Studying the Effect of Alpha-Synuclein and Parkinson's Disease Linked Mutants on Inter Pathway Connectivities

Sagnik Sen

Jadavpur University

Ashmita Dey (✉ ashmitadey19@outlook.com)

Jadavpur University

Ujjwal Maulik

Jadavpur University

Research Article

Keywords: Parkinson's Disease , alpha-synuclein , aberrations

Posted Date: April 5th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-378378/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Scientific Reports on August 11th, 2021.
See the published version at <https://doi.org/10.1038/s41598-021-95889-5>.

Studying the Effect of Alpha-Synuclein and Parkinson's Disease Linked Mutants on Inter Pathway Connectivities

Sagnik Sen^{1,+}, Ashmita Dey^{1,+,*}, and Ujjwal Maulik¹

¹Computer Science and Engineering, Jadavpur University, Kolkata, India

*ashmitadey19@outlook.com

+these authors contributed equally to this work

ABSTRACT

Parkinson's Disease is a common neurodegenerative disease. The differential expression of alpha-synuclein within Lewy Bodies leads to this disease. Some missense mutations of alpha-synuclein may result in functional aberrations. In this study, our objective is to verify the functional adaptation due to early and late-onset mutation which can trigger or control the rate of alpha-synuclein aggregation. In this regard, we have proposed a computational model to study the difference and/or similarities among the Wild type alpha-synuclein and two mutations G51D and E46K which are responsible for slow and fast aggregation respectively. Evolutionary sequence space analysis is also performed in this experiment. Subsequently, a comparative study has been performed between structural information and sequence space outcomes. The study shows the structural variability among the selected subtypes. This information assists inter pathway modeling due to mutational aberrations. Based on the structural variability, we have identified the protein-protein interaction partners for each protein that helps to increase the robustness of the inter-pathway connectivity. As per the inter-pathway networks, drug addiction has clear impacts on both the mutations i.e., G51D-slow, and E46K-fast and can be considered as the reason for early-onset Parkinson's Disease. Finally, three top pathways associated with drug addiction viz., Amphetamine addiction, alcoholism, and Cocaine addiction show the higher influence in the mutated pathway networks based on the PageRank Algorithm where Dopaminergic Synapse system is also found within the same list in terms of Parkinson's Disease pathogenicity.

Introduction

One of the common neurodegenerative disorders (NDs) is Parkinson's disease (PD) which has been reported to be associated with the protein alpha-synuclein(Asyn)^{1,2}. Asyn_s are usually known for their robust expression within lewy bodies (LB). For example, the association between expression levels of the Asyn especially through the Lewy Bodies, and functional disorder of the dopaminergic neuron are reported³. Propensity of the Asyn towards aggregation might be responsible for functional disorder of its protein partners⁴. These genetic aberrations due to mutations can lead to functional modifications. For example, the EGFR1 signaling pathway is highly dependent on the activities of the asyn proteins. Any mutational modification can lead to functional changes in EGFR1 signaling pathways. In 1997, a point mutation at G209A has been reported. The substitution has resulted in point mutation A53T which is an early onset mutation linked with PD. However, this mutation is highly debatable as threonine can naturally be identified at rodent synuclein homologs, followed by two more mutations Viz., E46K, and A30P. The observation has justified the occurrence of functional similarity and pathway orchestrations associated with synuclein homologs and paralogs. Early-onset and late-onset mutations Viz., A53T, E46K, H50Q and G51D are reported in⁵. Usually, alpha-synuclein has three domains. Among them, A53T, E46K and H50Q are shown to enhance the aggregation propensity^{6,7} whereas G51D is associated with slower rate of aggregation. Similarly, non-beta component of Alzheimer's disease amyloid plaque (NAC) domain i.e., residue range 61-95 plays significant role during aggregation (residue range 61-95). On other hand, β - and γ -synucleins⁸ helps to inhibit the aggregation process Asyn. Impacts of such sequential changes are significant. For example, the amyloid formation mechanism of alpha-synuclein is mostly associated with a monomeric primary level of nucleation⁹. Usually the aggregation rate is too slow to detect. However, the phenotypical modifications due to mentioned mutations would have been affecting the residues at NAC domain (which is the prime point of aggregation). Therefore, intra-molecular allosteric effects would lead mutant to behave like ancestral phenotypes. However, these activities are directly affecting the functional profiles of the targeted proteins. Modifications, enhanced due to phenotypical changes, are correlated with the ontology terms. These are shared as common semantics among the pathways separately for wild type (WT) protein their mutants and closest neighbors from each of the cases. During the pathogenic progression, phenotypical modifications can explain the internal orchestration of the pathways based on the mutated candidate and their interacting neighbors. Mostly, these

connections are observed for the mutated candidates by following the ancestral protein neighbors. A point mutation at a certain residual position cannot change the whole structure. However, it can affect the co-evolving residues. Therefore, the relation among pathways is modified which may lead to the initiation of neurodegenerative progression.

