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.