

RESEARCH ARTICLE

Did α -Synuclein and Glucocerebrosidase Coevolve? Implications for Parkinson's Disease

James M. Gruschus*

Laboratory of Structural Biophysics, NHLBI, NIH, Bethesda, Maryland, United States of America

* gruschuj@helix.nih.gov



OPEN ACCESS

Citation: Gruschus JM (2015) Did α -Synuclein and Glucocerebrosidase Coevolve? Implications for Parkinson's Disease. PLoS ONE 10(7): e0133863. doi:10.1371/journal.pone.0133863

Editor: Philipp J. Kahle, Hertie Institute for Clinical Brain Research and German Center for Neurodegenerative Diseases, GERMANY

Received: May 21, 2015

Accepted: July 3, 2015

Published: July 27, 2015

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by the Intramural Research Program at the National Institutes of Health, National Heart, Lung, and Blood Institute (NHLBI).

Competing Interests: The author has declared that no competing interests exist.

Abstract

Mutations in the *GBA1* gene are associated with increased risk of Parkinson's disease, and the protein produced by the gene, glucocerebrosidase, interacts with α -synuclein, the protein at the center of the disease etiology. One possibility is that the mutations disrupt a beneficial interaction between the proteins, and a beneficial interaction would imply that the proteins have coevolved. To explore this possibility, a correlated mutation analysis has been performed for all 72 vertebrate species where complete sequences of α -synuclein and glucocerebrosidase are known. The most highly correlated pair of residue variations is α -synuclein A53T and glucocerebrosidase G115E. Intriguingly, the A53T mutation is a Parkinson's disease risk factor in humans, suggesting the pathology associated with this mutation and interaction with glucocerebrosidase might be connected. Correlations with β -synuclein are also evaluated. To assess the impact of lowered species number on accuracy, intra and inter-chain correlations are also calculated for hemoglobin, using mutual information Z-value and direct coupling analyses.

Introduction

The synucleins are small (≤ 144 residues), membrane interacting proteins found predominantly in neurons, especially in their presynaptic regions, though small amounts have also been detected in non-neural tissue.[1] They are intrinsically disordered in solution and can adopt helical structure when bound to membranes. The most studied member of the synuclein family is α -synuclein (α -syn), due to its connection with Parkinson's disease (PD).[2–4] In PD α -syn forms amyloid, polymerizing into long fibrils via beta strand interactions between the monomers. These fibrils are the major component of Lewy bodies and neurites, proteinaceous inclusions seen in the neurons of Parkinson's disease patients.[5] As for its normal biological function, α -syn appears to be involved in the maintenance, exocytosis and regeneration of pre-synaptic vesicles,[6–10] interacting with the v-SNARE protein synaptobrevin-2 (also known as VAMP2).[11,12] Other potential interaction partners include Rab3a and Rab8b,[13,14] calmodulin,[15] dopamine transporter,[16] parkin, [17] phospholipase D1,[18] synphilin-1

(SNCAIP),[19] and tau.[20] In addition, α -syn interacts with the lysosomal enzyme glucocerebrosidase (GCase),[21] described below.

The synuclein family of proteins arose after the common ancestor of vertebrates diverged from other chordates. Typically, three genes for synuclein are present in each species. For example, humans have α , β and γ lampreys have three synucleins designated 3, DY and FD, and zebra fish have β , and two γ -synucleins, though many ray-finned fish also have a fourth synuclein with some α features.[22] Fig 1A shows a simplified evolutionary tree with the types of synuclein found in vertebrates. Most have α , β and γ types of synuclein. The exceptions are the jawless and ray-finned fishes, and curiously, monotreme and marsupial mammals, which appear to lack α -syn. Monotreme and marsupial genomes have not yet received the full extent of study that other genomes have, such as the zebra fish and human genomes. It might be that an α -syn gene does exist for non-placental mammals, but that it has somehow eluded discovery to date.

All synucleins bind membrane via their N-terminal regions, which include six (β) or seven (α and γ) imperfect amphipathic helix repeats (XKTKEGVXXXX) (Fig 1B). The α and β forms of synuclein share more sequence identity, being 64% identical in humans, while they both differ more from the γ form, with 49% identity between α or β and γ (Fig 1C). An important difference between the α and β forms is that α -syn has an additional 11 residue region that forms part of what is known as the non-amyloid- β component (NAC) region. The NAC region, residues 61–95, which was first detected as a component in amyloid beta plaques,[1] is part of the core beta sheet region in human α -syn amyloid fibrils.[23,24] In contrast, β -syn typically does not form fibrils and can even inhibit α -syn fibril formation.[25]

Mutations in several genes have been linked to increased risk for PD.[26] One of these genes is *GBA1*, the gene encoding glucocerebrosidase (GCase), a 60 kDa enzyme that cleaves

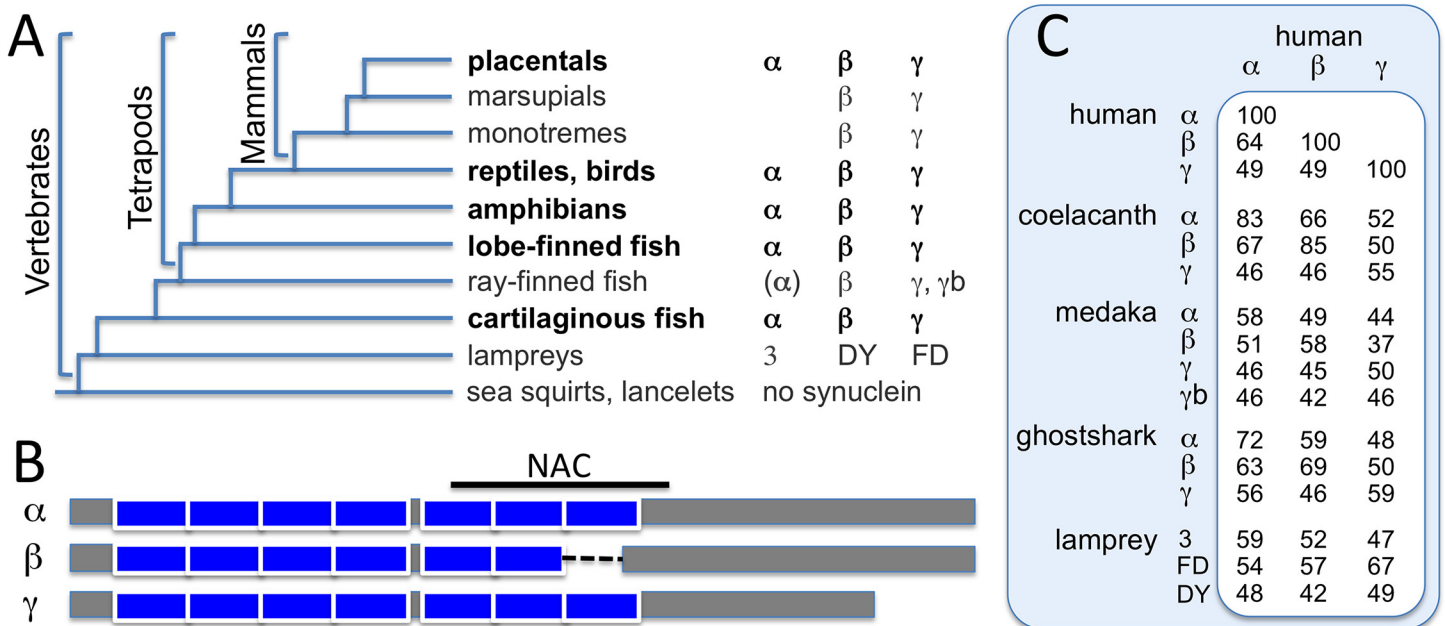


Fig 1. Synuclein evolutionary tree and sequence features. A) The types of synuclein proteins found in different branches of vertebrates are shown. Some ray-finned fish have a fourth synuclein that resembles α -syn in the NAC region. B) The sequence features of α -syn, β -syn and γ -syn are diagrammed. The boxes represent the imperfect amphipathic repeats, and the dashed line for β -syn indicates the gap in the NAC region, which is indicated by the line above α -syn. C) Table of sequence identities (%) comparing synucleins in tetrapods (human and coelacanth), ray-finned fish (medaka), cartilaginous fish (ghostshark) and jawless fish (lamprey).

doi:10.1371/journal.pone.0133863.g001

glucose from glucosyl-sphingolipids such as glucosylceramide. Inheritance of mutant *GBA1* from both parents results in Gaucher disease, the most common of the lysosomal storage diseases.[27] Carriers of *GBA1* mutations are found fivefold more frequently among PD patients than in non-PD controls,[28] and GCase has become an important therapeutic target for PD, as lowered activity has been linked to elevated α -syn levels.[29,30] The GCase enzyme is more ancient and widely expressed than α -syn. GCase is active in lysosomes, organelles found in nearly all types of animal cells, where proteins, nucleic acids, lipids and carbohydrates are broken down for reuse in the cell. The connection between GCase and α -syn might seem surprising, given that the substrate of GCase is a glycolipid, and that lysosomes and the presynaptic regions of neurons are distinct cellular milieus. Evidence suggests, however, that lysosomal degradation of α -syn plays an important role in maintaining normal α -syn levels in cells.[31] Intriguingly, α -syn interacts with wild type GCase under the acidic conditions found in lysosomes, and interaction is reduced for a common mutant form of GCase.[21] These observations hint that α -syn/GCase interaction might play some biological role that mutation of GCase disrupts.

