamik Sen, Ashutosh Kumar, Roland Riek, and Samir K. Maji. Structure based aggregation studies reveal the presence of helix-rich intermediate during alpha-Synuclein aggregation. *Scientific Reports*, 5(1) :9228, August 2015.
- [41] Michael J. Volles, Seung-Jae Lee, Jean-Christophe Rochet, Mark D. Shtilerman, Tomas T. Ding, Jeffrey C. Kessler, and Peter T. Lansbury. Vesicle permeabilization by protofibrillar alpha synuclein : implications for the pathogenesis and treatment of parkinson’s disease. *Biochemistry*, 40(26) :7812–7819, July 2001. Publisher : American Chemical Society.
- [42] Laura Tosatto, Mathew H. Horrocks, Alexander J. Dear, Tuomas P. J. Knowles, Mauro Dalla Serra, Nunilo Cremades, Christopher M. Dobson, and David Klenerman. Single-

- molecule FRET studies on alpha-synuclein oligomerization of Parkinson's disease genetically related mutants. *Scientific Reports*, 5(1) :16696, December 2015.
- [43] Amberley D. Stephens, Maria Zacharopoulou, Rani Moons, Giuliana Fusco, Neeleema Seetaloo, Anass Chiki, Philippa J. Woodhams, Ioanna Mela, Hilal A. Lashuel, Jonathan J. Phillips, Alfonso De Simone, Frank Sobott, and Gabriele S. Kaminski Schierle. Extent of N-terminus exposure of monomeric alpha-synuclein determines its aggregation propensity. *Nature Communications*, 11(1) :2820, December 2020.
- [44] Jie Li, Vladimir N. Uversky, and Anthony L. Fink. Conformational Behavior of Human alpha-Synuclein is Modulated by Familial Parkinson's Disease Point Mutations A30P and A53T. *NeuroToxicology*, 23(4) :553–567, October 2002.
- [45] Ghibom Bhak, Soonkoo Lee, Tae-Hwan Kim, Ji-Hye Lee, Jee Eun Yang, Keehyoung Joo, Jooyoung Lee, Kookheon Char, and Seung R. Paik. Morphological Evaluation of Metastable Oligomers of alpha-Synuclein with Small-Angle Neutron Scattering. *Scientific Reports*, 8 :14295, September 2018.
- [46] Nikolai Lorenzen, Søren Bang Nielsen, Alexander K. Buell, Jørn Døvling Kaspersen, Paolo Arosio, Brian Stougaard Vad, Wojciech Paslawski, Gunna Christiansen, Zuzana Valnickova-Hansen, Maria Andreassen, Jan J. Enghild, Jan Skov Pedersen, Christopher M. Dobson, Tuomas P. J. Knowles, and Daniel Erik Otzen. The Role of Stable alpha-Synuclein Oligomers in the Molecular Events Underlying Amyloid Formation. *Journal of the American Chemical Society*, 136(10) :3859–3868, March 2014. Publisher : American Chemical Society.
- [47] Serene W. Chen, Srdja Drakulic, Emma Deas, Myriam Ouberaï, Francesco A. Aprile, Rocío Arranz, Samuel Ness, Cintia Roodveldt, Tim Williams, Erwin J. De-Genst, David Klenerman, Nicholas W. Wood, Tuomas P. J. Knowles, Carlos Alfonso, Germán Rivas, Andrey Y. Abramov, José María Valpuesta, Christopher M. Dobson, and Nunilo Cremades. Structural characterization of toxic oligomers that are kinetically trapped during alpha-synuclein fibril formation. *Proceedings of the National Academy of Sciences*, 112(16) :E1994–E2003, April 2015. Publisher : National Academy of Sciences Section : PNAS Plus.
- [48] Lei Zhou and Dmitry Kurouski. Structural Characterization of Individual alpha-Synuclein Oligomers Formed at Different Stages of Protein Aggregation by Atomic Force Microscopy-Infrared Spectroscopy. *Analytical Chemistry*, 92(10) :6806–6810, May 2020.
