Biochemical and Biophysical Research Communications 441 (2013) 308–317

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

# Evolutionary aspects of the synuclein super-family and sub-families based on large-scale phylogenetic and group-discrimination analysis

Jiawen Yuan, Yuwu Zhao *

*Department of Neurology, Shanghai Sixth People's Hospital affiliated to Shanghai Jiao Tong University, Shanghai 200233, China*

ARTICLE INFO

*Article history:*
Received 23 September 2013
Available online 15 October 2013

*Keywords:*
Synuclein
Phylogenetic tree
His50
Parkinson's disease

ABSTRACT

Over the last decade, many genetic studies have suggested that the synucleins, which are small, natively unfolded proteins, are closely related to Parkinson's disease and cancer. Less is known about the molecular basis of this role. A comprehensive analysis of the evolutionary path of the synuclein protein family may reveal the relationship between evolutionarily conserved residues and protein function or structure. The phylogeny of 252 unique synuclein sequences from 73 organisms suggests that gamma-synuclein is the common ancestor of alpha- and beta-synuclein. Although all three sub-families remain highly conserved, especially at the N-terminal, nearly 15% of the residues in each sub family clearly diverged during evolution, providing crucial guidance for investigations of the different properties of the members of the superfamily. His50 is found to be an alpha-specific conserved residue (91%) and, based on mutagenesis, evolutionarily developed a secondary copper binding site in the alpha synuclein family. Surprisingly, this site is located between two well-known polymorphisms of alpha-synuclein, E46K and A53T, which are linked to early-onset Parkinson's disease, suggesting that the mutation-induced impairment of copper binding could be a mechanism responsible for alpha-synuclein aggregation.



## 1. Introduction

Synucleins have been studied for more than 2 decades and have received increasing attention after studies demonstrated that alpha-synuclein polymorphisms are genetically linked to Parkinson's disease [1,2], while the gamma-synucleins are over-expressed in breast tumours [3].

The synuclein family, similar to the majority of proteins, is organised hierarchically into three sub families: alpha, beta and gamma. In humans, the alpha-synuclein gene is located at 4q21.3–q22 [4], beta-synuclein is mapped to 5q35 [5], and gamma-synuclein is found at 10q23 [6]. All three members have five protein-coding exons; however, beta has 6 exons in total, whereas alpha has seven. So far, synucleins are only found in vertebrates. The alpha- and beta-synucleins are expressed mainly in the presynaptic terminals of brain tissue, while the gamma-synucleins are found primarily in the peripheral nervous system and retina [7]. Five isoforms have been reported in the alpha family, while the beta and gamma families have only a single isoform.

All synuclein proteins contain a highly conserved N-terminal with 4–7 repetitive 11-mer motifs. The repeats are degenerate, and less is known about the role of these domains. The C-terminal region is generally unfolded and acidic. No global conservation is found within the C-terminal. Instead, it is highly variable in size and sequence.

The normal function of synucleins remains unknown. However, a number of biological processes and cellular pathways involving the synucleins have been uncovered in recent years.

Alpha-synucleins are involved in the maintenance of the synaptic vesicle pool and in the regulation of dopamine biosynthesis, homeostasis and proteasome activities. Gamma-synucleins play a chaperone role in the stimulation of the oestrogen receptor-alpha signalling pathway and is over-expressed in several types of cancer. The level of its expression is considered a prognostic marker for the early identification of tumourigenesis [8]. Beta-synucleins can inhibit alpha-synuclein aggregation and fibril formation. It has also been shown that beta-synucleins can protect against oxidative stress via the inactivation of the c-Jun N-terminal kinase signalling pathway [9,10]. Numerous data have also indicated that the synucleins are involved in two large categories of human diseases: neurodegenerative diseases (NDDs) and cancer. For example, alpha-synuclein is associated with Parkinson's disease and Alzheimer's disease, and the over-expression of gamma-synucleins is correlated with the progression of breast cancer [11].

Nevertheless, the molecular basis and mechanisms underlying the formation of toxic forms of synuclein, the disruption of their normal functions and the contribution of aggregation to NDDs remain unknown [12].

* Corresponding author. Fax: +86 21 64085875.
*E-mail address:* zhaoyuwu2005@126.com (Y. Zhao).

0006-291X/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bbrc.2013.09.132

Generally, different members of a protein family are evolved from a single common ancestor gene through duplication and divergence. New functions are gained through the adaptation of beneficial changes in the protein sequence. These changes can be traced by analysing the sequences of all members of the superfamily using comparative bioinformatics and evolutionary analysis. This type of analysis can provide detailed information regarding the evolutionary path of the sub-families and increase understanding of the common and unique characteristics of the function and structure of different sub-families.

## 2. Methods

### 2.1. Data collection

To obtain complete synuclein protein sequence data, a 2-iteration PSI-Blast [13] was performed against the NCBI nr protein database using the sequence of human alpha, beta and gamma synuclein (downloaded from the Uniprot sequence database [14] as the query. The E-value cutoff was set to 1e-5. A domain search was performed using RPS-BLAST [15] against pfam and the NCBI CDD database [16].A sequence was assigned to the final list of the synuclein superfamily when it hit either pfam01387 in the pfam database or cl03193 in the CDD database, with an e-value <1e-5. Length control and the removal of redundant sequences were performed by perl scripts developed in-house.

### 2.2. Multiple sequence alignment

Multiple sequence alignments were created by MUSCLE [17] and refined by RASCAL [18].

### 2.3. Phylogeny trees

Phylogeny trees were inferred using FastTree [19] with a WAG + CAT model and rooted by midpoint [20].

### 2.4. Conservation analysis

Three conservation paradigms, absolute conservation, polar conservation and hydrophobic conservation (as defined in Liu et al. [21]), were used to assess the level of sequence conservation at each position in the alignment.

### 2.5. Discrimination analysis

Alignments were rebuilt for the alpha, beta and gamma subgroups separately, and an absolute conservation score was calculated for each position to obtain a group-conservation score. A position was assigned to a particular group-discriminated position when the following criteria were met:

1. The group-conservation score ($C_i$) of this i-th position (group) was >0.6
2. The majority of the amino acids used at this position in one group differed from those used in the other group.
3. The group-conservation score ($C_i$) of the first group (group1) exceeded the $C_i$ of the other group (group2) by 2-fold when the same amino acid was used at the i-th position.

In this study, 7 classes were defined: alpha-specific, beta-specific, gamma-specific, alpha–beta, beta–gamma, alpha–gamma and alpha–beta–gamma. For instance, a position assigned to the alpha–beta class indicates that this position is over 60% conserved in both the alpha and beta sub families; however, different residues evolved at this position.

The 3D structure of alpha-synuclein was obtained from the PDB protein structure database [22] and illustrated using RASMOL [23].

Solvent accessibility analysis was performed using SSpro 4.0 [24].

## 3. Results

### 3.1. Synuclein sequence data

The three search results were merged into a unique, non-redundant hit list that included 480 sequences. Three hits that were identified as synuclein binding proteins by domain search were deleted from the dataset. To ensure the quality of the sequence alignment, sequences that were shorter than 90 or longer than 150 amino acids in length were also removed from the dataset. Identical sequences from the same organism were removed from the dataset. The final dataset used for the analysis had 252 sequences. These sequences were from 73 organisms, all of which were from vertebrates.