In this study we explore the possibility that α -syn and GCase might have coevolved. That is, if during vertebrate evolution the interaction of α -syn and GCase acquired an important biological role, then subsequent mutations disrupting this interaction will have been selected against. Such selection should lead to two consequences. First, residues at the interaction interface should be more conserved. Second, if a mutation does occur at the interface, there is a chance that a corresponding mutation occurs in the interacting partner, one that either maintains the interaction or otherwise enhances fitness via a modified interaction. Of course, the first mutation must not affect fitness so much that it is selected out before the second mutation occurs. This means that when such correlated mutations do occur, their effect on interaction could be subtle when compared to other residue-residue interactions at the interface.

Several methods for identification of correlated mutations have been developed, reviewed in [32] with most used for *de novo* structure prediction,[33,34] though methods developed for inter-protein interaction are increasing in number.[35–38] The vetting process for these methods has often employed proteins for which 1000+ homolog sequences are known, and the importance of having very large numbers of sequences has been demonstrated repeatedly, with the minimum number of sequences for robust predictions estimated at 150 or higher, depending on protein length and the analysis employed.[32,39] However, this study analyzes the vertebrate specific protein α -syn, and only ~100 vertebrate genomes have been sequenced to date. To gain perspective on how effective these analyses are at identifying correlated mutations for vertebrate proteins and how the limitation imposed by lower number of sequences impacts the prediction, correlations between the alpha and beta chains of hemoglobin are also analyzed. [40] Hemoglobin is especially well suited for this analysis since its tetrameric structure has been confirmed by x-ray crystallography for 42 vertebrate species. Two of the most successful correlated mutation analyses are employed, direct coupling analysis (DCA) and mutual information (MI) Z-values,[41–44] and based on the results with hemoglobin, the MI Z-value approach has been chosen to identify putative correlated mutations in the 72 species where complete sequences of both α -syn and GCase are known.

Methods

Direct Coupling Analysis (DCA)

Direct coupling analysis was formulated to correct an issue encountered in other correlation analyses, namely, often ‘indirect’ correlations could obscure ‘direct’ correlations.[37] That is, in situations where A interacts with B and B interacts with C, sometimes the [A,C] correlation

coefficient could be larger than the [A,B] and [B,C] coefficients, due to extraneous factors such as differing residue mutation rates. To avoid this DCA takes all the single site and pairwise joint amino acid frequencies in the multiple sequence alignments (MSAs) of the proteins, combines them in a covariance matrix, and using a maximum entropy approach, corrects the joint frequencies in order to reduce the correlations of indirectly interacting pairs. Details of the derivation can be found in [41]. The correlation coefficients or “direct information” (DI) scores are given by the following mutual information formula:

$$DI_{ij} = \sum_{m,n} P_{ij}(m,n) \ln \left[\frac{P_{ij}(m,n)}{f_i(m)f_j(n)} \right]$$

where DI_{ij} is the DI score between residues positions i and j in the MSA, $P_{ij}(m,n)$ is the corrected joint frequency of amino acid m in position i and amino acid n in position j , and $f_i(m)$ and $f_j(n)$ are the amino acid frequencies at position i of amino acid m and position j of amino acid n , respectively, and the sum is taken over all amino acid types including gaps.

For the hemoglobin test case, the set of pairs considered to be in contact included all residue pairs with closest atom distance $\leq 7 \text{ \AA}$ in either oxy or deoxy forms. All sequences in this study were aligned using Clustal Omega.[45] All sequences were identified from blastp searches[46], repeated several times using the most divergent sequences in each set as query sequences, with incomplete sequences and any with unidentified residues eliminated from the set. For species with multiple entries for the same protein, the sequence with the most entries was chosen. The hemoglobin, α -syn, β -syn and GCase MSAs are provided in the Supporting Information (S1 Text). The DI values for hemoglobin were calculated with the MATLAB (The MathWorks, Inc.) script for DCA downloaded from dca.rice.edu. [41] For the intra-alpha chain DI calculation, only the alpha chain sequences were used, while for the inter-chain alpha-beta DI calculation, the input consisted of concatenated alpha and beta chain sequences, following protocol, [37] and the subset of DI values corresponding to inter-chain residue pairs used to assess the accuracy of predicted inter-chain contacts.

The DCA theta (Θ) and pseudocount_weight variables were optimized for both the intra and inter-chain calculations using the 314 species hemoglobin MSA. Varying the pseudocount_weight had little impact on accuracy, and the default value of 0.5 was used. Reducing theta improved both the intra and inter-chain prediction accuracies. The theta variable determines the sequence identity threshold in reweighting highly homologous sequences in the MSA.[41] For the intra-chain calculation, reducing theta from the default value of 0.2 to 0 resulted in the top-ranked correct contact prediction going from second at $\Theta = 0.2$ and 0.1 to first at $\Theta = 0$. The number of correctly predicted intra-chain contacts increased from 1 in the top ten and 20 in the top 100 ranked DI values for $\Theta = 0.2$, to 4 and 22 for $\Theta = 0.1$, to 6 and 24 at $\Theta = 0$. For the inter-chain calculation, the top-ranked correct contact prediction went from 24th at $\Theta = 0.2$ to second at $\Theta = 0.1$ and 0. The number of correctly predicted inter-chain contacts went from 0 in the top ten and 2 in the top 100 ranked DI values for $\Theta = 0.2$, to 1 and 1 for $\Theta = 0.1$, and to 1 and 3 for $\Theta = 0$. Because $\Theta = 0$ produced the most accurate predictions for both intra and inter chain contacts, this value was used to calculate the DI values in this study. A value of zero for theta means all hemoglobin sequences are weighted equally in the DI calculation.

Mutual Information (MI) Z-values

The expression for mutual information used in the MI Z-value analysis is the same as that used for the DI values in DCA, except the joint frequencies are not corrected for indirect correlations. The expression contains a correlation term $f_{ij}(m,n)/f_i(m)f_j(n)$ within an entropy-like

expression $f \ln[f]$. The entropy-like aspect of MI results in larger values when more pairs of correlated mutations occur, reflecting the likelihood that the correlation is not just due to chance. [39,43]

$$MI_{ij} = \sum_{m,n} f'_{ij}(m,n) \ln \left[\frac{f'_{ij}(m,n)}{f'_i(m)f'_j(n)} \right]$$

where $f'_{ij}(m,n)$ is the joint frequency that amino acid m occurs at position i and amino acid n occurs at position j in the same species, and $f'_i(m)$ and $f'_j(n)$ are the individual frequencies of amino acid m at position i and amino acid n at position j , and the sum is taken over all pairs of amino acid types including gaps. The frequencies include a pseudocount correction

$$f'_i(m) = \left(\frac{s}{s + 21\lambda} \right) \left(\frac{\lambda}{s} + f_i(m) \right)$$

$$f'_j(n) = \left(\frac{s}{s + 21\lambda} \right) \left(\frac{\lambda}{s} + f_j(n) \right)$$

$$f'_{ij}(m,n) = \left(\frac{s}{s + 42\lambda} \right) \left(\frac{\lambda}{s} f'_i(m) + \frac{\lambda}{s} f'_j(n) + f_{ij}(m,n) \right)$$

where λ is the pseudocount, s is the number of species in the sequence alignment, and $f_i(m)$ and $f_j(n)$ are the actual frequencies of amino acid m at position i in the first protein sequence and amino acid n at position j in the second protein sequence. That is, $f_i(m)$ equals the number of times amino acid m occurs at position i divided by the number of species, and similar for $f_j(n)$. Likewise, $f_{ij}(m,n)$ is the number of times the amino acid pair (m,n) occurs divided by the number of species. For the pseudocount, $\lambda = 1.5$ is used. A discussion of the importance of the pseudocount variable, especially in cases where sequencing errors might be present, is given in Appendix I in the Supporting Information (S2 Text). DCA also uses a pseudocount correction, but its implementation is more complex. [41]

MI Z-values, or Z-scores, tell one how many standard deviations the MI value for a particular pair of positions differ from the average MI for those positions with all other residue positions in the protein, and the Z-values are calculated as

$$Z_{ij} = 0.5[(MI_{ij} - MI_{iav})/\sigma_i + (MI_{ij} - MI_{jav})/\sigma_j]$$

where MI_{iav} is the average MI value between position i and all positions in the second protein sequence, MI_{jav} is the average MI value between position j and all positions in the first protein sequence, and σ_i and σ_j are the corresponding standard deviations. [42,43].