- [49] Ross A. Fredenburg, Carla Rospigliosi, Robin K. Meray, Jeffrey C. Kessler, Hilal A. Lashuel, David Eliezer, and Peter T. Lansbury. The impact of the E46K mutation on the properties of alpha-synuclein in its monomeric and oligomeric states. *Biochemistry*, 46(24) :7107–7118, June 2007.
- [50] Nunilo Cremades, Samuel I. A. Cohen, Emma Deas, Andrey Y. Abramov, Allen Y. Chen, Angel Orte, Massimo Sandal, Richard W. Clarke, Paul Dunne, Francesco A. Aprile, Carlos W. Bertocini, Nicholas W. Wood, Tuomas P. J. Knowles, Christopher M. Dobson,

- and David Klenerman. Direct Observation of the Interconversion of Normal and Toxic Forms of alpha-Synuclein. *Cell*, 149(5) :1048–1059, May 2012.
- [51] Xiaobo Mao, Michael Tianhao Ou, Senthilkumar S. Karuppagounder, Tae-In Kam, Xiling Yin, Yulan Xiong, Preston Ge, George Essien Umanah, Saurav Brahmachari, Joo-Ho Shin, Ho Chul Kang, Jianmin Zhang, Jinchong Xu, Rong Chen, Hyejin Park, Shaida A. Andrabi, Sung Ung Kang, Rafaella Araújo Gonçalves, Yu Liang, Shu Zhang, Chen Qi, Sharon Lam, James A. Keiler, Joel Tyson, Donghoon Kim, Nikhil Panicker, Seung Pil Yun, Creg J. Workman, Dario A. A. Vignali, Valina L. Dawson, Han Seok Ko, and Ted M. Dawson. Pathological alpha-synuclein transmission initiated by binding lymphocyte-activation gene 3. *Science*, 353(6307) :aah3374, September 2016. Publisher : American Association for the Advancement of Science.
- [52] Sampathkumar Krishnan, Eva Y. Chi, Stephen J. Wood, Brent S. Kendrick, Cynthia Li, William Garzon-Rodriguez, Jette Wypych, Theodore W. Randolph, Linda O. Narhi, Anja Leona Biere, Martin Citron, and John F. Carpenter. Oxidative Dimer Formation Is the Critical Rate-Limiting Step for Parkinson’s Disease alpha-Synuclein Fibrillogenesis. *Biochemistry*, 42(3) :829–837, January 2003. Publisher : American Chemical Society.
- [53] Zhengjian Lv, Alexey V. Krasnoslobodtsev, Yuliang Zhang, Daniel Ysselstein, Jean-Christophe Rochet, Scott C. Blanchard, and Yuri L. Lyubchenko. Direct Detection of alpha-Synuclein Dimerization Dynamics : Single-Molecule Fluorescence Analysis. *Biophysical Journal*, 108(8) :2038–2047, April 2015. Publisher : Elsevier.
- [54] Junping Yu, Sarka Malkova, and Yuri L. Lyubchenko. alpha-Synuclein Misfolding : Single Molecule AFM Force Spectroscopy Study. *Journal of Molecular Biology*, 384(4) :992–1001, December 2008.
- [55] Alexey V. Krasnoslobodtsev, Ivan L. Volkov, Josephat M. Asiago, Jagadish Hindupur, Jean-Christophe Rochet, and Yuri L. Lyubchenko. alpha-Synuclein Misfolding Assessed with Single Molecule AFM Force Spectroscopy : Effect of Pathogenic Mutations. *Biochemistry*, 52(42) :7377–7386, October 2013. Publisher : American Chemical Society.
- [56] Yuliang Zhang, Mohtadin Hashemi, Zhengjian Lv, Benfeard Williams, Konstantin I. Popov, Nikolay V. Dokholyan, and Yuri L. Lyubchenko. High-speed atomic force microscopy reveals structural dynamics of alpha-synuclein monomers and dimers. *The Journal of Chemical Physics*, 148(12) :123322, March 2018. Publisher : American Institute of Physics.