In this study, one frame has been proposed by considering the sequential-structural modifications, and pathway semantics networks for Asyn and their mutants. The experiment has been initiated, focusing on one accelerating mutation and one slow mutation from the early onset of PD. Firstly, the sequential trait of the synuclein family is observed. Then evolutionary conserved co-evolving patches are studied by performing the Direct Coupling Analysis (DCA). Subsequently the information fetched from sequence space has been mapped to the structure. For two mutated cases, two distinct structures has been predicted applying *ab-initio* method. For the wild type protein, the ensemble model has been selected from the known structures. There after, the structure network has been formed based on normal mode scores of each residue. Comparing these information, the effective residues due to mutations have been fetched. Subsequently, BLAST helps to find the sequential similarity with ancestors from the same family. This may provide different lists of neighboring interaction partners for the wild type and mutant. Finally, weighted networks are formed based on pathway semantic relations. The pathways and their relations represent the nodes and the edges of the network. The weight of the edges are computed based on the semantic similarities.

Results

It is known that α -Synuclein is the candidate protein of Parkinson's disease. In order to understand the effect of this protein and its mutated samples towards the neurodegenerative diseases, a frame is proposed based on sequential and structural changes along with the pathway semantics.

Sequence Analysis

In sequence space, the family trait of the Synuclein family is analyzed in terms of order and disorder. In this regard, Shannon entropy is calculated for each protein belongs to this family, shown in Fig 1. Moreover, the entropic score of consensus sequence is 2.58. This suggests that the family trait of Synuclein family is ordered, whereas many proteins belong to this family are disorder in nature. This gives an in-depth idea of evolution.

Figure 1. The changing rate of Shannon entropic score of the protein members from synuclein protein family (Pfam id. PF01387)

To support the evolutionary changes Direct Coupling Analysis (DCA) is performed. The co-varying residue patches are represented in Fig 2(a). DCA score indicates the change in one residue is responsible for the change in other coupled residues. The result depicts that the residue changing for point mutation is also responsible for the change of its co-varying patches which eventually contributes to slow as well as accelerating aggregation of α -Synuclein. From the analysis, among five mutations E46K, H50Q and G51D mutated samples show conservation throughout the evolution. In parallel, the residue patches are clustered depending on their DCA score and residue in one cluster denotes same evolution rate. Interestingly, the mutated residues are present in one cluster shown in Fig 2(b), which indicates that the residue present in that cluster are varying accordingly.

Figure 2. DI scores based on top DCA contacts, (a) The co-evolved residue patches at NAC domain (b) A weighted network G_{DCA} and corresponding color modules based on over all residual co-variation from DI score

Selection of structure and analysis

α -Synuclein has multiple dynamic structure with frequent mutation. Due to this, an ensemble of all the structures is performed to understand the fluctuation rate, shown in Fig. 3. Moreover, a heatmap (Fig. 4) of all the structures also implies the minimum energy and fluctuation rate of structure ID 1XQ8. In this regard, we selected this PDB structure for the further analysis.

In order to have a comprehensive grasp of the changes in sequence space, the PDB structures of α -Synuclein and the mutated samples are constructed from the IntFOLD protein prediction server¹⁰. Based on the PDB models, structure networks are established. The structure network provides a new insight towards the nature and essential factors of structure-function dynamicity and folding process of a protein. During structure network analysis community detection and betweenness centrality of the protein, structures are executed.

The community detection of a protein is crucial to understand evolution. In this study, every residue pair, obtained from structure network (as shown in Fig 6 (a),(b) and (c)) belongs to the same cluster or in a different cluster with a strong association between them as shown in Fig 7 for WT-human, G51D-slow and E46K-fast. This result also supports the DCA score.

Furthermore, the residue present in modules are compared with cluster residue of Fig 2 (b) in which the mutated residues are available. The common residues present in both the cluster in the stretch of the NAC domain are mutated accordingly for three models in Fig 6. The stretch are further consider for performing BLAST. The highest similarity sequence is identified and applied for establishing a PPI network.

In Fig 8, the betweenness centrality is calculated to reveal the influence of a particular node on the internal dynamics of a different structure. Betweenness centrality is a measure based on shortest paths. The centrality measuring method denotes the degrees to which node stand between each-other. A node having higher centrality indicates better control over the network, this is because through the particular node more information will pass in the network.

In Fig 5A, the certain courses of van der Waals interaction are shown of the domain 40-95 where mutations mostly occur for three sequential models. Along with the non-covalent force the hydrophobic profiling is represented through a graph in Fig 5B.