Results

Hemoglobin test case: intra *versus* inter-chain and the effect of reducing the number of sequences

Both DCA and MI Z-value analyses performed quite well at predicting intra-alpha chain contacts for the 314 species hemoglobin multiple sequence alignment (MSA) (Fig 2A and 2C). The DCA analysis performed best, with six of its top ten ranked DI values corresponding to contacting residues. Details on the optimization of the DCA parameters are given in the Methods. Both analyses gave the same number one ranked pair, the contacting residues 30 and 50. The situation is completely reversed for the inter alpha-beta chain contact prediction (Fig 2B and 2D), with both analyses performing more poorly, and with the MI Z-value analysis

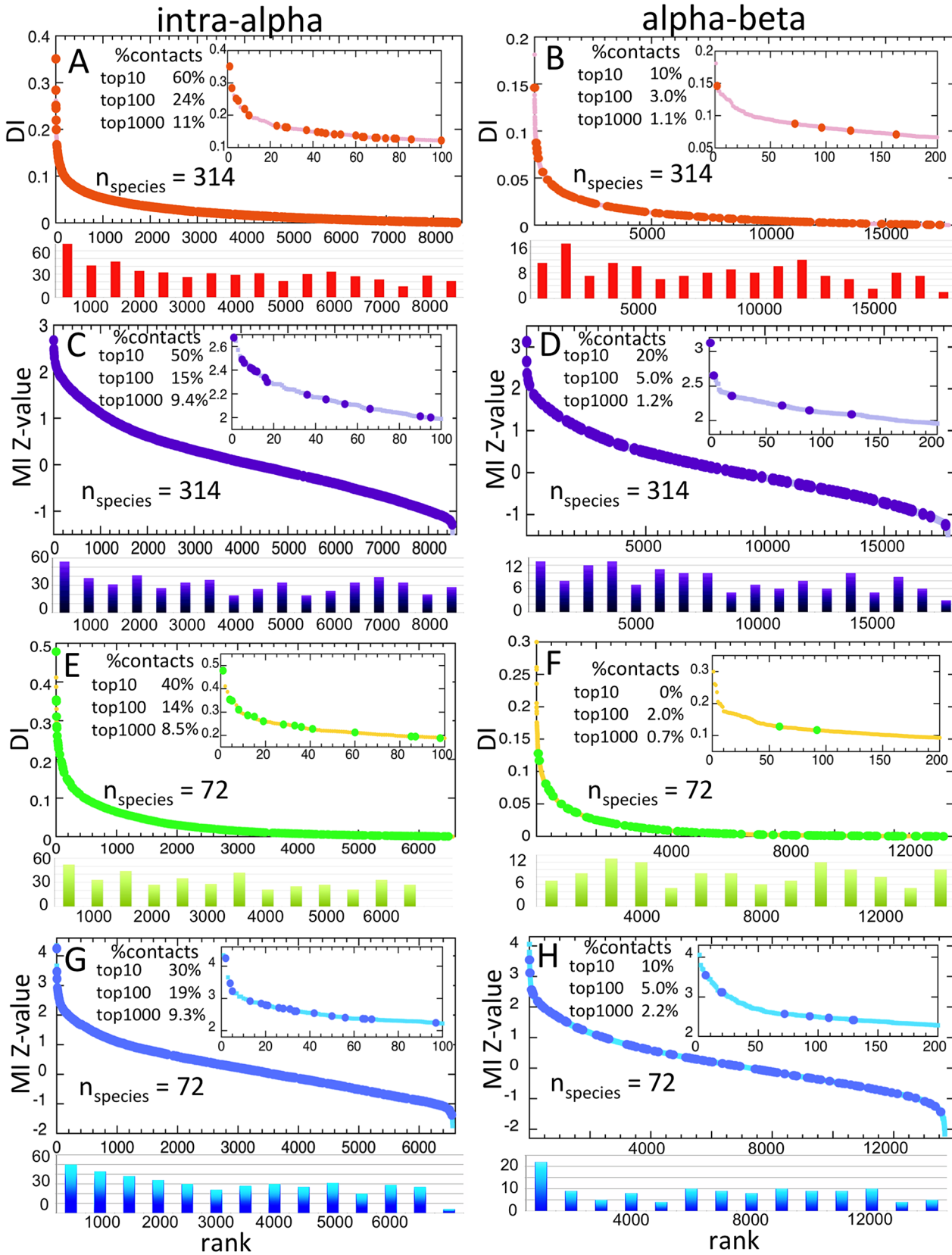


Fig 2. DCA and MI Z-value correlated mutation analyses of hemoglobin. A) Intra-chain and B) inter-chain DI values (y-axis) for the 314 species hemoglobin MSA, ranked largest to smallest on the x-axis, between all non-invariant residues (light red dots), with those corresponding to contacts shown as orange circles. The inset shows an expanded view of the top 100 (intra-chain) or top 200 (inter-chain) results. The percentages of correctly predicted contacts for the top 10, 100, and 1000 results are shown. Because of the large number of contacts, histograms are shown below, with the number of contacts binned every 500 DI values (intra-chain) or 1000 DI values (inter-chain). C) Intra-chain and D) inter-chain MI Z-values for the 314 species hemoglobin MSA, shown similarly to A) and B). E) Intra-chain and F) inter-chain DI values for the 72 species hemoglobin MSA, corresponding most closely with species for which complete α -syn and GCCase sequences are known, shown similarly to A) and B). G) Intra-chain and H) inter-chain MI Z-values for the 72 species hemoglobin MSA, shown similarly to A) and B). Note that for the 72 species hemoglobin MSA, there are fewer non-invariant residues, and thus fewer potential intra- and inter-chain pairs along the x-axis.

doi:10.1371/journal.pone.0133863.g002

outperforming DCA, with two of the top ten Z-values corresponding to contacting residues. In this case, the number one Z-value ranked pair, corresponding to contacting residues 111 and 115 in the alpha and beta chains, differed from the top correct DI pair, ranked at number two, corresponding to residues 120 and 33. Mathematically, however, one expects the inter-chain case to be more challenging; hemoglobin has 537 intra-chain *versus* 149 inter-chain residue contacts but 8515 intra-chain *versus* 17685 inter-chain potential residue pairs. Viewed in this light, the correct prediction by the MI Z-value analysis of 20% inter-chain contacts is more encouraging. In fact, the 20% correct prediction is comparable to that obtained by intra-protein methods when one restricts the set of residue pairs to include only long-range contacts (> 24 residues separation in sequence).[47]

To test the effect of reducing the number of sequences, the DCA and MI Z-value analyses were repeated using a subset of 72 of the 314 hemoglobin proteins corresponding most closely with the 72 species for which complete α -syn and GCCase sequences are known (Fig 2E and 2H). For the intra-chain case, both analyses performed more poorly than with the 314 species MSA, but still performed moderately well, with DCA again outperforming the MI Z-value analysis with 4 of the top ten DI values corresponding to contacts. Interestingly, the number of correct predictions in the top 100 Z-values actually increased, however. Both analyses also performed more poorly for the inter-chain calculation, especially DCA. The MI Z-value analysis still gave one correctly predicted pair in the top ten, and the same number in the top 100 as for the 314 species case. The spread of DI and Z-values increased significantly for the 72 species MSA, due to the larger percentage of invariant residues and the greater effect of statistical fluctuations in the smaller data set. Notably for the smaller species set, the DI top correctly predicted inter-chain pair is the same as the Z-value top pair, alpha 111 with beta 115. The possibility that this could happen by chance is very low. The results show that while it is still possible to obtain correctly predicted correlated residue pairs with fewer sequences, increased “noise” can potentially reduce the number of highest-ranked, correctly predicted pairs. Given its better performance in predicting inter-chain hemoglobin contacts, the MI Z-value analysis has been chosen as the main method for the correlated mutation analysis of α -syn and GCCase. Still, one should not necessarily generalize, and the implications of the intra and inter-chain results are addressed further in the Discussion.