- [57] Lucas. Navigating the dynamic landscape of alpha-synuclein morphology : a review of the physiologically relevant tetrameric conformation.
- [58] Ulf Dettmer, Andrew J. Newman, Victoria E. von Saucken, Tim Bartels, and Dennis Selkoe. KTKEGV repeat motifs are key mediators of normal alpha-synuclein tetramerization : Their mutation causes excess monomers and neurotoxicity. *Proceedings of the National Academy of Sciences of the United States of America*, 112(31) :9596–9601, August 2015.

- [59] Yoann Cote, Patrice Delarue, Harold A. Scheraga, Patrick Senet, and Gia G. Maisuradze. From a Highly Disordered to a Metastable State : Uncovering Insights of alpha-Synuclein. *ACS Chemical Neuroscience*, 9(5) :1051–1065, May 2018. Publisher : American Chemical Society.
- [60] Thomas Gurry, Orly Ullman, Charles K. Fisher, Iva Perovic, Thomas Pochapsky, and Collin M. Stultz. The Dynamic Structure of alpha-Synuclein Multimers. *Journal of the American Chemical Society*, 135(10) :3865–3872, March 2013. Publisher : American Chemical Society.
- [61] Tobias S. Ulmer, Ad Bax, Nelson B. Cole, and Robert L. Nussbaum. Structure and Dynamics of Micelle-bound Human alpha-Synuclein\*. *Journal of Biological Chemistry*, 280(10) :9595–9603, March 2005.
- [62] Chi-Yuan Cheng, Jobin Varkey, Mark R. Ambroso, Ralf Langen, and Songi Han. Hydration dynamics as an intrinsic ruler for refining protein structure at lipid membrane interfaces. *Proceedings of the National Academy of Sciences*, 110(42) :16838–16843, October 2013. Publisher : National Academy of Sciences Section : Biological Sciences.
- [63] B. I. Giasson, I. V. Murray, J. Q. Trojanowski, and V. M. Lee. A hydrophobic stretch of 12 amino acid residues in the middle of alpha-synuclein is essential for filament assembly. *The Journal of Biological Chemistry*, 276(4) :2380–2386, January 2001.
- [64] K. Ueda, H. Fukushima, E. Masliah, Y. Xia, A. Iwai, M. Yoshimoto, D. A. Otero, J. Kondo, Y. Ihara, and T. Saitoh. Molecular cloning of cDNA encoding an unrecognized component of amyloid in Alzheimer disease. *Proceedings of the National Academy of Sciences*, 90(23) :11282–11286, December 1993. Publisher : National Academy of Sciences Section : Research Article.
- [65] Ian V. J. Murray, Benoit I. Giasson, Shawn M. Quinn, Vishwanath Koppaka, Paul H. Axelsen, Harry Ischiropoulos, John Q. Trojanowski, and Virginia M.-Y. Lee. Role of alpha-synuclein carboxy-terminus on fibril formation in vitro. *Biochemistry*, 42(28) :8530–8540, July 2003.
- [66] Pratibha Kumari, Dhiman Ghosh, Agathe Vanas, Yanick Fleischmann, Thomas Wiegand, Gunnar Jeschke, Roland Riek, and Cédric Eichmann. Structural insights into alpha-synuclein monomer–fibril interactions. *Proceedings of the National Academy of Sciences*, 118(10) :e2012171118, March 2021.
- [67] Matthew M. Dedmon, Kresten Lindorff-Larsen, John Christodoulou, Michele Vendruscolo, and Christopher M. Dobson. Mapping Long-Range Interactions in alpha-Synuclein using Spin-Label NMR and Ensemble Molecular Dynamics Simulations. *Journal of the American Chemical Society*, 127(2) :476–477, January 2005. Publisher : American Chemical Society.