The sequence with highest similarity of each sequential models after BLAST are used to understand the interaction with other protein by PPI network. From the BLAST result, it is found that mutation G51D-slow and E46K-fast have more similarity with Pan paniscus and Pan troglodytes respectively. With the cutoff, 0.5 top eleven including SNCA (α -Synuclein) are selected. The details list of associated proteins with their scoring depending on the shortest distance from SNCA are listed in Fig 9(a). Among the eleven associated proteins, a Venn diagram is performed. Fig 9(b) shows six proteins are common in three sequential models, which depicts that, these proteins remain connected through the evolution of α -Synuclein protein. it is concluded that protein may remain to interact but their connection scores change from one species to other.

Pathway semantic similarity

The six proteins are further studied in details to understand the role of pathways during evolution. From Reactome¹¹ and KEGG pathway databases¹² associated pathways of the six proteins are curated. We consider those pathways having more than two proteins including α -Synuclein. This is observed that every three combinations have the same pathways but the

Figure 3. Square Fluctuation Map on PDB Structures of Alpha Synuclein where each color represents one monomeric chain.

association of proteins of pathway differ from one other. The biological process is considered for each pathway and a semantic similarity graph is constructed shown in Fig 10. The red, blue and yellow nodes represent pathway Amyloid fiber formation, Parkin-Ubiquitin Proteasomal System pathway and Ectoderm Differentiation respectively. Moreover, the thick edges of the pathway graphs represent the connection weight between two pathways depending on the common biological terms.

An adjacency matrix of each type is obtained during pathway semantic calculation. This matrix is used as input for PR algorithm to rank the pathways according to their importance. From the resulted rank list we noticed that cocaine addiction, alcoholism and dopaminergic synapse secured a higher position in both G51D-slow and E46K-slow mutations. Whereas in E46K-slow mutation amphetamine addiction is also equally important according to the rank list. However in the WT-type parkin ubiquitin proteasomal system pathway ranked high which is absent in both the mutations. A barplot of each mutation of alpha-synuclein is reported in Supplementary Figure 1.

Discussion

The dynamic nature of the α -synuclein lies on the NAC domain. As this domain has a higher propensity of aggregation, it can be considered as the most unstable region. As discussed before, the β -synuclein and γ -synuclein proteins are paralogous to α -synuclein. However, α -synuclein has two more domains. SE scores on each member from the synuclein family indicate that overall trait of the family is highly ordered. In the case of α -synuclein, a certain range of disorder has been observed. This trait indicates the unstructured region of the α -synuclein. Furthermore, from coupling analysis, 291 potential coupling pairs have been found. Among them, very few have fallen under NAC domain range. However, the mutated samples, namely, G51D and E46K, are not known as part of the aggregation domain. So, the preliminary level of aggregations is somehow influenced by such mutations. Residual association within the protein is majorly controlled by the covalent and non-covalent interactions. In

Figure 4. A Heatmap to show fluctuating affinity among the monomeric PDB structures of the Alpha Synuclein

Figure 5. Comparative study on WT-human, G51D-slow and E46K-fast based on A. Variation in Van der waal's clashes and B. Hydrophobic Index at different residual points

folding purposes, non-covalent forces such as weak van der Waals and ionic interactions play a vital role. Also, the hydrophobic residual shields guard the hydrophilic residues against misfolding. Following this theory, the abruption at non-covalent forces

Figure 6. The structure network of (a) WT-human, (b) G51D-slow, (c) E46K-fast models

Figure 7. The community cluster modules of structures (a) WT-human, (b) G51D-slow, (c) E46K-fast after performing modularity detection

Figure 8. The betweenness centrality graph of model (a) WT-human, (b) G51D-slow, (c) E46K-fast

WT-human		G51D-slow		E46K-fast	
Associations	Score	Associations	Score	Associations	Score
STUB1	0.996	PINK1	0.933	SLC6A3	0.939
SNCAIP	0.992	TH	0.908	PARK7	0.909
SLC6A3	0.990	LRRK2	0.861	DYRK1A	0.847
UCHL1	0.985	MAPT	0.840	LRRK2	0.842
APP	0.983	APP	0.818	SNCAIP	0.808
LRRK2	0.971	DYRK1A	0.788	MAPT	0.749
HSPA4	0.970	SLC6A3	0.775	PINK1	0.724
FYN	0.968	PARK2	0.725	ATP13A2	0.696
PARK2	0.965	SNCAIP	0.717	TH	0.696
PARK7	0.960	PARK7	0.706	PARK2	0.693

Figure 9. The associating proteins of α -Synuclein for three sequential models are listed in (a) and in (b) the common proteins among them are shown through a Venn diagram

is not allowing the proper folding activity. In Fig 5A, the certain courses of non-covalent interaction have been shown at the affected targeted region at three sequential models. With that, hydrophobic profiling of the models are assist to describe the unstructured nature. In Fig 5B, the individual hydrophobic profiling of the WT and two mutants have been shown. Strikingly, the profiling of WT and E46k-fast are more similar. Therefore, the folding pattern of these two is supposed to be similar.