### 3.2. The large-scale phylogenetic tree

Reveals a clear relationship among the three sub families. The results corroborate the family tree structure of Lavedan et al. [25], but not George et al. [7]. Alpha and beta are more closely related to each other because they are clustered in one sub tree with a shorter branch length than gamma, which has a longer evolutionary history (Fig. 1).

### 3.3. Subfamily

Based on the clear phylogeny, sequences were classified into three subfamilies: 103 sequences for alpha-synuclein, 72 for beta-synuclein and 77 for gamma-synuclein, which were derived from 63, 45 and 44 organisms, respectively (Fig. 2A).

In 20 organisms (31%), there were only alpha-synucleins present; we call these alpha-specific organisms. In contrast, beta-specific and gamma-specific organisms were found only 5 times (7%) and 2 times (3%), respectively.

### 3.4. Conservation analysis

The conservation of each amino acid in the synuclein superfamily and three sub families was calculated based on identity and class similarity (see Section 2).

In total, 79 positions exceeded a conservation level of 60%. Across all three sub-families, 42 residues were over 90% conserved. In addition, 37 were located in the N-terminal (residues 1–59), and 5 were within the known non-amyloid-beta component of the Alzheimer's disease amyloid plaques (NAC) region (residues 60–95), indicating strong selection effects on these positions.

These conserved residues comprise 50% of the 6 known 11-mer repeats that are similar to the apolipoprotein-like class-A2 helix that mediates binding to phospholipid vesicles [25].

The first three repeats (I, II and III) were much more conserved than the other three. Each of the first three repeats had only two residues that were not well conserved.

In repeat II, two residues (27A and 31G) of human alpha synuclein were alpha-specific residues (Table 1). In the beta and gamma groups, T and E were originally used in these two positions. These evolutionary changes are a clear relaxation of the binding affinity of the peptide and could indicate a functional change of the alpha and beta–gamma sub families.

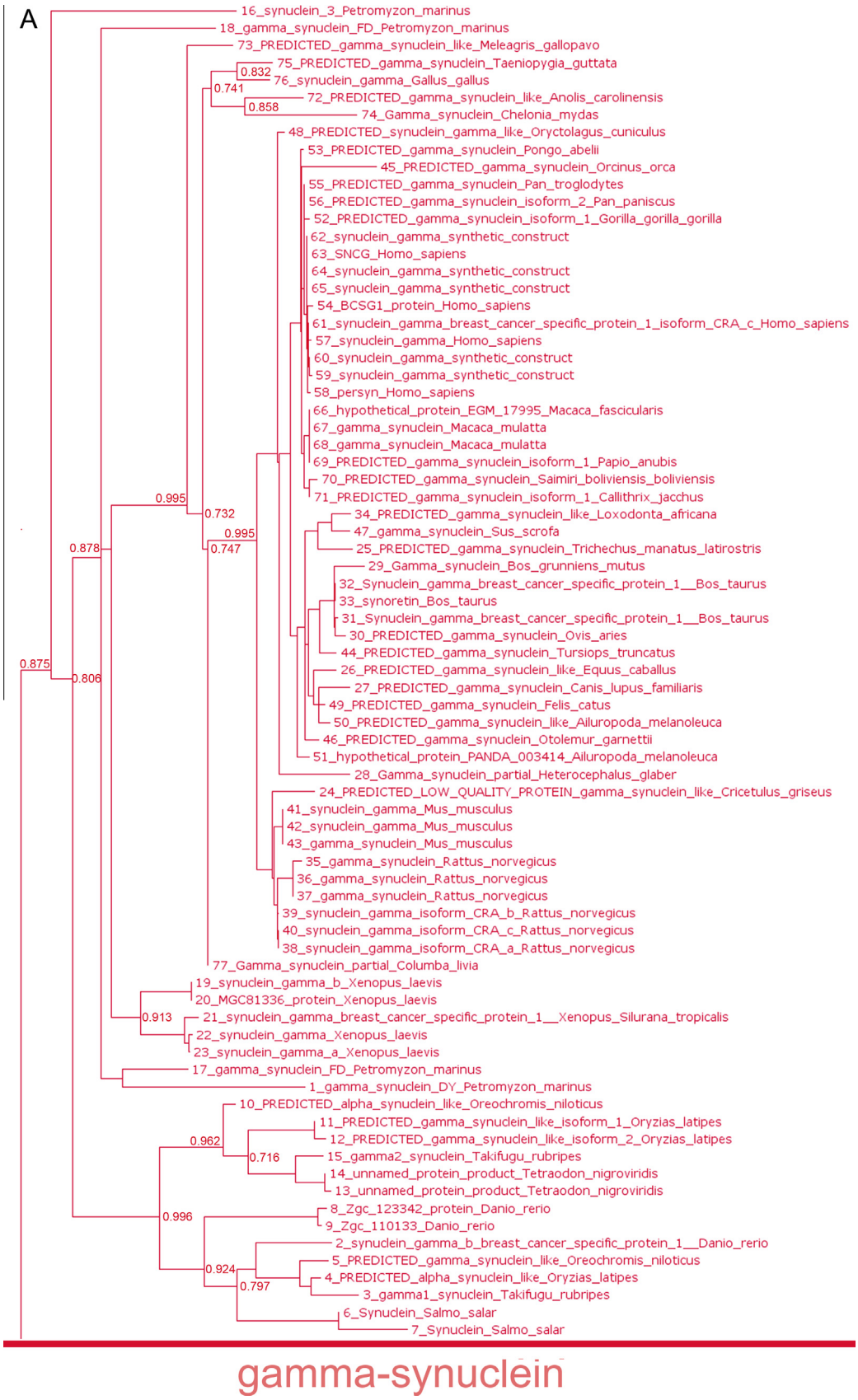
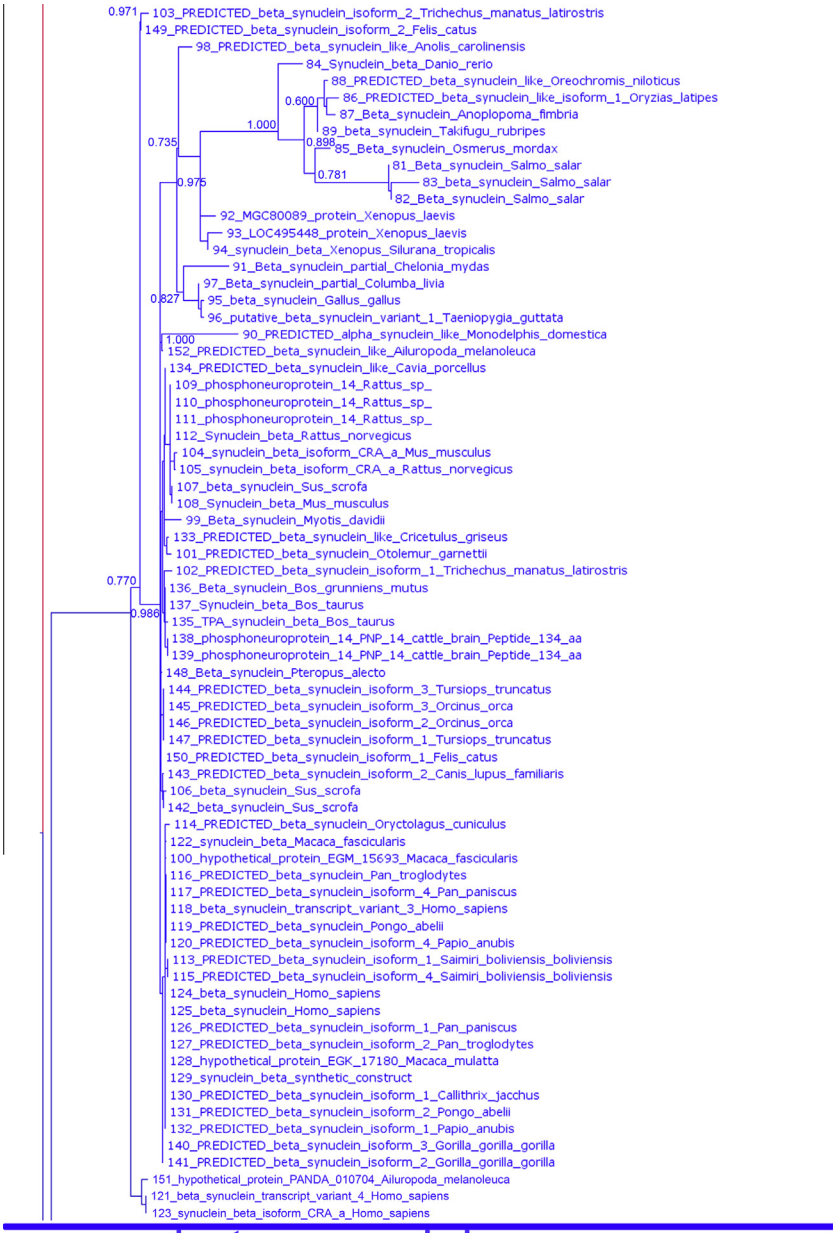