Correlated mutation analysis of α -syn and GCCase

An MI Z-value correlated mutation analysis was performed using sequences from 72 vertebrate species where complete sequences of both α -syn and GCCase have been reported. The α -syn-like sequences found in some ray-finned fish were not included; these sequences lack the conserved C-terminal region known to interact with GCCase.[21] Fig 3 shows the top ten MI Z-value ranked correlated pairs. The choice of ten is somewhat arbitrary, and views differ on the number of top-ranked predicted pairs that should be considered as candidates for correlated residues.[39,48] The analysis of hemoglobin shows the odds of actual contacts beyond the top 10 ranked pairs falls rapidly, from 50% for the top ten intra-alpha pairs dropping to 15% for

the top 100, and to 9% for the top 1000 (Fig 2C). The trend is even more pronounced for the alpha-beta inter-chain results, with 20% for the top ten, 5% for the top 100, and 1.2% for the top 1000 (Fig 2D).

The top ten MI Z-values for α -syn and GCase (5.0–9.6) are much higher than those obtained for the alpha and beta hemoglobin correlated mutation analysis of the 72 species subset (3.4–4.1). However, this does not mean the α -syn, GCase correlations are “stronger” than the hemoglobin ones. The higher Z-values are a consequence of the greater sequence conservation of α -syn and GCase compared to hemoglobin. Human and coelacanth α -syn sequences are 81% identical, 63% identical for GCase, but only 51% identical for hemoglobin. MI Z-values show how many standard deviations MI values differ from average, and the more conserved the set of sequences is, the more that correlated residue pairs can potentially differ from the baseline average. The actual MI values are more similar, 0.57–0.71 versus 0.58–0.67, for the top ten α -syn/GCase and alpha/beta chain Z-value ranked pairs, respectively. Correlations were also calculated using the DCA method. Consistent with the hemoglobin results, several top ten Z-value α -syn, GCase pairs are amongst the top 100, but not top 10 DCA pairs. To be precise, Z-value ranked pairs #1, #2, #3, #5, #6, and #9 correspond to DCA ranked pairs #43, #46, #20, #32, #51, and #81, respectively.

#1 [α -Syn 53, GCase 115]. The top MI Z-value ranked pair is α -syn residue 53 with GCase residue 115. This pair is intriguing because the A53T and A53E mutations in humans are associated with PD.[49,50] Even more curious is the fact that except for apes, Old World monkeys and one New World monkey (squirrel monkey, *S. boliviensis*), nearly all other species have threonine at position 53, the same amino acid as the PD-associated mutation A53T. Only the Australian ghost shark *C. milii* has a different amino acid, serine. Unfortunately, GCase residue 115 of *C. milii* is unknown; its sequence is known for only residues 231–497. The Z-value of 9.6 for this pair is quite high, much higher than the highest values seen for hemoglobin. The value is high because this pattern of residue variation, seen just for apes and Old World monkeys, is not seen for any other residues in α -syn or GCase. Intra-protein Z-value and DCA correlations for α -syn and GCase were also examined, confirming that the pattern is unique.

To test whether the [α -syn 53, GCase 115] Z-value is influenced by over-representation of Old World monkey and ape sequences, the calculation was redone with the closest human relatives removed one by one. Removing chimpanzee (*P. troglodytes*), orangutan (*P. abelii*) and gibbon (*N. leucogenys*) sequences had a negligible effect on the Z-value. Only after also removing baboon (*P. anubis*) and rhesus macaque (*M. mulatta*), leaving just human and green monkey (*C. sabaeus*), did the Z-value fall to number two in the rankings. Thus, the high Z-value does not appear to be significantly inflated by over-representation of Old World monkey and ape sequences.

Fig 4A shows the locations of the top ten ranked GCase residues in the crystal structure of human GCase (1OGS).[51] Interestingly, G115 is not exposed in the structure, though it is near the surface, and its alpha carbon is situated such that a glutamate substituted at this position would almost certainly be surface-exposed. Since G115 is buried for human GCase, this residue cannot directly contact α -syn residue A53 in humans unless there is a change in GCase structure upon interaction. However, while residue contacts are used as a proxy to test how well different analyses identify correlated mutation pairs, it is not necessarily the case that all truly correlated pairs are in direct contact. For instance, residues could be in contact in some species with a particular pair of amino acids and not in others with a different pair. Alternatively, a correlated pair might not be in contact in any species, but instead allosterically influence protein interaction.[52,53] Thus, many scenarios leading to the [α -syn 53, GCase 115] correlated pair are possible.

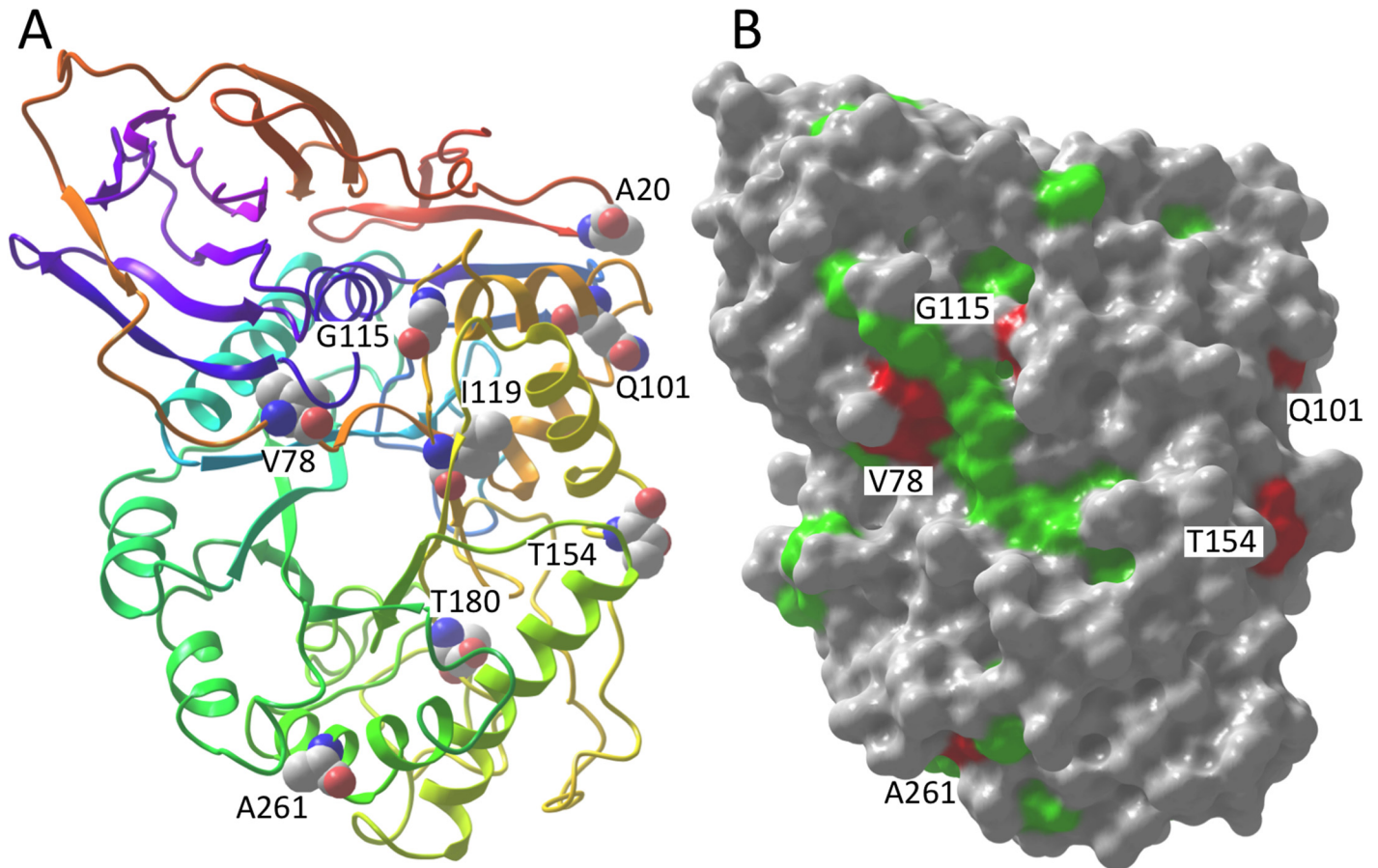


Fig 4. GCCase residues for the top ranked correlated pairs and conserved surface regions A) A ribbon structure of human GCCase (pdb 1OGS) with the residues from the top correlated pairs shown. B) The surface corresponding to invariant GCCase residues is shown in green, with the surfaces of selected correlated pair residues in red and labeled. A large patch of invariant surface lies between V78 and G115.

doi:10.1371/journal.pone.0133863.g004

It is also possible that the [α -syn 53, GCCase 115] pair do not interact, neither directly nor allosterically, and that the correlation arose solely by chance. Since only one correlated mutation is seen for this pair, the odds that the pair arose by chance are not insignificant. An estimation of the probability in a simplified test case is given in Appendix II of the Supporting Information (S3 Text), yielding roughly a 1 in 3 chance that such a correlation might have arisen between non-interacting residues. This sort of chance correlation is an aspect of what is termed phylogenetic noise or bias, [54,55] and since it may be relevant to the putative [α -syn 53, GCCase 115] correlated pair, a bit more explanation is warranted here.