Figure 10. The pathway semantic graph of (a)WT-Human, (b) G51D-slow and (c) E46K-fast mutations of Alpha-synuclein. In the graph, the color nodes represent a particular pathway associated with the mutation type and the edges represents the weighted connection between the pathways. The bold edges indicate the higher association between those pathways.

On other hand, generic extensive effects of the co-evolution at a group of residues have been shown through network G_{DCA} (shown in Fig 2(a)). As the modules on the network are the result of betweenness centrality based clustering, it is expected that the members of the modules are completely or partially co-varied due to the mutation at residual points at G51D-slow and E46K-fast. Interestingly, the effects of the mutations are conserved in two colour modules mostly in terms of evolutionary information. Although it is generic information in terms of Synuclein family, it would provide clear evidence about the directly or partially affected residues sequentially. To make a comprehensive understanding, the sequence space information is mapped to individual structural information corresponding to the structure networks WT, G51D-slow and E46K-fast (Shown in Fig 6). From the colour modules of the G_{NMA} , it is observed that the distribution of the clusters is giving proper evidence of the aggregation. As each of the networks has the same tree cutter threshold, the distribution of the modules should be unbiased. Following this trait, it has been observed that G51D-slow has a higher propensity of aggregation than the other two. Mohite et

al.¹³ shows the strong effect of G51D phenotypical changes. Mapping the sequence space information to structure space, highly potential residue sets have been fetched. From Fig 2a, the coupling propensity of the NAC domain has shown. Interestingly, the distribution of co-evolutionary patches of the NAC domain is large. Comparing with individual structure networks, the individual lists of the affected residues due to an individual set of mutations are considered to be found the ancestral sequential traits. In the case of G51D-slow, the mutated sequence is more similar to Pan Paniscus. Similarly, the mutated sequence of E46K-fast is more similar to Pan Troglodytes. Eventually, the sequence traits are indicating the type of structural aberration from the WT. Subsequently, structural modifications in each case can provide a different list of top interaction partners. In Fig 9(a), the list of interaction partners for the individual types are given. The complete list of significant interaction partners is constant. Only the ranking of the partners is being changed due to structural modifications. However, common partners have been selected from the top list of the interaction partners. In Fig 9(b) six common partners have been taken from the list. In the case of non-pathogenic pathways, few of the pathways are usually performed in every eukaryotic cell. However, the activities of the pathways are revised based on their metabolome. In Fig 10, three networks based on semantic similarity have been shown. From the networks, all three common non-pathogenic pathways associated mostly with the common top neighbors of α -synuclein (with few protein sample specific neighbors) are semantically associated with each other. Namely BP has been considered for the similarity calculation. In WT-human, the non-pathogenic pathways have almost homogeneous association where the connection between blue and yellow node is slightly strong. In E46K-fast, this association between blue and yellow node is extremely strong. The activities of Parkin-Ubiquitin Proteasomal System pathway and Ectoderm Differentiation are highly influential in the activity of the Dopaminergic Neurons which is key point of PD initiations^{14,15}. However, strong association between the selected pathways has shown in semantic networks. This information is comparable with the module distribution of the G_{NMA} where the propensity of aggregation is highly strong in G51D-slow.