**Fig. 1.** Phylogeny of the synuclein superfamily. (A) Maximum likelihood phylogeny based on 252 synuclein sequences from 73 organisms. (B) Distance phylogeny from a 2001 review with 16 sequences. (C) Distance phylogeny from a 1998 review with only 12 sequences. Clearly, the evolutionary relationships among the 3 sub-families revealed by the comprehensive tree are consistent with the tree from 1998.



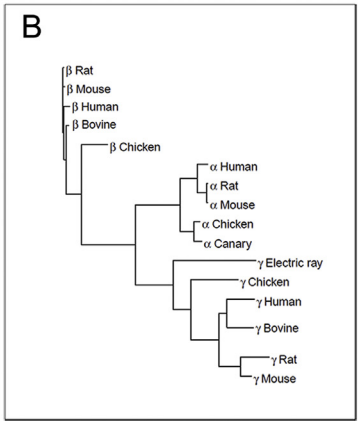




**Fig. 1** (*continued*)

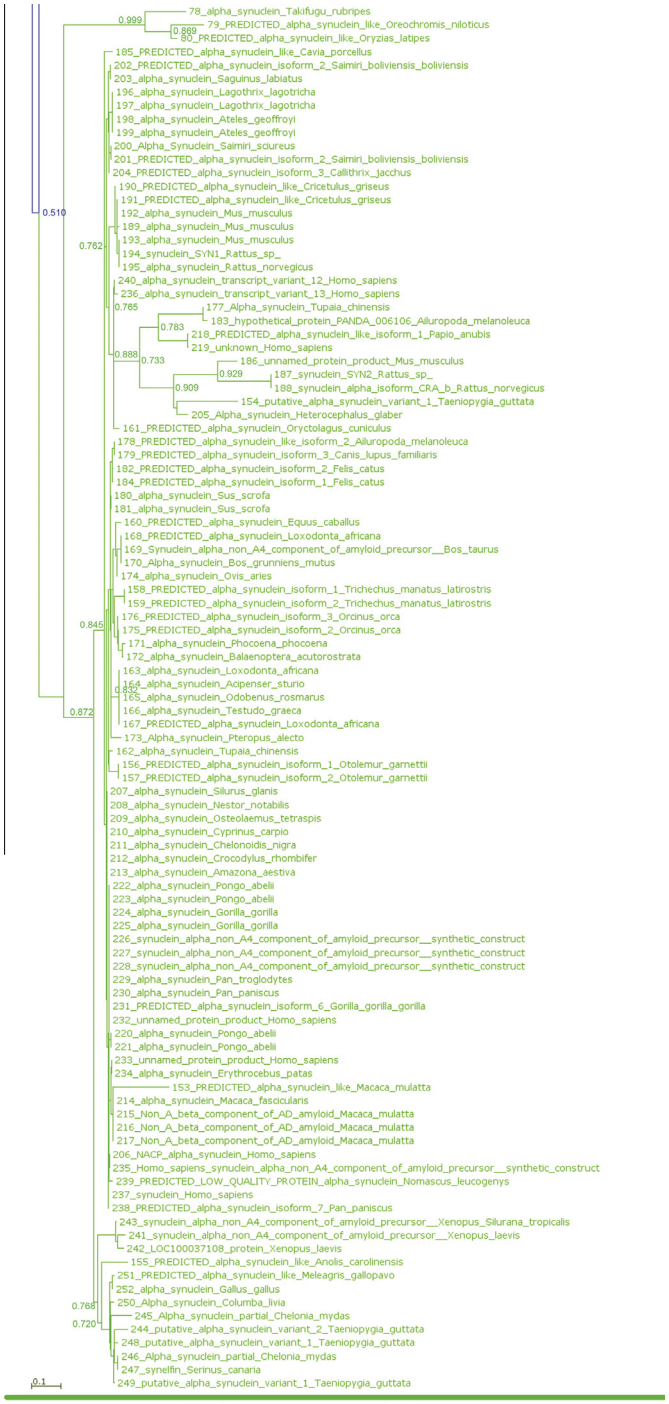



**Fig. 1** (*continued*)

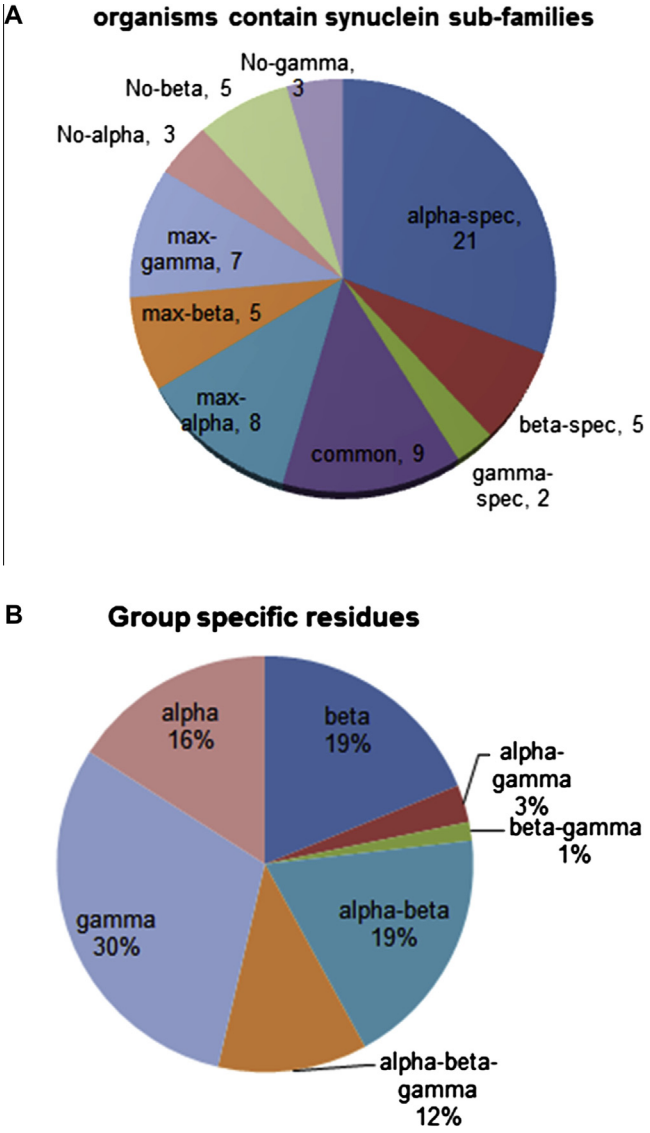



**Fig. 2.** (A) Organisms that contain the various sub families of synucleins. Common: organisms that contain all three sub families, with all of the sub families having the same number of copies; Alpha/beta/gamma-spec: Only alpha/beta/gamma-synuclein is found in the organism; No-alpha/beta/gamma: alpha/beta/gamma-synuclein is not observed in the organism; Max-alpha/beta/gamma: all three sub families are found in the organism, with a greater copy number in one sub family (alpha/beta/gamma-synuclein) than in the other two sub families. (B) Distribution of group-specific residues.

In repeat III, 2 residues, 38L and 42S, which were located on the second loop-linking helix, were not well conserved (64% and 59%, respectively). However, 38L was 99% conserved when considering hydrophobic regions, and 42S was 86% neutral. Together with residues 39–41, which were 95% conserved on average, this finding may indicate the importance of the loop in enabling lipid binding with a large shift of helix structure (Fig. 3, Table 1).

Repeats IV and V were considerably less conserved than repeats I, II and III. In each repeat, fewer than 5 positions remained over 90% conserved. Especially in repeat IV, only two positions exceeded a 90% conservation level. The other positions were accumulating mutations. It is noteworthy that repeat IV was generally lacking in isoforms 2–5 and that repeats I to IV were connected directly, while, in contrast, there was a link insertion of 4 amino acids be-

**Table 1**
Absolute conservation positions within repeat I, II and III.

| ID | Alignment Pos | Consensus | Level | Alpha | | Gamma | | Beta | | Type |
|---|---|---|---|---|---|---|---|---|---|---|
| Repeat I | 32 | ***K*** | ***0.46*** | ***10*** | ***K*** | ***10*** | ***I*** | ***10*** | ***M*** | ***Absolute*** |
| | 33 | A | 0.97 | 11 | A | 11 | A | 11 | A | Absolute |
| | 34 | K | 0.95 | 12 | K | 12 | K | 12 | K | Absolute |