In general, the MI Z-value approach removes much of the influence of phylogenetic bias. In particular, the more common a pattern of sequence variation, due to shared ancestry for instance, the lower the corresponding Z-values for residue pairs displaying the common pattern. [56] For example, differences in the α -syn sequences between mammals and non-mammals are fairly common, likewise for GCCase, so residue pairs displaying a mammal/non-mammal sequence difference pattern score lower Z-values. On the other hand, if the pattern is rare, as is the case with [α -syn 53, GCCase 115], the Z-value will be high. Therefore, additional criteria are needed to identify which potential correlated pairs are less likely to have arisen by chance.

One common criterion to distinguish pairs that are less likely to be due to chance is that the correlated mutation pair occurs in at least two non-sister groups, that is, the mutation pair must have arisen at least twice independently.[57,58] The [α -syn 53, GCase 115] pair fails this requirement since the correlated pair of mutations occurs only in the Old World monkeys and apes clade. More sequences from cartilaginous fish could help to make a stronger case for correlation, since the one currently known, that of the *C. milii*, is the only α -syn sequence with neither alanine nor threonine in position 53. As for GCase 115, the African clawed frog (*X. laevis*) has the GCase E115G substitution, like apes and Old World monkeys, but it retains α -syn T53, and thus does not provide additional support for a true correlation. Therefore, despite α -syn 53 and GCase 115 being the top-ranked correlated pair, it cannot be ruled out that the correlation might have arisen by chance.

#2 [α -Syn 124, GCase 78]. The second highest MI Z-value ranked pair is α -syn 124 with GCase 78. At first glance, this potential correlated pair is not very compelling. Specifically, the mutations for GCase 78 are all conservative substitutions, valine, isoleucine, leucine and methionine. Furthermore, the side chain of residue 78 is buried in the GCase structure, with only the backbone exposed (Fig 4). On the other hand, this residue precedes K79, a surface-exposed, invariant residue, and mutations of V78 and K79 are associated with Gaucher disease.[27,59] Even though correlated mutation analyses cannot directly identify contacts involving invariant residues, correlations involving neighboring non-invariant residues could still allow one to deduce their presence. Note that α -syn 124 lies between two invariant glutamates, E123 and E126 (Fig 5), which could potentially form a salt-bridge with GCase K79.

The α -syn A124T substitution is only present when GCase 78 is isoleucine. There are more occurrences of isoleucine for GCase 78 than threonine for α -syn 124, suggesting that the α -syn A124T substitution could be accommodated only after GCase I78 was already present. Valine is the most common residue for GCase 78, so perhaps the size difference between isoleucine and valine at residue 78 perturbs the positioning of K79 in a way that alters its interactions with α -syn residue 124 and its neighboring glutamates. The threonine, isoleucine correlated mutation occurs not just in one clade, as was the case with [α -syn 53, GCase 115], but in three: bats, afrotheres and amphibians. None of these three clades are sister groups. Therefore, the chance that the same random mutations occurred in these unrelated groups is quite small, suggesting that the [α -syn 124, GCase 78] correlation arises from a real interaction.

#3, #5, #6 & #9 α -Syn 68 & 95 with GCase 101 & 154. The third highest MI Z-value ranked pair belongs to a group of four of the top ten pairs, between α -syn residues 68 and 95 and GCase residues 101 and 154. Groups of correlated pairs are not unusual; in fact, they are expected. While the aim of the DCA and related methods has been to disentangle direct from indirect correlations, one quite successful correlated mutation analysis, SCA, capitalizes on indirect correlations, identifying correlated “sectors” of proteins, that is, sets of residues that are inter-correlated.[48] The α -syn 68 & 95 and GCase 101 & 154 pairs also rank in the top ten intra-protein correlations in both the DCA and Z-value analyses. In the structure of human GCase, residues 101 and 154 are separated by 11 Å, both surface-exposed with most of the space between them unoccupied (see Fig 4A). Thus, it is plausible that either α -syn residue 68 or 95 or both might lie between the two GCase residues. In addition, the T154 side chain is hydrogen-bonded to the backbone at the beginning of a helix, thus a T154V mutation could cause a significant local structural change.

For the number three ranked pair, [α -syn 68, GCase 101], the putative correlated mutation involves an α -syn G68E mutation in Laurasiatheres, Afrotheres and Xenarthrans, and a GCase Q101R mutation in Laurasiatheres, suggesting the possible evolution of a salt bridge. One problem with this particular correlated pair is that it fails the phylogenetic bias test. For instance, a single G68E mutation in α -syn in the common ancestor of placental mammals,

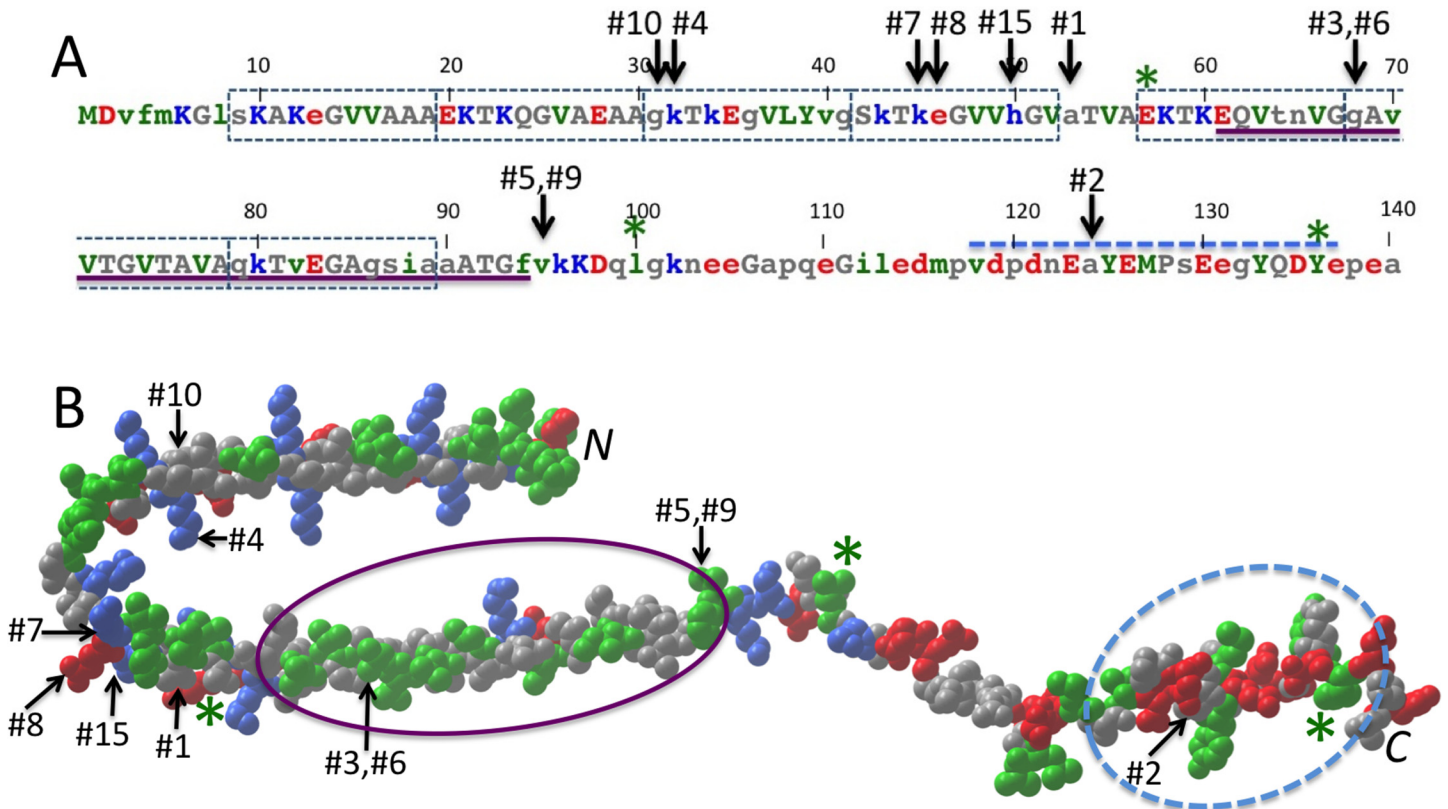


Fig 5. Sequence and structure of α -syn showing top correlated residues and regions known to interact with GCCase A) The α -syn residues of the top ranked correlated pairs with GCCase are indicated. Invariant residues are capitalized, negatively charged residues in red, positively charged in blue, hydrophobic in green, all others gray. The seven imperfect repeat regions are indicated by the dashed boxes and the NAC region is underlined (purple). In solution only residues 118–137 (dashed blue line) interact with GCCase. In the presence of lipid vesicles, fluorescent labels at residues 57, 100, and 136 (green asterisks) showed interaction with GCCase, indicating a much larger region of interaction than in solution. B) The α -syn structure with residues of the top ranked correlated pairs indicated. The residue coloring is the same as for the α -syn sequence. The NAC region is indicated by the purple oval, and the C-terminal region that interacts with GCCase in solution is indicated by the dashed blue oval. The green asterisks indicate the locations at residues 57, 100, and 136 where fluorescent labels showed interaction with GCCase in the presence of lipid vesicles. The structure shown is micelle bound α -syn (pdb 1XQ8).

doi:10.1371/journal.pone.0133863.g005

which happened to later revert back to glycine for most Supraprimates, plus a single Q101R mutation in GCCase in the common ancestor of Laurasiatheres is all that is needed. Hence, there is a significant probability that [α -syn 68, GCCase 101] could be correlated by chance.