Aging is one of the prime reasons for losing the cognitive senses. WT type Asyn is usually coming up with cellular precipitation which leads to dementia. However, the early onset mutation can trigger the precipitation within 30-40 years of the age range. Two mutational subtypes i.e., G51D-slow and E46K-fast, are studied to unveil the functional and molecular orchestration during such mutations. The associated list of pathways individually for three samples has been considered. The pathways are connected through the sharing of interacting neighbors. Here, the list of protein interacting neighbors is modifying with mutations which brings us the distinct list of pathways as well as semantic connectivity. Six networks are based on two cognitive parameters i.e., Cellular component and, Biological process. From the networks, it is clearer that all the associated pathways are sharing almost the same subcellular localization. In WT type, the pathways excluding Parkinson's Disease and Alzheimer's Disease are associated with two distinct types of the function where some pathways are associated with prevention from unusual activity within the neurons whereas the rest of the pathways are associated with the pathogenic progression of the diseases. For example, monoamine transport systems are neurotransmitters that prevent the extracellular vesicles of the neuronal cells from excess dopamine, serotonin, noradrenaline, etc. Dopaminergic pathway, found in the list, is one of the prime pathways from the neuronal cells which helps to control the cognitive senses. Monoamine transport systems are the key modulator of the dopaminergic pathways. It can also control the effects of nanomolar elements such as cocaine, amphetamine, etc. On the other hand, pathways like Parkin-Ubiquitin Proteasomal System regulates the misfolding of the proteins which are further responsible for the pathogenic progression of the diseases such as PD. The pathway semantic networks (shown in Fig 11a) show the connectivity within the pathways based on sharing list of biological processes and cellular components. In the BP based network, the Parkin-Ubiquitin Proteasomal System pathway is the highest-ranked pathway as per PR outcomes. This shows the influence of the pathway in the network which increases the possibility of the Asyn misfolding. The pathways of the selected mutations Viz., G51D-slow and E46K-fast, are mostly associated with external nanomolar elements. More elaborately, dopaminergic pathways are strongly associated with the proteins which are differentially regulated due to cocaine addiction, alcoholism, amphetamine addiction¹⁶⁻¹⁸. The pathway semantic networks for E46K-fast shows the aforementioned strong connectivity. For both the mutation, it is clear that all the pathways are sharing almost the same cellular components. However, the amphetamine addiction pathway is disconnected in terms of BP for G51D-slow. Other than addiction dependent pathways, mitophagy is one more pathway that has been observed during both mutations. As per the previous researches, mitophagy elements propagate the possibility of the neuroinflammation and immune-precipitation of Asyn which further leads to PD¹⁹. Generally, the early onset of PD is strongly associated with drug addiction²⁰. Also, the mutations can inhibit the activities of the neurotransmitters such as monoamine transport systems²¹. Due to that reason, the excess hormones in the cellular systems cannot be removed. Outcomes of the PR algorithm shows affinity with the aforementioned analysis. As per the PR algorithm, the G51D-slow network has three equally influential pathways viz., cocaine addiction, Dopaminergic synapse, and Alcoholism whereas E46K-fast has four equally influential pathways viz., Cocaine addiction, Amphetamine addiction, Dopaminergic synapse, and Alcoholism. The algorithm supports the involvement of the addiction-related pathways at the early onset of PD. Therefore, any sort of addiction can increase the favorable condition for PD pathogenicity. Dopaminergic synapse plays important roles to control the cognitive senses. As mentioned earlier, the addiction pathways are highly connected with dopaminergic synapses. Therefore, dopaminergic pathways can be regulated by the aforementioned pathways.

Conclusion

From this study, the sequential variation of α -synuclein is observed depending on two mutational conditions, considering the structural consequences. The objective of the study is also to observe changes in non-pathogenic pathways due to structural aberration. From the outcomes, it is observed that G51D is a more influential early onset mutation for the pathogenic progression of PD. Similarly, three non-pathogenic pathways play vital roles in every possible mutational perspectives viz., Amyloid fiber formation, Parkin-Ubiquitin Proteasomal System pathway, and Ectoderm Differentiation. However, G51D mediated mutation disrupts the common trait of association among the mentioned pathways. Finally, the proposed frame can provide a comprehensive outlook on the mutation mediated structural aberrations and their affects on functional pathways.

Method

This study aims to understand the structural changes along with the pathway semantics of α -Synuclein concerning evolution. The detailed method is described in Fig 11.

Figure 11. The flowchart of the proposed framework

Sequence Space Analysis

Primarily, the Synuclein family is considered to unveil the family trait with the help of Shannon entropy (SE) calculation. It is evident that Shannon entropy is directly proportional to the rate of disorder. This implicates if the entropic score increases it signifies the higher disorderness of a protein. In²², 2.9 is considered as the threshold of Shannon entropy score, i.e. sequence with score less than 2.9 is treated as ordered sequence and vice-versa. This idea is utilized in this study and applied on each protein sequences of Synuclein family. Shannon entropy is calculated as follows:

$$Shannon_E(i) = - \sum_{i=1}^L H_i \log_2 H_i \quad (1)$$

where H_i was the probability of given amino acids and L was the number of letters in a sequence. The summation run over the 20 residues that normally were present in a protein sequence. The probability H_i represent the composition of the consensus sequence. So the entropy range lied between 0 and the $\log_2(20) = 4.32$.

Moreover, multiple sequence alignment is performed of the protein sequences belong to PF01387. It is noticed that multiple sequence alignment have a significant role in the field of structure and function analysis of biological sequences. Due to this, full alignment FASTA format sequences were collected, which comprises of tree orderings and all lower case letters including dashes to indicate the gaps. Additionally, a consensus sequence is generated from the aligned sequences. The consensus sequence is a set of amino acids with their occurrence frequency. These frequencies indicate the signature amino acids remain conserved throughout the evolution. In this regard, Shannon entropy is calculated for the consensus sequence also.

The aligned sequences are further considered for DCA. This is a statistical framework which possesses the idea of direct co-evolution coupling among the residue pairs. The drawback of the Mutual Information (MI) is, unable to extricate direct

correlation from indirect ones. This problem is easily solved by Direct Information (DI) theory. Depending on the DI theory, DCA is calculated. Hence, the computation of the DCA score of an aligned sequence indicates how directly the selected coinciding residues are coupled with each other and contributed toward evolution. The DI score is defined as:

$$DI_{s,t} = \sum XX' * F_{s,t}^{(dir)} * (X, X') * \ln \frac{F_{s,t}^{(dir)}}{F_s(X) F_t(X')} \quad (2)$$

Here, Here $F_{st}^{(dir)}$ represents reweighted frequency counts to introduce two residues for DI. where $F(X, X')$ is considered as joint probability, and $F(X)$ and $F(X')$ are individual probability. $F_s(X)$ and $F_t(X')$ are for amino acid type A at s^{th} position and similarly B at t^{th} position.