| | 35 | E | 0.93 | 13 | E | 13 | E | 13 | E | Absolute |
| | 36 | G | 0.98 | 14 | G | 14 | G | 14 | G | Absolute |
| | 37 | V | 0.93 | 15 | V | 15 | V | 15 | V | Absolute |
| | 38 | V | 0.93 | 16 | V | 16 | V | 16 | V | Absolute |
| | ***39*** | ***A*** | ***0.74*** | ***17*** | ***A*** | ***17*** | ***G*** | ***17*** | ***A*** | ***Absolute*** |
| | 40 | A | 0.97 | 18 | A | 18 | A | 18 | A | Absolute |
| | ***41*** | ***A*** | ***0.78*** | ***19*** | ***A*** | ***19*** | ***V*** | ***19*** | ***A*** | ***Absolute*** |
| | 42 | E | 0.98 | 20 | E | 20 | E | 20 | E | Absolute |
| Repeat II | 43 | K | 0.97 | 21 | K | 21 | K | 21 | K | Absolute |
| | 44 | T | 0.97 | 22 | T | 22 | T | 22 | T | Absolute |
| | 45 | K | 0.97 | 23 | K | 23 | K | 23 | K | Absolute |
| | 46 | Q | 0.88 | 24 | Q | 24 | Q | 24 | Q | Absolute |
| | 47 | G | 0.97 | 25 | G | 25 | G | 25 | G | Absolute |
| | 48 | V | 0.96 | 26 | V | 26 | V | 26 | V | Absolute |
| | ***49*** | ***T*** | ***0.46*** | ***27*** | ***A*** | ***27*** | ***T*** | ***27*** | ***T*** | ***Absolute*** |
| | 50 | E | 0.9 | 28 | E | 28 | E | 28 | E | Absolute |
| | 51 | A | 0.98 | 29 | A | 29 | A | 29 | A | Absolute |
| | 52 | A | 0.98 | 30 | A | 30 | A | 30 | A | Absolute |
| | ***53*** | ***E*** | ***0.55*** | ***31*** | ***G*** | ***31*** | ***E*** | ***31*** | ***E*** | ***Absolute*** |
| Repeat III | 54 | K | 0.95 | 32 | K | 32 | K | 32 | K | Absolute |
| | 55 | T | 0.98 | 33 | T | 33 | T | 33 | T | Absolute |
| | 56 | K | 0.98 | 34 | K | 34 | K | 34 | K | Absolute |
| | 68 | E | 0.96 | 35 | E | 35 | E | 35 | E | Absolute |
| | 69 | G | 0.98 | 36 | G | 36 | G | 36 | G | Absolute |
| | 70 | V | 0.96 | 37 | V | 37 | V | 37 | V | Absolute |
| | ***71*** | ***L*** | ***0.64*** | ***38*** | ***L*** | ***38*** | ***M*** | ***38*** | ***L*** | ***Absolute*** |
| | 72 | Y | 0.94 | 39 | Y | 39 | Y | 39 | Y | Absolute |
| | 73 | V | 0.96 | 40 | V | 40 | V | 40 | V | Absolute |
| | 99 | G | 0.93 | 41 | G | 41 | G | 41 | G | Absolute |
| | ***100*** | ***S*** | ***0.59*** | ***42*** | ***S*** | ***42*** | ***A*** | ***42*** | ***S*** | ***Absolute*** |

Positions highlighted in bold italic are positions that were less than 90% conserved. Columns from left to right are: repeat index, alignment position of the synuclein superfamily, consensus residue, conservation level, real position in human alpha-synuclein, residue used in alpha-synuclein, real position in human gamma-synuclein, residue used in gamma-synuclein, real position in human beta-synuclein, residue used in beta-synuclein, conservation type.