The number five ranked pair, [α -syn 95, GCCase 101] also fails the phylogenetic bias test, but the number six and nine ranked pairs, [α -syn 68, GCCase 154] and [α -syn 95, GCCase 154] have an additional mutation that suggests the correlation might not be due to chance. One Afrothere, the elephant shrew (*E. edwardii*), has the same GCCase T154V mutation as seen for Laurasiatheres. The other three GCCase sequences for Afrotheres have T154, suggesting the T154V mutation in *E. edwardii* might have arisen after the other mammal clades split from Afrotheres. In this case, elephant shrews and the Laurasiatheres would not be sister groups, suggesting the correlations with T154 might be real.

#8 and #15 [α -syn 46, GCCase 180] and [α -syn 50, GCCase 261]. The number eight and fifteen ranked pairs include residues corresponding to the PD-associated α -syn mutations E46K and H50Q. Also, the number seven ranked pair [α -syn 45, GCCase 180] shows a similar substitution pattern as α -syn 46. The seven and eight ranked pairs show correlated mutations occurring for one amphibian (*A. mexicanum*), one turtle (*C. mydas*) and one lizard (*A. carolinensis*), but not in the other amphibian and reptile sequences, so it is likely the correlated

mutations arose independently. However, the mutations are conservative, α -syn K45R, α -syn E46D, and GCase T180S. Moreover, GCase residue 180 is completely buried, so if these pairs are truly correlated, the interaction would have to be allosteric. The number fifteen ranked pair also passes the phylogenetic bias test, in this case the H50Q, A261V variation seen for penguin (*A. forsteri*) and coelacanth (*L. chalumnae*), and additionally H50Q, A261I seen for anole (*A. carolinensis*). Curiously, as with the A53T mutation, the PD associated H50Q mutation mirrors the substitution seen in the different species. The two other PD-related α -syn mutations, A30P and G51D, occur in invariant positions in the species studied, so the analysis cannot address them. Finally, the number four and ten ranked pairs involve variations only in bovines and cetaceans, respectively, so they fail the phylogenetic bias test.

Conserved surface regions of α -syn and GCase

Protein interaction interfaces are known to contain more conserved residues than non-interacting regions of protein surfaces,[60] hence α -syn and GCase were searched for conserved regions. α -Syn is intrinsically disordered in solution, and while transient intra-protein contacts occur,[61] all its residues are potential candidates for contacts with binding partners. The most conserved regions encompass residues 6–85, containing the seven imperfect amphipathic repeats, and residues 123–136, the region that interacts with GCase in solution (Fig 5).[21,62] All but the number five and nine ranked correlated pairs have α -syn residues that lie in these conserved regions.

In contrast to α -syn, GCase is a folded protein with a well-defined surface that can be searched for conserved regions. The largest region lies in and around the enzyme active site, with invariant residues comprising an area of roughly 300 Å² of the surface. The second largest area, covering roughly 200 Å², lies on the opposite side of GCase and includes invariant residues 79, 227, 228, 429, 452, and 454. This second region is shown in Fig 4B, along with the surface regions of GCase residues from selected top Z-value ranked pairs. The region lies between GCase residues V78 and G115, from the number one and two ranked correlated pairs, supporting the hypothesis that α -syn has coevolved with GCase to maintain contact with this region.

Comparison with β -syn

The closest homolog of α -syn is β -syn, and correlations between them and with GCase were also examined by MI Z-value analysis in the 55 species where complete sequences of all three are known. α -Syn and β -syn are known to interact *in vitro* and *vivo* [25], and both are present in high amounts in the presynaptic spaces of central nervous system neurons. β -Syn can also interact with GCase, inhibiting its activity *in vitro*, though 4-fold less so than α -syn.[63] Fig 6 shows the top five ranked MI Z-values for the α -syn, β -syn and GCase analyses. Interestingly, the most highly correlated α -syn, β -syn pairs involve α -syn 53, just as with α -syn and GCase. The top two pairs are α -syn 53 with β -syn residues 86 and 121, with Z-values of 9.7 and 9.0, respectively, though keep in mind that the smaller data set, 55 vs. 72 species, is expected to yield larger Z-values. In particular, residue 86 involves a Lys to Arg substitution in apes, Old and New World monkeys, rabbit (*O. cuniculus*) and horse (*E. caballus*), that is, the same species as for the α -syn A53T substitution (Fig 3) plus marmoset (*C. jacchus*), rabbit and horse. β -Syn residue 121 involves a Glu to Asp substitution in apes, Old and New World monkeys, mouse (*M. musculus*), rat (*R. norvegicus*), and bat (*E. fuscus*). Recall that the top ranked α -syn, GCase pair showed the GCase E115G substitution only in apes, Old World monkeys and frog (*X. laevis*). Because the β -syn residues have even more substitutions than α -syn 53, they are not as strongly correlated with GCase 115, with Z-values of 7.6 and 7.4 for β -syn 86 and 121, respectively. As a final note, like α -syn 53, the most common residue for β -syn 53 is alanine, though

analysis of metazoan G-protein coupled receptors, where using MIP, a method similar to MI Z-values, a drop in accuracy of 52% to 28% was seen analyzing alignments consisting of 283 and 107 species.[64]

The DCA and MI Z-value methods differ in their formulation and implementation, but both utilize mutual information terms, and both are “global” correlation methods, that is, the final residue pair correlations are the result of a calculation influenced by all the other residue pairs. In DCA this occurs in the initial stage, when the joint amino acid frequencies are adjusted so that DI values exclude contributions from indirect correlations.[37] In the MI Z-value method, the globalization occurs in the final step, when the residue pair MI Z-value is calculated by how many standard deviations its MI value deviates from the MI values of other pairs with one residue in common. One difference that might influence to the inter-chain result is that inter-chain DI values are calculated in the context of both intra and inter-chain pairs, while MI Z-values are calculated with inter-chain pairs only.

Of course, there are many other methods of correlated mutation analysis, and hemoglobin is just one complex. Any generalization of these results should await tests with additional methods and complexes, and factors such as the impact of interacting partners with differing mutational rates must be assessed. Hemoglobin was chosen since as a protein present in vertebrates, the species list could be reduced to closely match those in the α -syn and GCCase sequence alignments. While hemoglobin may be unique among vertebrate protein complexes in the number sequences known and structures solved,[40] there are many bacterial protein complexes where 1000+ sequences and multiple structures are known. DCA, the related evolutionary couplings (EVCouplings), and other methods have been successfully applied to these bacterial systems.[35,37,38]

Comparison with experimental data on α -syn/GCCase interaction

Several experiments have shown a physical interaction between α -syn and GCCase. Their interaction has been measured both in human tissue homogenates and *in vitro*. GCCase has been co-immunoprecipitated with α -syn in lysates prepared from brain tissue, and they co-localize in neuroblastoma cells that over-express both proteins.[21] *In vitro*, residues 118–137 in the C-terminal region of α -syn contact GCCase in solution,[21] and when the proteins are membrane-bound, a much larger α -syn region, including residues 57, 100 and 136, interacts with GCCase.[63] Fig 5 shows the α -syn sequence with the regions that interact with GCCase indicated. In solution, only the number two ranked correlated pair [α -syn 124, GCCase 78] involves an α -syn residue in the observed interacting region. However, for α -syn interacting with GCCase in the presence of lipid vesicles, eight of the top ten correlated pairs (#1 –#3, #5 –#9) involve α -syn residues near ones observed to interact.