The cascading effect of the co-evolution is not conserved in a couple of residues. It is highly possible that mutational changes at a residual point can affect a distant residue by means of a cascading effect. To understand such an effect, a weighted network G_{DCA} has been defined considering DI score between residue. The weighted undirected network $G_{DCA} = (V_{res}, E_{DI})$ where V_{res} are set of nodes, consist of residue whereas E_{DI} are weighted edges, consists of DI score between the residues. Subsequently, the color modules have been formed applying Girvan-Newman algorithm²³.

Root Mean Square Fluctuation

The information related to sequence space or structure space individually is not enough to understand the changes. For the study, the WT type protein is taken from Protein Data Bank where the respective PDB id is 1XQ8. Due to the dynamic structure and frequent mutations, the structure possess lowest energy is considered. In this regard, Root Mean Square Fluctuation (RMSF) is performed to measure the particle deviation. In RMSF, a mean over time is considered for a residue r at the current position and some reference position. The definition of the RMSF is given in Equation 4.

$$RMSF_r = \left(\frac{1}{S} \sum_{t_n=1}^S mod(R_i(t_n) - R_i^{re})^2 \right)^{.05} \quad (3)$$

Where S is time over which the mean has been taken for reference position of the particle i , R_i^{re} . The RMSF has been observed based on the reference position of the particle i over time.0

Structure Network Analysis

To grasp this observation we have built the structure of those sequences. These complex models are analyzed with the help of a structure network²⁴. The interaction between the elements of the networks is represented through nodes and edges. Generally, a secondary structure and folding arrangement mechanism are used to understand the structure of a protein. Another promising methodology for understanding the structure is through network. The equation is represented below.

$$E_{ab} = \left[\frac{sc_{ab}}{\sqrt{(sc_a * sc_b)}} \right] * 100 \geq E_c \quad (4)$$

E_c is the threshold of interaction strength, the default value is 4%. Here, sc_{ab} was the number of side chain atom pairs of residues a and b . sc_a and X_b were the normalization factor for residues types a and b ²⁵.

In this paper, depending on the normal mode analysis (NMA) a correlation matrix is obtained. The matrix is applied to establish a full residue weighted network $G_{NMA} = (V_{res}, E_{NM})$ where V_{res} is a set of nodes representing residues whereas E_{NM} is set of weighted edges where weights are obtained from correlation matrix. Consequently, the network is split into a highly correlated coarse-grained community cluster network by using Girvan-Newman²³ clustering method where the highly interacting residues were clumped together in the clusters with a threshold value 0.7.

Pathway Semantic

Furthermore, mutation G51D and E46K i.e., slow and accelerating mutation on early onset are considered for further analysis. During mutation, co-varying residue cluster according to the DCA score are compared with the structure network module. The residues common in both the cases and also belong with the residue number 51 and 46 are mutated accordingly. As it is known that residue range 61-95 is the NAC domain of α -Synuclein, the residue stretch from 40-95 is utilized for performing Basic Local Alignment Tool (BLAST). The protein sequences are matched with the sequence database and statistical significance is calculated for the matched areas. The highest similarity sequence is selected and Protein-Protein Interaction (PPI)²⁶ is executed to detect the associated proteins. PPI is carried out for α -Synuclein of three species with a threshold value of 0.5. Among the associated proteins of WT-human, G51D-slow and E46K-fast, a Venn diagram is executed to find the common proteins. The pathways of the three sets are pondered for computing pathway semantics. Pathways responsible for at least two

proteins including α -Synuclein are considered. Here the semantic similarity is calculated over a set of a biological process responsible for the selected pathways. Wang method²⁷ is applied to evaluate the similarity established on graph-like structure of gene ontology (GO). The aggregated contribution is done by the semantic value of GO term T to the terms in DAG_T which is semantic of GO term T is firstly defined in Wang method. GO terms closer to T in DAG_T implies more contribution toward it semantics. Hence, it is defined that the contribution of GO term p to the semantics of T as the S – value of GO term p related to GO term T . For whatever term of p in DAG_T , the S – value associated GO term T , $S_T(p)$ is calculated as:

$$\begin{cases} S_T(T) = 1 \\ S_T(p) = \max(c_e * S_T(p') \mid p' \in \text{children of } (p)) \text{ if } p \neq T \end{cases} \quad (5)$$

Here C_e is defined as semantic contribution factor for the edge $e \in E_T$ linking GO term p with its child term p' . After calculating the S – value for the GO term in DAG_T , the semantic value of GO term T , $SV(T)$ is defined as:

$$SV(T) = \sum_{p \in X_T} S_T(p) \quad (6)$$

For two given GO term such as T and Q , semantic similarity is calculated between them is as follows:

$$SS_w(T, Q) = \frac{\sum_{p \in X_T \cap X_Q} S_T(p) + S_Q(p)}{SV(T) + SV(Q)} \quad (7)$$

Here $S_T(P)$ is the S – value of GO term t related to term T and $S_Q(p)$ is the S – value of GO term p related to term Q .