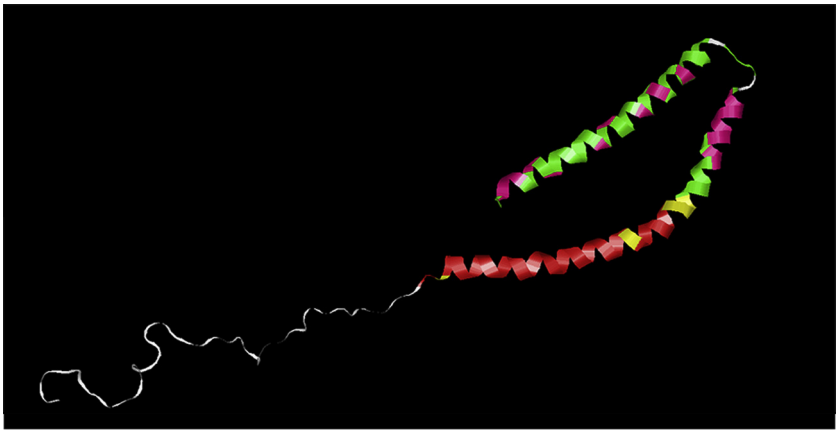


**Fig. 3.** 3D structure of alpha-synuclein (PDB:1xq8). The N-terminal (residues 1–59) is coloured pink and the NAC region (residues 60–95) is coloured red. Residues within the N-terminal that were over 90% conserved are coloured green and conserved residues within the NAC region are coloured yellow. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

tween IV and V. Interestingly, the link was also well conserved and could be an important structural component that allows synucleins to correctly bind to targets. Further experimental work is needed to determine whether this short link binds directly to the targets or acts as a switch from a native to active state (Table 2).

Repeat VI was 11 residues downstream of repeat V, and none if its residues were over 90% conserved. A deletion of the first 5 residues of repeat VI was observed in the beta sub family. Interestingly, exactly 11 residues remained between repeat V and repeat VI, and Zhao et al. [26] defined it as the 7th repeat, although with less similarity than repeats I–VI. In secondary structure, this 11-mer comprises a helical pattern similar to repeats V and VI (Table 3).

### 3.5. Group discrimination analysis

To fully understand the evolutionary path of the three subgroups, it is important to perform an overall investigation of the

**Table 2**
Absolute conservation positions within repeat IV and V.

| ID | Alignment Pos | Consensus | Level | Alpha | | Gamma | | Beta | | Type |
|---|---|---|---|---|---|---|---|---|---|---|
| Repeat IV | ***101*** | ***K*** | ***0.89*** | ***43*** | ***K*** | ***43*** | ***K*** | ***43*** | ***K*** | ***Absolute*** |
| | 102 | T | 0.9 | 44 | T | 44 | T | 44 | T | Absolute |
| | ***103*** | ***K*** | ***0.72*** | ***45*** | ***K*** | ***45*** | ***K*** | ***45*** | ***R*** | ***Absolute*** |
| | ***104*** | ***E*** | ***0.81*** | ***46*** | ***E*** | ***46*** | ***E*** | ***46*** | ***E*** | ***Absolute*** |
| | ***105*** | ***G*** | ***0.75*** | ***47*** | ***G*** | ***47*** | ***N*** | ***47*** | ***G*** | ***Absolute*** |
| | ***106*** | ***V*** | ***0.87*** | ***48*** | ***V*** | ***48*** | ***V*** | ***48*** | ***V*** | ***Absolute*** |
| | ***107*** | ***V*** | ***0.85*** | ***49*** | ***V*** | ***49*** | ***V*** | ***49*** | ***V*** | ***Absolute*** |
| | ***108*** | ***H*** | ***0.41*** | ***50*** | ***H*** | ***50*** | ***Q*** | ***50*** | ***Q*** | ***Absolute*** |
| | ***109*** | ***G*** | ***0.63*** | ***51*** | ***G*** | ***51*** | ***S*** | ***51*** | ***G*** | ***Absolute*** |
| | 110 | V | 0.9 | 52 | V | 52 | V | 52 | V | Absolute |
| | ***111*** | ***T*** | ***0.51*** | ***53*** | ***A*** | ***53*** | ***T*** | ***53*** | ***A*** | ***Absolute*** |
| Linker | ***112*** | ***T*** | ***0.49*** | ***54*** | ***T*** | ***54*** | ***S*** | ***54*** | ***S*** | ***Absolute*** |
| | 114 | V | 0.97 | 55 | V | 55 | V | 55 | V | Absolute |
| | 115 | A | 0.97 | 56 | A | 56 | A | 56 | A | Absolute |
| | 116 | E | 0.91 | 57 | E | 57 | E | 57 | E | Absolute |
| Repeat V | 117 | K | 0.96 | 58 | K | 58 | K | 58 | K | Absolute |
| | 118 | T | 0.96 | 59 | T | 59 | T | 59 | T | Absolute |
| | 119 | K | 0.9 | 60 | K | 60 | K | 60 | K | Absolute |
| | 120 | E | 0.92 | 61 | E | 61 | E | 61 | E | Absolute |
| | 121 | Q | 0.93 | 62 | Q | 62 | Q | 62 | Q | Absolute |
| | ***122*** | ***A*** | ***0.54*** | ***63*** | ***V*** | ***63*** | ***A*** | ***63*** | ***A*** | ***Absolute*** |
| | ***123*** | ***T*** | ***0.34*** | ***64*** | ***T*** | ***64*** | ***N*** | ***64*** | ***S*** | ***Absolute*** |
| | ***124*** | ***N*** | ***0.38*** | ***65*** | ***N*** | ***65*** | ***A*** | ***65*** | ***H*** | ***Absolute*** |
| | ***125*** | ***V*** | ***0.68*** | ***66*** | ***V*** | ***66*** | ***V*** | ***66*** | ***L*** | ***Absolute*** |
| | ***126*** | ***G*** | ***0.75*** | ***67*** | ***G*** | ***67*** | ***S*** | ***67*** | ***G*** | ***Absolute*** |
| | ***127*** | ***G*** | ***0.61*** | ***68*** | ***G*** | ***68*** | ***E*** | ***68*** | ***G*** | ***Absolute*** |

Positions highlighted in bold italic are positions that were less than 90% conserved. The linker indicated the linker sequence located between repeat IV and V. Columns are the same as in Table 1.