- [68] Hang Yu, Wei Han, Wen Ma, and Klaus Schulten. Transient beta-hairpin formation in alpha-synuclein monomer revealed by coarse-grained molecular dynamics simulation.

- The Journal of Chemical Physics*, 143(24) :243142, December 2015. Publisher : American Institute of Physics.
- [69] SeongByeong Park, Jeseong Yoon, Soonmin Jang, Kyunghee Lee, and Seokmin Shin. The role of the acidic domain of  $\alpha$ -synuclein in amyloid fibril formation : a molecular dynamics study. *Journal of Biomolecular Structure and Dynamics*, 34(2) :376–383, may 2015.
- [70] Nicholas I. Brodie, Konstantin I. Popov, Evgeniy V. Petrotchenko, Nikolay V. Dokholyan, and Christoph H. Borchers. Conformational ensemble of native  $\alpha$ -synuclein in solution as determined by short-distance crosslinking constraint-guided discrete molecular dynamics simulations. *PLOS Computational Biology*, 15(3) :e1006859, mar 2019.
- [71] Igor F. Tsigelny, Pazit Bar-On, Yuriy Sharikov, Leslie Crews, Makoto Hashimoto, Mark A. Miller, Steve H. Keller, Oleksandr Platoshyn, Jason X.-J. Yuan, and Eliezer Masliah. Dynamics of  $\alpha$ -synuclein aggregation and inhibition of pore-like oligomer development by  $\alpha$ -synuclein. *FEBS Journal*, 274(7) :1862–1877, mar 2007.
- [72] Karnesh Jain, Othman Ghribi, and Jerome Delhommelle. Folding free-energy landscape of  $\alpha$ -synuclein (35–97) via replica exchange molecular dynamics. *Journal of Chemical Information and Modeling*, 61(1) :432–443, dec 2020.
- [73] Orkid Coskuner and Olivia Wise-Scira. Structures and Free Energy Landscapes of the A53T Mutant-Type  $\alpha$ -Synuclein Protein and Impact of A53T Mutation on the Structures of the Wild-Type  $\alpha$ -Synuclein Protein with Dynamics. *ACS Chemical Neuroscience*, 4(7) :1101–1113, April 2013.
- [74] Rafael Ramis, Joaquín Ortega-Castro, Rodrigo Casasnovas, Laura Mariño, Bartolomé Vilanova, Miquel Adrover, and Juan Frau. A Coarse-Grained Molecular Dynamics Approach to the Study of the Intrinsically Disordered Protein  $\alpha$ -Synuclein. *Journal of Chemical Information and Modeling*, 59(4) :1458–1471, April 2019.
- [75] Orly Ullman, Charles K. Fisher, and Collin M. Stultz. Explaining the Structural Plasticity of  $\alpha$ -Synuclein. *Journal of the American Chemical Society*, 133(48) :19536–19546, December 2011.
- [76] Masataka Yamauchi and Hisashi Okumura. Dimerization of  $\alpha$ -Synuclein Fragments Studied by Isothermal–Isobaric Replica-Permutation Molecular Dynamics Simulation. *Journal of Chemical Information and Modeling*, 61(3) :1307–1321, March 2021. Publisher : American Chemical Society.
- [77] Kamlesh Kumar Sahu, Michael T. Woodside, and Jack A. Tuszynski.  $\alpha$ -synuclein dimer structures found from computational simulations. *Biochimie*, 116 :133–140, September 2015.
- [78] Adam Liwo, Adam K. Sieradzian, Agnieszka G. Lipska, Cezary Czaplewski, InSuk Joung, Wioletta Żmudzińska, Anna Hałabis, and Stanisław Ołdziej. A general method for the derivation of the functional forms of the effective energy terms in coarse-grained energy functions of polymers. III. Determination of scale-consistent backbone-local and correlation potentials in the UNRES force field and force-field calibration and validation. *The*



- Journal of Chemical Physics*, 150(15) :155104, April 2019. Publisher : American Institute of Physics.