While the precise location on GCCase where α -syn interacts is not known, α -syn binds more weakly to GCCase with the common N370S mutation.[21] The weaker binding was measured in the absence of lipid, so the interaction must involve C-terminal α -syn residues 118–137. Residue 370 is not solvent exposed, but based on crystal structures, it is thought to influence the surrounding structure, including a loop near the active site.[65] Of the GCCase residues in the top ten Z-value ranked pairs, the one closest to N370 is V78 of the [α -syn 124, GCCase 78] pair, with a distance of 10 Å. Taken together, the weaker binding to the N370S mutant and the correlated mutation analysis are consistent with the α -syn C-terminal region interacting with a GCCase surface region that includes V78.

Biological implications

The most interesting pairs of residues identified by the correlated mutation analysis are the number one, eight and fifteen ranked pairs, [α -syn 53, GCase 115], [α -syn 46, GCase 180] and [α -syn 50, GCase 261], since mutations of these α -syn residues are known Parkinson's disease risk factors.[49,50,66,67] In addition, the GCase residue of the number two ranked pair is associated with a Gaucher disease mutation, V78A.[27] In a correlated mutation study involving over 1000 human proteins, residues with known disease-associated mutations occurred in highly ranked intra-protein correlated pairs with a much higher frequency than expected by chance, especially at the protein surface.[68] The results here hint that this observation will likely extend to inter-protein correlated pairs too.

It could not be ruled out that the [α -syn 53, GCase 115] correlation might be a phylogenetic artifact, via genetic drift or perhaps due to evolutionary selection related to primate neural development or function, but not involving a direct α -syn interaction with GCase. Nevertheless, experiment has shown that α -syn 53 must be near GCase when membrane bound (Fig 5), [63] and GCase 115 lies adjacent to the second largest region of conserved GCase surface (Fig 4B), consistent with this region being involved in intermolecular interaction, so an interaction between these residues is plausible. In addition, the other α -syn mutant-associated correlations appear less likely to have arisen by chance. If any of these mutant-associated pairs are truly correlated, then the implications are two-fold. First, this would mean that α -syn and GCase have co-evolved to maintain a beneficial interaction that involves these pairs in some way. Second, it suggests that GCase might be directly involved in PD pathology caused by the alteration in interaction due to their mutation.

The set of residue pairs [α -syn 68, GCase 154] and [α -syn 95, GCase 154] also have possible implications regarding how the α -syn/GCase protein interaction might be related to PD etiology. Residues 68 and 95 of α -syn lie in the NAC region of α -syn (Fig 5), a portion of α -syn critical for the amyloid formation in PD.[23] GCase and α -syn interact in the lysosome, one of the cellular locations where α -syn is degraded, so interaction with the NAC region could affect how α -syn gets degraded. GCase mutations that interfere with this putative interaction or that result in less GCase reaching the lysosome could alter the accessibility of α -syn NAC region to lysosomal proteases, which could promote PD pathology. GCase also has an enzymatic cofactor, saposin C, that can compete with α -syn binding, rescuing α -syn-induced inhibition of the enzyme, thus GCase mutations that modify saposin C interaction could also impact α -syn interaction.[62,69]

The analysis of β -syn showed weaker correlation than α -syn with GCase, but also revealed a strong correlation between α -syn residue 53 and β -syn residues 86 and 121, the same α -syn residue in the top α -syn, GCase correlated pair. Colocalization of α -syn and β -syn is observed the presynaptic termini of neurons, and β -syn can mitigate α -syn induced toxicity and accumulation.[25] *In vitro* β -syn can interact with both α -syn and GCase, which might suggest some sort of competitive interaction.[25,63] However, there is no experimental evidence of any significant β -syn, GCase interaction *in vivo*. [31] The residue substitutions in β -syn residues 86 and 121 might have occurred earlier in primate evolution than those in α -syn 53, since the β -syn substitutions are seen in an additional primate species, marmoset (*C. jacchus*). The substitution for GCase residue 115 occurs for just Old World monkeys and apes, that is, one less primate species than the α -syn 53 substitution, so the GCase substitution could have occurred later. Thus, a causal chain is possible, with the β -syn 53 substitution leading to the α -syn 53 substitution, which then led to the GCase substitution. However, while it is conceivable that the α -syn A53T mutation could somehow alter interaction with β -syn, possibly impacting PD etiology, the genetic link between PD and β -syn is weak at best.[70]

Conclusion

The coevolutionary analysis of α -syn and GCase has shown that PD-associated alpha-synuclein mutations mirror highly correlated residue substitutions across vertebrate species, providing evidence linking altered α -syn/GCase interaction to PD pathology. Correlations between α -syn and β -syn hint that the same α -syn residue of the top ranked α -syn, GCase pair, α -syn 53, might also be involved in interactions with β -syn. The GCase residues of the two top ranked pairs lie adjacent one of the largest, highly conserved regions of the GCase surface, as one would expect for interacting proteins that have coevolved. While the predicted pairs are consistent with current experimental knowledge of α -syn/GCase interaction, more mutational and structural work is needed to verify whether the correlated pairs truly influence the complex *in vivo*. This study illustrates how sequences of non-human genomes can potentially help us understand human disease, but also highlights how more genome sequencing, especially of vertebrate genomes, is still needed for more robust analyses.

Supporting Information

S1 Table. Hemoglobin DI values.

(TXT)

S2 Table. Hemoglobin MI Z-values.

(TXT)

S3 Table. α -syn, GCase MI Z-values.

(TXT)

S4 Table. β -syn, GCase MI Z-values.

(TXT)

S5 Table. α -syn, β -syn MI Z-values.

(TXT)

S6 Table. α -syn, GCase DI values.

(TXT)

S1 Text. Multiple sequence alignments for hemoglobin, α -syn, GCase and β -syn.

(DOCX)

S2 Text. Appendix I. Sequence errors and the importance of the pseudocount correction.

(DOCX)

S3 Text. Appendix II. Probability estimate of a chance correlation.

(DOCX)

Acknowledgments

The author thanks Dr. Jennifer C. Lee and Dr. Ehud Goldin for their helpful suggestions.

Author Contributions

Conceived and designed the experiments: JMG. Performed the experiments: JMG. Analyzed the data: JMG. Contributed reagents/materials/analysis tools: JMG. Wrote the paper: JMG.