**Table 3**
Absolute conservation positions within repeat VI and the linker sequence located between repeats V(indicated in Table 2) and VI.

| ID | Alignment pos | Consensus | Level | Alpha | | Gamma | | Beta | | Type |
|---|---|---|---|---|---|---|---|---|---|---|
| Repeat ? | 128 | A | 0.94 | 69 | A | 69 | A | 69 | A | Absolute |
| | ***129*** | ***V*** | ***0.89*** | ***70*** | ***V*** | ***70*** | ***V*** | ***70*** | ***V*** | ***Absolute*** |
| | ***130*** | ***V*** | ***0.71*** | ***71*** | ***V*** | ***71*** | ***V*** | ***71*** | ***F*** | ***Absolute*** |
| | ***131*** | ***T*** | ***0.44*** | ***72*** | ***T*** | ***72*** | ***S*** | ***72*** | ***S*** | ***Absolute*** |
| | ***132*** | ***G*** | ***0.75*** | ***73*** | ***G*** | ***73*** | ***S*** | ***73*** | ***G*** | ***Absolute*** |
| | ***133*** | ***V*** | ***0.64*** | ***74*** | ***V*** | ***74*** | ***V*** | ***73*** | – | ***Absolute*** |
| | ***134*** | ***T*** | ***0.41*** | ***75*** | ***T*** | ***75*** | ***N*** | ***73*** | – | ***Absolute*** |
| | ***135*** | ***A*** | ***0.41*** | ***76*** | ***A*** | ***76*** | ***T*** | ***73*** | – | ***Absolute*** |
| | ***136*** | ***V*** | ***0.69*** | ***77*** | ***V*** | ***77*** | ***V*** | ***73*** | – | ***Absolute*** |
| | ***137*** | ***A*** | ***0.67*** | ***78*** | ***A*** | ***78*** | ***A*** | ***73*** | – | ***Absolute*** |
| | ***138*** | ***Q*** | ***0.44*** | ***79*** | ***Q*** | ***79*** | ***T*** | ***73*** | – | ***Absolute*** |
| Repeat VI | ***139*** | ***K*** | ***0.65*** | ***80*** | ***K*** | ***80*** | ***K*** | ***73*** | – | ***Absolute*** |
| | ***140*** | ***T*** | ***0.69*** | ***81*** | ***T*** | ***81*** | ***T*** | ***73*** | – | ***Absolute*** |
| | ***141*** | ***V*** | ***0.71*** | ***82*** | ***V*** | ***82*** | ***V*** | ***73*** | – | ***Absolute*** |
| | ***142*** | ***E*** | ***0.71*** | ***83*** | ***E*** | ***83*** | ***E*** | ***73*** | – | ***Absolute*** |
| | ***143*** | ***G*** | ***0.52*** | ***84*** | ***G*** | ***84*** | ***E*** | ***73*** | – | ***Absolute*** |
| | ***144*** | ***A*** | ***0.86*** | ***85*** | ***A*** | ***85*** | ***A*** | ***74*** | ***A*** | ***Absolute*** |
| | ***145*** | ***G*** | ***0.69*** | ***86*** | ***G*** | ***86*** | ***E*** | ***75*** | ***G*** | ***Absolute*** |
| | ***146*** | ***N*** | ***0.7*** | ***87*** | ***S*** | ***87*** | ***N*** | ***76*** | ***N*** | ***Absolute*** |
| | ***147*** | ***I*** | ***0.89*** | ***88*** | ***I*** | ***88*** | ***I*** | ***77*** | ***I*** | ***Absolute*** |
| | ***148*** | ***A*** | ***0.86*** | ***89*** | ***A*** | ***89*** | ***A*** | ***78*** | ***A*** | ***Absolute*** |
| | ***149*** | ***A*** | ***0.74*** | ***90*** | ***A*** | ***90*** | ***V*** | ***79*** | ***A*** | ***Absolute*** |

Positions highlighted in bold italic are positions that were less than 90% conserved. Repeat ? indicated the linker sequence located between repeats V and VI, as it is an exact 11-mer and comprises a helix in secondary structure. Columns are the same as in Table 1.

group-discriminated residues in addition to the global conservation analysis. Residues that are conserved cross the entire superfamily are crucial for structure and common molecular functions, while some beneficial mutations are fixed in sub-families during evolution and define new conformational changes or sub-family-specific functions.

To fully uncover functionally or structurally important residues in each sub family, mutations were clustered into 7 classes: alpha-specific, beta-specific, gamma-specific, alpha–beta, beta–gamma, alpha–gamma and alpha–beta–gamma (see Section 2). Of the group-specific mutations, 30% were gamma-specific, while 16% and 19% belonged to the alpha and beta class, respectively, indicating a clear direction for a mutagenesis study investigating the various functions of the sub-families (Fig. 2B).

As shown in Table 4, a list of lab-induced mutagenesis studies of synucleins [14], each of them can be clearly classified, which is useful in interpreting the mutagenesis results. For instance, position 39 was conserved across the synuclein superfamily. However, the Y39F mutant did not exhibit any effect on osmotic stress-induced phosphorylation, which could indicate the importance of this position is structural rather than due to the phosphorylation site at 39Y. Position 2 is a common copper-binding site in all synu-

**Table 4**
Classification of known lab-induced mutagenesis studies.

| Type | Con.value | | Position | Length | Experimental information |
|---|---|---|---|---|---|
| Common | | Mutagenesis | 2 | 1 | D → A: Impairs copper-binding. Ref. [25] |
| Common | 0.94 | Mutagenesis | 39 | 1 | Y → F: No effect on osmotic stress-induced phosphorylation. Ref. [21] |
| Alpha | 0.91 | Mutagenesis | 50 | 1 | H → A: Impairs copper-binding. Ref. [25] |
| 67,68 Gamma<br>71 Beta | – | Mutagenesis | 67–71 | 5 | Missing: Reduces polymerisation into amyloid fibrils. Ref. [23] |
| 71 Beta<br>72 Alpha<br>73 Gamma<br>79 Alpha–gamma | – | Mutagenesis | 71–82 | 12 | Missing: Impairs polymerisation into amyloid fibrils. Ref. [23] |
| Not conserved | 0.55 | Mutagenesis | 76–77 | 2 | Missing: Impairs polymerisation into amyloid fibrils. Ref. [23] |
| Not conserved | 0.41 | Mutagenesis | 76 | 1 | Missing: Does not affect polymerisation into amyloid fibrils. |
| Not conserved | 0.69 | Mutagenesis | 77 | 1 | Missing: Does not affect polymerisation into amyloid fibrils. Ref. [23] |
| Not conserved | 0.64 | Mutagenesis | 78 | 1 | Missing: Does not affect polymerisation into amyloid fibrils. Ref. [23] |
| Gamma | 0.95 | Mutagenesis | 85–94 | 10 | Missing: Reduces polymerisation into amyloid fibrils. Ref. [23] |
| Gamma | 0.83 | Mutagenesis | 125 | 1 | Y → F: Abolishes osmotic stress-induced phosphorylation. Ref. [21] |
| Not conserved | 0.75 | Mutagenesis | 133 | 1 | Y → F: No effect on osmotic stress-induced phosphorylation. Ref. [21] |
| Not conserved | 0.79 | Mutagenesis | 136 | 1 | Y → F: No effect on osmotic stress-induced phosphorylation. Ref. [21] |

cleins; however, position 50 should be an extended copper-binding site that evolved in the alpha sub family (Table 4).

### 3.6. Discriminated residues of alpha synuclein

Eleven alpha-only discriminated positions were uncovered. 50H is a known copper-binding site based on a lab-induced mutagenesis, and H50A exhibits significantly impaired copper binding [27]. Another copper-binding site is 2D, which is 95% conserved across the entire synuclein superfamily [27]. Copper binding sites are generally used for structural stabilisation. The employment of a positively charged histidine at this position, instead of a polar, uncharged glutamine, could be important for stabilising polymerisation in the alpha sub family.

87S is a phosphoserine site for CK2, which evolved from polar uncharged asparagines. 63V and 72T were 100% conserved in the alpha group (Table 5).