- [79] Adam Liwo, Cezary Czaplewski, Jarosław Pillardy, and Harold A. Scheraga. Cumulant-based expressions for the multibody terms for the correlation between local and electrostatic interactions in the united-residue force field. *The Journal of Chemical Physics*, 115(5) :2323–2347, August 2001. Publisher : American Institute of Physics.
- [80] Gia G. Maisuradze, Patrick Senet, Cezary Czaplewski, Adam Liwo, and Harold A. Scheraga. Investigation of Protein Folding by Coarse-Grained Molecular Dynamics with the UNRES Force Field. *The Journal of Physical Chemistry A*, 114(13) :4471–4485, April 2010. Publisher : American Chemical Society.
- [81] Ana V. Rojas, Adam Liwo, and Harold A. Scheraga. Molecular dynamics with the united-residue force field : ab initio folding simulations of multichain proteins. *The Journal of Physical Chemistry B*, 111(1) :293–309, dec 2006.
- [82] Rui Zhou, Gia G. Maisuradze, David Suñol, Toni Todorovski, Maria J. Macias, Yi Xiao, Harold A. Scheraga, Cezary Czaplewski, and Adam Liwo. Folding kinetics of WW domains with the united residue force field for bridging microscopic motions and experimental measurements. *Proceedings of the National Academy of Sciences*, 111(51) :18243–18248, December 2014. ISBN : 9781420914115 Publisher : National Academy of Sciences Section : Biological Sciences.
- [83] Adam K. Sieradzan, Anatolii Korneev, Alexander Begun, Khatuna Kachlishvili, Harold A. Scheraga, Alexander Molochkov, Patrick Senet, Antti J. Niemi, and Gia G. Maisuradze. Investigation of Phosphorylation-Induced Folding of an Intrinsically Disordered Protein by Coarse-Grained Molecular Dynamics. *Journal of Chemical Theory and Computation*, 17(5) :3203–3220, May 2021. Publisher : American Chemical Society.
- [84] Ewa Gołaś, Gia G. Maisuradze, Patrick Senet, Stanisław Ołdziej, Cezary Czaplewski, Harold A. Scheraga, and Adam Liwo. Simulation of the Opening and Closing of Hsp70 Chaperones by Coarse-Grained Molecular Dynamics. *Journal of Chemical Theory and Computation*, 8(5) :1750–1764, May 2012. Publisher : American Chemical Society.
- [85] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314(1-2) :141–151, nov 1999.
- [86] Mey Khalili, Adam Liwo, Anna Jagielska, and Harold A. Scheraga. Molecular Dynamics with the United-Residue Model of Polypeptide Chains. II. Langevin and Berendsen-Bath Dynamics and Tests on Model alpha-Helical Systems. *The Journal of Physical Chemistry B*, 109(28) :13798–13810, July 2005. Publisher : American Chemical Society.
- [87] Wouter G. Touw, Coos Baakman, Jon Black, Tim A. H. te Beek, E. Krieger, Robbie P. Joosten, and Gert Vriend. A series of PDB-related databanks for everyday needs. *Nucleic Acids Research*, 43(D1) :D364–D368, January 2015.

- [88] Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure : Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12) :2577–2637, dec 1983.
- [89] M. Heinig and D. Frishman. STRIDE : a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Research*, 32(Web Server) :W500–W502, jul 2004.
- [90] Michael Feig, John Karanicolas, and Charles L. Brooks. MMTSB Tool Set : enhanced sampling and multiscale modeling methods for applications in structural biology. *Journal of Molecular Graphics and Modelling*, 22(5) :377–395, May 2004.
- [91] Piotr Rotkiewicz and Jeffrey Skolnick. Fast procedure for reconstruction of full-atom protein models from reduced representations. *Journal of Computational Chemistry*, 29(9) :1460–1465, 2008. \_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.20906>.