Moreover, based on the semantic similarity of GO terms, Best-Match Average (BMC)²⁸ strategy is performed to compute semantic similarity among sets of GO terms associated with the protein associated with a particular pathway and column, which defined as:

$$S_{BMA}(G1, G2) = \frac{\sum_{i=1}^i \max_{1 \leq n \leq j} S(go1_m, go2_n) + \sum_{j=1}^j \max_{1 \leq m \leq i} S(go1_m, go2_n)}{i + j} \quad (8)$$

where gene $G1$ annotated by GO terms sets $GO1 = (go1_1, go1_2 \dots go1_i)$ and $G2$ annotated by $GO2 = (go2_1, go2_2 \dots go2_j)$.

Ranking the pathways

The pathway semantic similarity graphs reveal the semantic strength between two pathway. In order to understand the importance of the pathways PageRank (PR)²⁹ Algorithm introduced by Google is applied on the resulted networks. This algorithm used probability distribution based on the weight among different nodes. In following equation 9, pathways are represented as nodes and the edges are weighted. The node rank has been defined as:

$$PR(n) = \sum_{m \in A_n} \frac{PR(m)}{E(m)} \quad (9)$$

where the rank of node n is relied on the PR values for each connected node $m \in A_n$, divided by $E(m)$, edges from node n . Therefore, according to the the PR values the pathways are ranked based on their importance in the network.

References

1. Pujols, J. & et.al. Small molecule inhibits α -synuclein aggregation, disrupts amyloid fibrils, and prevents degeneration of dopaminergic neurons. *Proc. Natl. Acad. Sci. United States Am.* **115** (2018).
2. Stefanis, L. α -Synuclein in Parkinson's Disease. *Cold Spring Harb. Perspectives Medicine* **2**, a009399 (2012).
3. Venda, L. L., Cragg, S. J., Buchman, V. L. & Martins, R. W. α -Synuclein and dopamine at the crossroads of Parkinson's disease. *Trends Neurosci.* **33**, 559–568 (2010).
4. Laura, M., Giuseppe, N., Lesley, C., Michele, V. & Annalisa, P. Functional interactions as a survival strategy against abnormal aggregation. *The FASEB J.* **25**, 45–54 (2010).
5. Ozansoy, M. & Basa, A. N. The Central Theme of Parkinson's Disease: α -Synuclein. *Mol. Neurobiol.* **47**, 460–465 (2013).
6. Flagmeier, P. *et al.* Mutations associated with familial Parkinson's disease alter the initiation and amplification steps of α -synuclein aggregation. *Proc. Natl. Acad. Sci. United States Am.* **113**, 10328–10333 (2016).