### 3.7. 50H and disease

It is known that certain polymorphisms in alpha-synuclein are major risk factors for sporadic Parkinson's disease (PD). Three well-known point mutations, A30P, E46K and A53T, result in dominant familial parkinsonism [28]. By highlighting these points and the alpha-specific copper-binding site 50H in the 3D structure, a tight connection between these residues was exposed: 50H is surrounded by E46, A53 and T54, the mutation E46K changes the electric field of the copper-binding region, and the mutation A53T results in narrowed accessibility of this region. Together, these changes could impair copper binding and contribute to alpha-synuclein aggregation (Fig. 4A).

Solvent accessibility analysis of alpha-synuclein predicts that position 30A will be a non-accessible site when in its natural state. Experiments have shown that monomeric wild type alpha-synuclein and two mutants, A53T and E46K, can form ion channels in planar bilayer membranes, but A30P cannot [29], suggesting that a mutation of A30P might expose this position and induce a large conformational change in nearby regions (Fig. 4B).

**Table 5**
Position of alpha-specific residues compared to their corresponding positions in the other two sub families.

| Group | Alignment pos | Alpha | | | Beta | | | Gamma | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Alpha | 49 | 27 | A | 0.96 | 27 | T | 0.75 | 27 | T | 0.78 |
| Alpha | 53 | 31 | G | 0.9 | 31 | E | 0.96 | 31 | E | 0.79 |
| Alpha | 108 | 50 | H | 0.91 | 50 | Q | 0.71 | 50 | Q | 0.6 |
| Alpha | 112 | 54 | T | 0.94 | 54 | S | 0.74 | 54 | S | 0.69 |
| Alpha | 122 | 63 | V | 1 | 63 | A | 0.89 | 63 | A | 0.92 |
| Alpha | 131 | 72 | T | 1 | 72 | S | 0.88 | 72 | S | 0.58 |
| Alpha | 146 | 87 | S | 0.62 | 76 | N | 0.99 | 87 | N | 0.92 |
| Alpha | 157 | 98 | D | 0.98 | 87 | E | 0.81 | 98 | E | 0.75 |
| Alpha | 164 | 102 | K | 0.93 | 93 | L | 0.82 | 100 | L | 0.74 |
| Alpha | 166 | 104 | E | 0.79 | 95 | P | 0.88 | 102 | P | 0.64 |
| Alpha | 198 | 121 | D | 0.72 | 115 | E | 0.93 | 116 | E | 0.49 |
| Alpha | 206 | 127 | M | 0.84 | 121 | D | 0.51 | 122 | A | 0.58 |

## 4. Discussion

Although a number of hits with other proteins can be found with low E-value searches, such as the class A2 lipid-binding domains of the apolipoproteins [30], chaperone 14-3-3 [31] and several small heat-shock-proteins [32], the coverage of these hits in the original sequences is too short to be considered proof of a common ancestor. No ancestor or precursor in prokaryotic organisms can be identified, either. The evolutionary history of the synuclein family remains unclear. A hypothesis has described the evolution of the synuclein family as taking place in two steps: gene duplication and recombination events in one synuclein ancestor [29].

For the first time, by combining sequence similarity search with the HMM model of the synuclein family, 252 synucleins from 73 organisms were identified. The maximum likelihood phylogenetic tree based on this dataset suggests that gamma synuclein should be the common ancestor of the synuclein family. The fact that the alpha branch and the beta branch are both direct descendants of the gamma branch in the rooted tree is direct evidence supporting the hypothesis mentioned above.

Alpha-synucleins, but not gamma-synucleins, were the most frequently observed synucleins across organisms in this dataset. In 21 organisms, only alpha-synucleins have been found, including some primates, such as *nomascus leucogenys* and *erythrocebus patas*. With the knowledge of the gollia, chimpanzee and human genomes, it is likely that sequences from the other two sub families will be found in these organisms in the future.

11-Mer repeats are typical structure motifs found in synucleins. Because 50% of them are highly conserved (over 90% identical), these residues should be crucial for either structure or function. Other residues seem to be undergoing divergence; some becoming more flexible through the accumulation of various mutations, while others are becoming group-specific residues via adaptive evolution.

In this study, a comprehensive list of such residues in each group was provided. Some of them have been proven to be func-

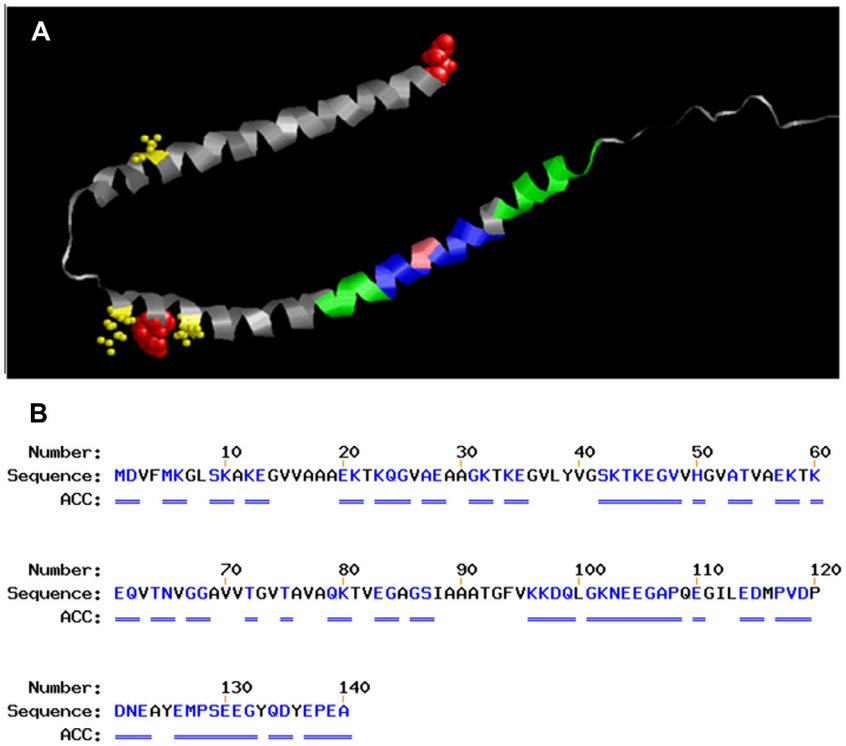



**Fig. 4.** (A) Yellow bubbles indicate known disease-associated mutations: A30P in Park1, E46 K in park1 and DLB, which exhibits significantly increased binding to negatively charged phospholipid liposomes, and A53T in Park1; red bubbles indicate the 2 copper-binding sites. 50H is located between E46 K and A53T. (B) Solvent accessibility prediction for the human alpha-synuclein sequence. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

tionally important, but more remains to be discovered. This study provides investigational clues for studying the different properties of the members of the synuclein family.

## Competing interests

The authors declare that they have no competing interests.

## Acknowledgment

We thank MHBI Bioinformatics Inc. for technical support.

## References

[1] R. Kruger, W. Kuhn, T. Muller, D. Woitalla, M. Graeber, S. Kosel, H. Przuntek, J.T. Epplen, L. Schols, O. Riess, Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease, Nat. Genet. 18 (1998) 106–108.