- [92] Naomi K. Fox, Steven E. Brenner, and John-Marc Chandonia. SCOPe : Structural Classification of Proteins—extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Research*, 42(D1) :D304–D309, January 2014.
- [93] John-Marc Chandonia, Naomi K Fox, and Steven E Brenner. SCOPe : classification of large macromolecular structures in the structural classification of proteins—extended database. *Nucleic Acids Research*, 47(D1) :D475–D481, January 2019.
- [94] G. Labesse, N. Colloc'h, J. Pothier, and J.-P. Mornon. P-SEA : a new efficient assignment of secondary structure from Calpha trace of proteins. *Bioinformatics*, 13(3) :291–295, June 1997.
- [95] Adrien Guzzo, Patrice Delarue, Ana Rojas, Adrien Nicolai, Gia G. Maisuradze, and Patrick Senet. Missense Mutations Modify the Conformational Ensemble of the alpha-Synuclein Monomer Which Exhibits a Two-Phase Characteristic. *Frontiers in Molecular Biosciences*, 8 :1104, 2021.
- [96] Jackson S. Turner, Julian Q. Zhou, Julianna Han, Aaron J. Schmitz, Amena A. Rizk, Wafaa B. Alsoussi, Tingting Lei, Mostafa Amor, Katherine M. McIntire, Philip Meade, Shirin Strohmeier, Rafael I. Brent, Sara T. Richey, Alem Haile, Yuhe R. Yang, Michael K. Klebert, Teresa Suessen, Sharlene Teefey, Rachel M. Presti, Florian Krammer, Steven H. Kleinstein, Andrew B. Ward, and Ali H. Ellebedy. Human germinal centres engage memory and naive b cells after influenza vaccination. *Nature*, 586(7827) :127–132, aug 2020.
- [97] Jing Huang, Sarah Rauscher, Grzegorz Nawrocki, Ting Ran, Michael Feig, Bert L de Groot, Helmut Grubmüller, and Alexander D MacKerell. CHARMM36m : an improved force field for folded and intrinsically disordered proteins. *Nature Methods*, 14(1) :71–73, nov 2016.
- [98] Robert B. Best and Gerhard Hummer. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *The Journal of Physical Chemistry B*, 113(26) :9004–9015, jun 2009.

- [99] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L. Klepeis, Ron O. Dror, and David E. Shaw. Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins : Structure, Function, and Bioinformatics*, 78(8) :1950–1958, mar 2010.
- [100] Tian Qiu, Qian Liu, Yong-Xiang Chen, Yu-Fen Zhao, and Yan-Mei Li. Abeta42 and abeta40 : similarities and differences. *Journal of Peptide Science*, 21(7) :522–529, may 2015.
- [101] Ana Rojas, Adam Liwo, Dana Browne, and Harold A. Scheraga. Mechanism of fiber assembly : Treatment of abeta peptide aggregation with a coarse-grained united-residue force field. *Journal of Molecular Biology*, 404(3) :537–552, dec 2010.
- [102] Ulf Dettmer, Andrew J. Newman, Victoria E. von Saucken, Tim Bartels, and Dennis Selkoe. KTKEGV repeat motifs are key mediators of normal alpha-synuclein tetramerization : Their mutation causes excess monomers and neurotoxicity. *Proceedings of the National Academy of Sciences*, 112(31) :9596–9601, August 2015. Publisher : National Academy of Sciences Section : Biological Sciences.
- [103] Wei Wang, Iva Perovic, Johnathan Chittuluru, Alice Kaganovich, Linh T. T. Nguyen, Jingling Liao, Jared R. Auclair, Derrick Johnson, Anuradha Landaru, Alana K. Simorellis, Shulin Ju, Mark R. Cookson, Francisco J. Asturias, Jeffrey N. Agar, Brian N. Webb, ChulHee Kang, Dagmar Ringe, Gregory A. Petsko, Thomas C. Pochapsky, and Quyen Q. Hoang. A soluble alpha-synuclein construct forms a dynamic tetramer. *Proceedings of the National Academy of Sciences*, 108(43) :17797–17802, October 2011. Publisher : National Academy of Sciences Section : Biological Sciences.