7. Sakai, R. *et al.* E46K mutant α -synuclein is more degradation resistant and exhibits greater toxic effects than wild-type α -synuclein in *Drosophila* models of Parkinson's disease. *PLoS ONE* **14**, e0218261 (2019).
8. Liberles, J. S. Evolution of Structurally Disordered Proteins Promotes Neostructuralization. *Mol. Biol. Evol.* **28**, 59–62 (2011).
9. Oliveira, G. A. P. & Silva, J. L. Alpha-synuclein stepwise aggregation reveals features of an early onset mutation in Parkinson's disease. *Commun. Biol.* **2** (2019).
10. McGuffin, L. J. & *et.al.* Intfold: an integrated server for modelling protein structures and functions from amino acid sequences. *Nucleic Acids Res.* **43** (2015).
11. Croft, D. & *et.al.* Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* **39** (2011).
12. Kanehisa, M. & Goto, S. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **28** (2000).
13. Mohite, G. M. & *et.al.* Parkinson's Disease Associated α -Synuclein Familial Mutants Promote Dopaminergic Neuronal Death in *Drosophila melanogaster*. *ACS Chem. Neurosci.* **9** (2018).
14. Lim, K. L. & Tan, J. M. Role of the ubiquitin proteasome system in Parkinson's disease. *BMC Biochem.* **8** (2007).
15. Jadalannagari, S. & *et. al.* Ectodermal Differentiation of Wharton's Jelly Mesenchymal Stem Cells for Tissue Engineering and Regenerative Medicine Applications. *Tissue Eng. Part B Rev.* **21** (2015).
16. Illes, A. *et al.* Open Access Dynamic interaction of genetic risk factors and cocaine abuse in the background of Parkinsonism—a case report. *BMC Neurol.* **19** (2019).
17. Bettiol, S. S., Rose, T. C., Hughes, C. J. & Smith, L. A. Alcohol Consumption and Parkinson's Disease Risk: A Review of Recent Findings. *J. Park. Dis.* **5** (2015).
18. Christine, C. W., Garwood, E. R., Schrock, L. E., Austin, D. E. & McCulloch, C. E. Parkinsonism in patients with a history of amphetamine exposure. *Mov. Disord.* **25** (2010).
19. Gao, F. *et al.* Mitophagy in Parkinson's Disease: Pathogenic and Therapeutic Implications. *Front. Neurol.* **8**, 1–6 (2017).
20. Dagher, A. & Robbins, T. W. Personality, Addiction, Dopamine: Insights from Parkinson's Disease. *Neuron* **61** (2009).
21. Oaks, A. W. & Sidhu, A. Synuclein Modulation of Monoamine Transporters. *FEBS Lett.* **585** (2011).
22. Romero, P. & *et.al.* Sequence complexity of disordered protein. *Proteins* **42** (2000).
23. Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. United States Am.* **99** (2002).
24. Greene, L. H. Protein structure networks. *Briefings Funct. Genomics* **11** (2012).
25. Brinda, K. V. & Vishveshwara, S. A network representation of protein structures: implications for protein stability. *Biophys. J.* **89** (2005).
26. Makino, T. & Gojobori, T. Evolution of protein-protein interaction network. *Genome Dyn.* **3** (2007).
27. Wang, J. Z. & *et.al.* A new method to measure the semantic similarity of GO terms. *Bioinformatics* **23** (2007).
28. Pesquita, C. & *et.al.* Metrics for GO based protein semantic similarity: a systematic evaluation. *BMC Bioinforma.* **9** (2008).
29. Page, L., Brin, S., Motwani, R. & Winograd, T. The pagerank citation ranking: Bringing order to the web. Tech. Rep., Stanford InfoLab (1999).

Acknowledgement

The work of SS and AD is supported by DST-INSPIRE fellowship.

Author contributions statement

S.S. and A.D. conceived and conducted the experiment. S.S., A.D., and U.M. analysed the results and drafted the manuscript. All authors reviewed the manuscript.

Competing interests

The authors declare that they have no competing interests.

Figures

Figure 1

The changing rate of Shannon entropic score of the protein members from synuclein protein family (Pfam id. PF01387)

(a)

(b)

Figure 2

DI scores based on top DCA contacts, (a) The co-evolved residue patches at NAC domain (b) A weighted network GDCA and corresponding color modules based on over all residual co-variation from DI score

Figure 3

Square Fluctuation Map on PDB Structures of Alpha Synuclein where each color represents one monomeric chain

Figure 4

A Heatmap to show fluctuating affinity among the monomeric PDB structures of the Alpha Synuclein

Figure 5

Comparative study on WT-human, G51D-slow and E46K-fast based on A. Variation in Vaan der waal's clashes and B. Hydropathic Index at different residual points

Figure 6

The structure network of (a) WT-human, (b) G51D-slow, (c) E46K-fast models

Figure 7

The community cluster modules of structures (a) WT-human, (b) G51D-slow, (c) E46K-fast after performing modularity detection

Figure 8

The betweenness centrality graph of model (a) WT-human, (b) G51D-slow, (c) E46K-fast

WT-human		G51D-slow		E46K-fast	
Associations	Score	Associations	Score	Associations	Score
STUB1	0.996	PINK1	0.933	SLC6A3	0.939
SNCAIP	0.992	TH	0.908	PARK7	0.909
SLC6A3	0.990	LRRK2	0.861	DYRK1A	0.847
UCHL1	0.985	MAPT	0.840	LRRK2	0.842
APP	0.983	APP	0.818	SNCAIP	0.808
LRRK2	0.971	DYRK1A	0.788	MAPT	0.749
HSPA4	0.970	SLC6A3	0.775	PINK1	0.724
FYN	0.968	PARK2	0.725	ATP13A2	0.696
PARK2	0.965	SNCAIP	0.717	TH	0.696
PARK7	0.960	PARK7	0.706	PARK2	0.693

Figure 9

The associating proteins of α -Synuclein for three sequential models are listed in (a) and in (b) the common proteins among them are shown through a Venn diagram

Figure 10

The pathway semantic graph of (a)WT-Human, (b) G51D-slow and (c) E46K-fast mutations of Alpha-synuclein. In the graph, the color nodes represent a particular pathway associated with the mutation type and the edges represents the weighted connection between the pathways. The bold edges indicate the higher association between those pathways.

Figure 11

The flowchart of the proposed framework

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supp1.pdf](#)