[2] M.H. Polymeropoulos, C. Lavedan, E. Leroy, S.E. Ide, A. Dehejia, A. Dutra, B. Pike, H. Root, J. Rubenstein, R. Boyer, E.S. Stenroos, S. Chandrasekharappa, A. Athanassiadou, T. Papapetropoulos, W.G. Johnson, A.M. Lazzarini, R.C. Duvoisin, G. Di Iorio, L.I. Golbe, R.L. Nussbaum, Mutation in the alpha-synuclein gene identified in families with Parkinson's disease, Science 276 (1997) 2045–2047.

[3] H. Ji, Y.E. Liu, T. Jia, M. Wang, J. Liu, G. Xiao, B.K. Joseph, C. Rosen, Y.E. Shi, Identification of a breast cancer-specific gene, BCSG1, by direct differential cDNA sequencing, Cancer Res. 57 (1997) 759–764.

[4] D. Campion, C. Martin, R. Heilig, F. Charbonnier, V. Moreau, J.M. Flaman, J.L. Petit, D. Hannequin, A. Brice, T. Frebourg, The NACP/synuclein gene: chromosomal assignment and screening for alterations in Alzheimer disease, Genomics 26 (1995) 254–257.

[5] M.G. Spillantini, A. Divane, M. Goedert, Assignment of human alpha-synuclein (SNCA) and beta-synuclein (SNCB) genes to chromosomes 4q21 and 5q35, Genomics 27 (1995) 379–381.

[6] N.N. Ninkina, M.V. Alimova-Kost, J.W. Paterson, L. Delaney, B.B. Cohen, S. Imreh, N.V. Gnuchev, A.M. Davies, V.L. Buchman, Organization, expression and polymorphism of the human persyn gene, Hum. Mol. Genet. 7 (1998) 1417–1424.

[7] J.M. George, The synucleins, Genome Biol. 3 (2002) REVIEWS3002.

[8] M. Ahmad, S. Attoub, M.N. Singh, F.L. Martin, O.M. El-Agnaf, Gamma-synuclein and the progression of cancer, FASEB J. 21 (2007) 3419–3430.

[9] M. Hashimoto, E. Rockenstein, M. Mante, M. Mallory, E. Masliah, Beta-Synuclein inhibits alpha-synuclein aggregation: a possible role as an anti-parkinsonian factor, Neuron 32 (2001) 213–223.

[10] M. Hashimoto, L.J. Hsu, E. Rockenstein, T. Takenouchi, M. Mallory, E. Masliah, Alpha-synuclein protects against oxidative stress via inactivation of the c-Jun N-terminal kinase stress-signaling pathway in neuronal cells, J. Biol. Chem. 277 (2002) 11465–11472.

[11] A. Surguchov, Synucleins: are they two-edged swords?, J Neurosci. Res. 91 (2013) 161–166.

[12] H.A. Lashuel, C.R. Overk, A. Oueslati, E. Masliah, The many faces of alpha-synuclein: from structure and toxicity to therapeutic target, Nat. Rev. Neurosci. 14 (2013) 38–48.

[13] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.

[14] C. UniProt, Reorganizing the protein space at the Universal Protein Resource (UniProt), Nucleic Acids Res. 40 (2012) D71–D75.

[15] A. Marchler-Bauer, S.H. Bryant, CD-Search: protein domain annotations on the fly, Nucleic Acids Res. 32 (2004) W327–W331.

[16] A. Marchler-Bauer, S. Lu, J.B. Anderson, F. Chitsaz, M.K. Derbyshire, C. DeWeese-Scott, J.H. Fong, L.Y. Geer, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, J.D. Jackson, Z. Ke, C.J. Lanczycki, F. Lu, G.H. Marchler, M. Mullokandov, M.V. Omelchenko, C.L. Robertson, J.S. Song, N. Thanki, R.A. Yamashita, D. Zhang, N. Zhang, C. Zheng, S.H. Bryant, CDD: a Conserved Domain Database for the functional annotation of proteins, Nucleic Acids Res. 39 (2011) D225–D229.

[17] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797.

[18] J.D. Thompson, J.C. Thierry, O. Poch, RASCAL: rapid scanning and correction of multiple sequence alignments, Bioinformatics 19 (2003) 1155–1161.

[19] M.N. Price, P.S. Dehal, A.P. Arkin, FastTree: computing large minimum evolution trees with profiles instead of a distance matrix, Mol. Biol. Evol. 26 (2009) 1641–1650.

[20] P.N. Hess, C.A. De Moraes Russo, An empirical test of the midpoint rooting method, Biol. J. Linn. Soc. 92 (2007) 669–674.

[21] X. Liu, Comprehensive bioinformatic analysis of kinesin classification and prediction of structural changes from a closed to an open conformation of the motor domain, lmu, 2009.

[22] J.L. Sussman, D. Lin, J. Jiang, N.O. Manning, J. Prilusky, O. Ritter, E.E. Abola, Protein Data Bank (PDB): database of three-dimensional structural

information of biological macromolecules, Acta Crystallogr. D Biol. Crystallogr. 54 (1998) 1078–1084.
[23] R. Sayle, RasMol v 2.5: A Molecular Visualisation Program, Biomol. Struct. Glaxo Research and Development Greenford. Roger Sayle and Biomol. Struct. (1994).
[24] J. Cheng, A.Z. Randall, M.J. Sweredoski, P. Baldi, SCRATCH: a protein structure and structural feature prediction server, Nucleic Acids Res. 33 (2005) W72–W76.
[25] C. Lavedan, The synuclein family, Genome Res. 8 (1998) 871–880.
[26] M. Zhao, D. Cascio, M.R. Sawaya, D. Eisenberg, Structures of segments of alpha-synuclein fused to maltose-binding protein suggest intermediate states during amyloid formation, Protein Sci. 20 (2011) 996–1004.
[27] C.G. Dudzik, E.D. Walter, G.L. Millhauser, Coordination features and affinity of the Cu(2)+ site in the alpha-synuclein protein of Parkinson's disease, Biochemistry 50 (2011) 1771–1777.
[28] F. Wang, X. Feng, J. Ma, H. Zou, P. Chan, A common A340T variant in PINK1 gene associated with late-onset Parkinson's disease in Chinese, Neurosci. Lett. 410 (2006) 121–125.
[29] A. Surguchov, Molecular and cellular biology of synucleins, Int. Rev. Cell. Mol. Biol. 270 (2008) 225–317.
[30] D.F. Clayton, J.M. George, The synucleins: a family of proteins involved in synaptic function, plasticity, neurodegeneration and disease, Trends Neurosci. 21 (1998) 249–254.
[31] N. Ostrerova, L. Petrucelli, M. Farrer, N. Mehta, P. Choi, J. Hardy, B. Wolozin, Alpha-Synuclein shares physical and functional homology with 14–3-3 proteins, J. Neurosci. 19 (1999) 5782–5791.
[32] T.D. Kim, E. Choi, H. Rhim, S.R. Paik, C.H. Yang, Alpha-synuclein has structural and functional similarities to small heat shock proteins, Biochem. Biophys. Res. Commun. 324 (2004) 1352–1359.