- [104] Agata Rekas, Robert B. Knott, Anna Sokolova, Kevin J. Barnham, Keyla A. Perez, Colin L. Masters, Simon C. Drew, Roberto Cappai, Cyril C. Curtain, and Chi L. L. Pham. The structure of dopamine induced alpha-synuclein oligomers. *European Biophysics Journal*, 39(10) :1407–1419, September 2010.
- [105] P. H. Weinreb, W. Zhen, A. W. Poon, K. A. Conway, and P. T. Lansbury. NACP, a protein implicated in Alzheimer’s disease and learning, is natively unfolded. *Biochemistry*, 35(43) :13709–13715, October 1996.
- [106] W. Sean Davidson, Ana Jonas, David F. Clayton, and Julia M. George. Stabilization of alpha-Synuclein Secondary Structure upon Binding to Synthetic Membranes \*. *Journal of Biological Chemistry*, 273(16) :9443–9449, April 1998. Publisher : Elsevier.
- [107] Katsuya Araki, Naoto Yagi, Rie Nakatani, Hiroshi Sekiguchi, Masatomo So, Hisashi Yagi, Noboru Ohta, Yoshitaka Nagai, Yuji Goto, and Hideki Mochizuki. A small-angle X-ray scattering study of alpha-synuclein from human red blood cells. *Scientific Reports*, 6(1) :30473, July 2016. Number : 1 Publisher : Nature Publishing Group.
- [108] Kelly A. Conway, James D. Harper, and Peter T. Lansbury. Accelerated in vitro fibril formation by a mutant -synuclein linked to early-onset parkinson disease. *Nature Medicine*, 4(11) :1318–1320, nov 1998.

- 
- [109] Eric A. Greenbaum, Charles L. Graves, Amanda J. Mishizen-Eberz, Michael A. Lupoli, David R. Lynch, S. Walter Englander, Paul H. Axelsen, and Benoit I. Giasson. The e46k mutation in  $\alpha$ -synuclein increases amyloid fibril formation. *Journal of Biological Chemistry*, 280(9) :7800–7807, mar 2005.
- [110] Mathew H. Horrocks, Laura Tosatto, Alexander J. Dear, Gonzalo A. Garcia, Marija Iljina, Nunilo Cremades, Mauro Dalla Serra, Tuomas P. J. Knowles, Christopher M. Dobson, and David Klenerman. Fast Flow Microfluidics and Single-Molecule Fluorescence for the Rapid Characterization of  $\alpha$ -Synuclein Oligomers. *Analytical Chemistry*, 87(17) :8818–8826, September 2015.
- [111] Orkid Coskuner and Olivia Wise-Scira. Structures and free energy landscapes of the A53T mutant-type  $\alpha$ -synuclein protein and impact of A53T mutation on the structures of the wild-type  $\alpha$ -synuclein protein with dynamics. *ACS chemical neuroscience*, 4(7) :1101–1113, July 2013.
- [112] Elisa A. Waxman, Joseph R. Mazzulli, and Benoit I. Giasson. Characterization of Hydrophobic Residue Requirements for  $\alpha$ -Synuclein Fibrillization. *Biochemistry*, 48(40) :9427–9436, October 2009.
- [113] Ciaran P. A. Doherty, Sabine M. Ulamec, Roberto Maya-Martinez, Sarah C. Good, Jemma Makepeace, G. Nasir Khan, Patricija van Oosten-Hawle, Sheena E. Radford, and David J. Brockwell. A short motif in the n-terminal region of  $\alpha$ -synuclein is critical for both aggregation and function. *Nature Structural & Molecular Biology*, 27(3) :249–259, mar 2020.