References

1. Ueda K, Fukushima H, Masliah E, Xia Y, Iwai A, Yoshimoto M, et al. (1993) Molecular cloning of cDNA encoding an unrecognized component of amyloid in Alzheimer disease. *Proc Natl Acad Sci U S A* 90: 11282–11286. PMID: [8248242](#)
2. Breydo L, Wu JW, Uversky VN (2012) Alpha-synuclein misfolding and Parkinson's disease. *Biochim Biophys Acta* 1822: 261–285. doi: [10.1016/j.bbadis.2011.10.002](#) PMID: [22024360](#)
3. Stefanis L (2012) alpha-Synuclein in Parkinson's disease. *Cold Spring Harb Perspect Med* 2: a009399. doi: [10.1101/cshperspect.a009399](#) PMID: [22355802](#)
4. Cookson MR (2012) Evolution of neurodegeneration. *Curr Biol* 22: R753–761. doi: [10.1016/j.cub.2012.07.008](#) PMID: [22975006](#)
5. Spillantini MG, Schmidt ML, Lee VM, Trojanowski JQ, Jakes R, Goedert M (1997) Alpha-synuclein in Lewy bodies. *Nature* 388: 839–840. PMID: [9278044](#)
6. Abeliovich A, Schmitz Y, Farinas I, Choi-Lundberg D, Ho WH, Castillo PE, et al. (2000) Mice lacking alpha-synuclein display functional deficits in the nigrostriatal dopamine system. *Neuron* 25: 239–252. PMID: [10707987](#)
7. Cabin DE, Shimazu K, Murphy D, Cole NB, Gottschalk W, McIlwain KL, et al. (2002) Synaptic vesicle depletion correlates with attenuated synaptic responses to prolonged repetitive stimulation in mice lacking alpha-synuclein. *J Neurosci* 22: 8797–8807. PMID: [12388586](#)
8. Nemani VM, Lu W, Berge V, Nakamura K, Onoa B, Lee MK, et al. (2010) Increased expression of alpha-synuclein reduces neurotransmitter release by inhibiting synaptic vesicle recluster after endocytosis. *Neuron* 65: 66–79. doi: [10.1016/j.neuron.2009.12.023](#) PMID: [20152114](#)
9. Bellani S, Sousa VL, Ronzitti G, Valtorta F, Meldolesi J, Chieregatti E (2010) The regulation of synaptic function by alpha-synuclein. *Commun Integr Biol* 3: 106–109. PMID: [20585500](#)
10. Vargas KJ, Makani S, Davis T, Westphal CH, Castillo PE, Chandra SS (2014) Synucleins regulate the kinetics of synaptic vesicle endocytosis. *J Neurosci* 34: 9364–9376. doi: [10.1523/JNEUROSCI.4787-13.2014](#) PMID: [25009269](#)
11. Burre J, Sharma M, Tsetsenis T, Buchman V, Etherton MR, Sudhof TC (2010) Alpha-synuclein promotes SNARE-complex assembly in vivo and in vitro. *Science* 329: 1663–1667. doi: [10.1126/science.1195227](#) PMID: [20798282](#)
12. Diao J, Burre J, Vivona S, Cipriano DJ, Sharma M, Kyoung M, et al. (2013) Native alpha-synuclein induces clustering of synaptic-vesicle mimics via binding to phospholipids and synaptobrevin-2/VAMP2. *Elife* 2: e00592. doi: [10.7554/eLife.00592](#) PMID: [23638301](#)
13. Yin G, Lopes da Fonseca T, Eisbach SE, Anduaga AM, Breda C, Orcellet ML, et al. (2014) alpha-Synuclein interacts with the switch region of Rab8a in a Ser129 phosphorylation-dependent manner. *Neurobiol Dis* 70: 149–161. doi: [10.1016/j.nbd.2014.06.018](#) PMID: [24983211](#)
14. Chen RH, Wislet-Gendebien S, Samuel F, Visanji NP, Zhang G, Marsilio D, et al. (2013) alpha-Synuclein membrane association is regulated by the Rab3a recycling machinery and presynaptic activity. *J Biol Chem* 288: 7438–7449. doi: [10.1074/jbc.M112.439497](#) PMID: [23344955](#)
15. Gruschus JM, Yap TL, Pistolesi S, Maltsev AS, Lee JC (2013) NMR structure of calmodulin complexed to an N-terminally acetylated alpha-synuclein peptide. *Biochemistry* 52: 3436–3445. doi: [10.1021/bi400199p](#) PMID: [23607618](#)
16. Lee FJ, Liu F, Pristupa ZB, Niznik HB (2001) Direct binding and functional coupling of alpha-synuclein to the dopamine transporters accelerate dopamine-induced apoptosis. *FASEB J* 15: 916–926. PMID: [11292651](#)
17. Choi P, Golts N, Snyder H, Chong M, Petrucelli L, Hardy J, et al. (2001) Co-association of parkin and alpha-synuclein. *Neuroreport* 12: 2839–2843. PMID: [11588587](#)
18. Ahn BH, Rhim H, Kim SY, Sung YM, Lee MY, Choi JY, et al. (2002) alpha-Synuclein interacts with phospholipase D isozymes and inhibits pervanadate-induced phospholipase D activation in human embryonic kidney-293 cells. *J Biol Chem* 277: 12334–12342. PMID: [11821392](#)
19. Neystat M, Rzhetskaya M, Kholodilov N, Burke RE (2002) Analysis of synphilin-1 and synuclein interactions by yeast two-hybrid beta-galactosidase liquid assay. *Neurosci Lett* 325: 119–123. PMID: [12044636](#)
20. Jensen PH, Hager H, Nielsen MS, Hojrup P, Gliemann J, Jakes R (1999) alpha-synuclein binds to Tau and stimulates the protein kinase A-catalyzed tau phosphorylation of serine residues 262 and 356. *J Biol Chem* 274: 25481–25489. PMID: [10464279](#)
21. Yap TL, Gruschus JM, Velayati A, Westbroek W, Goldin E, Moaven N, et al. (2011) Alpha-synuclein interacts with Glucocerebrosidase providing a molecular link between Parkinson and Gaucher diseases. *J Biol Chem* 286: 28080–28088. doi: [10.1074/jbc.M111.237859](#) PMID: [21653695](#)

22. Uemura N, Koike M, Ansai S, Kinoshita M, Ishikawa-Fujiwara T, Matsui H, et al. (2015) Viable Neurodegenerative Gaucher Disease Model in Medaka (*Oryzias latipes*) Displays Axonal Accumulation of Alpha-Synuclein. *PLoS Genet* 11: e1005065. doi: [10.1371/journal.pgen.1005065](https://doi.org/10.1371/journal.pgen.1005065) PMID: [25835295](https://pubmed.ncbi.nlm.nih.gov/25835295/)
23. El-Agnaf OM, Bodles AM, Guthrie DJ, Harriott P, Irvine GB (1998) The N-terminal region of non-A beta component of Alzheimer's disease amyloid is responsible for its tendency to assume beta-sheet and aggregate to form fibrils. *Eur J Biochem* 258: 157–163. PMID: [9851705](https://pubmed.ncbi.nlm.nih.gov/9851705/)
24. Vilar M, Chou HT, Luhrs T, Maji SK, Riek-Loher D, Verel R, et al. (2008) The fold of alpha-synuclein fibrils. *Proc Natl Acad Sci U S A* 105: 8637–8642. doi: [10.1073/pnas.0712179105](https://doi.org/10.1073/pnas.0712179105) PMID: [18550842](https://pubmed.ncbi.nlm.nih.gov/18550842/)
25. Hashimoto M, Rockenstein E, Mante M, Mallory M, Masliah E (2001) beta-Synuclein inhibits alpha-synuclein aggregation: a possible role as an anti-parkinsonian factor. *Neuron* 32: 213–223. PMID: [11683992](https://pubmed.ncbi.nlm.nih.gov/11683992/)
26. Klein C, Westenberger A (2012) Genetics of Parkinson's disease. *Cold Spring Harb Perspect Med* 2: a008888. doi: [10.1101/cshperspect.a008888](https://doi.org/10.1101/cshperspect.a008888) PMID: [22315721](https://pubmed.ncbi.nlm.nih.gov/22315721/)
27. Hruska KS, LaMarca ME, Scott CR, Sidransky E (2008) Gaucher disease: mutation and polymorphism spectrum in the glucocerebrosidase gene (GBA). *Hum Mutat* 29: 567–583. doi: [10.1002/humu.20676](https://doi.org/10.1002/humu.20676) PMID: [18338393](https://pubmed.ncbi.nlm.nih.gov/18338393/)
28. Sidransky E, Nalls MA, Aasly JO, Aharon-Peretz J, Annesi G, Barbosa ER, et al. (2009) Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. *N Engl J Med* 361: 1651–1661. doi: [10.1056/NEJMoa0901281](https://doi.org/10.1056/NEJMoa0901281) PMID: [19846850](https://pubmed.ncbi.nlm.nih.gov/19846850/)
29. Sardi SP, Clarke J, Kinnecom C, Tamsett TJ, Li L, Stanek LM, et al. (2011) CNS expression of glucocerebrosidase corrects alpha-synuclein pathology and memory in a mouse model of Gaucher-related synucleinopathy. *Proc Natl Acad Sci U S A* 108: 12101–12106. doi: [10.1073/pnas.1108197108](https://doi.org/10.1073/pnas.1108197108) PMID: [21730160](https://pubmed.ncbi.nlm.nih.gov/21730160/)
30. Sybertz E, Krainc D (2014) Development of targeted therapies for Parkinson's disease and related synucleinopathies. *J Lipid Res* 55: 1996–2003. doi: [10.1194/jlr.R047381](https://doi.org/10.1194/jlr.R047381) PMID: [24668939](https://pubmed.ncbi.nlm.nih.gov/24668939/)
31. Mak SK, McCormack AL, Manning-Bog AB, Cuervo AM, Di Monte DA (2010) Lysosomal degradation of alpha-synuclein in vivo. *J Biol Chem* 285: 13621–13629. doi: [10.1074/jbc.M109.074617](https://doi.org/10.1074/jbc.M109.074617) PMID: [20200163](https://pubmed.ncbi.nlm.nih.gov/20200163/)
32. de Juan D, Pazos F, Valencia A (2013) Emerging methods in protein co-evolution. *Nat Rev Genet* 14: 249–261. doi: [10.1038/nrg3414](https://doi.org/10.1038/nrg3414) PMID: [23458856](https://pubmed.ncbi.nlm.nih.gov/23458856/)
33. Marks DS, Colwell LJ, Sheridan R, Hopf TA, Pagnani A, Zecchina R, et al. (2011) Protein 3D structure computed from evolutionary sequence variation. *PLoS One* 6: e28766. doi: [10.1371/journal.pone.0028766](https://doi.org/10.1371/journal.pone.0028766) PMID: [22163331](https://pubmed.ncbi.nlm.nih.gov/22163331/)
34. Kosciolk T, Jones DT (2014) De novo structure prediction of globular proteins aided by sequence variation-derived contacts. *PLoS One* 9: e92197. doi: [10.1371/journal.pone.0092197](https://doi.org/10.1371/journal.pone.0092197) PMID: [24637808](https://pubmed.ncbi.nlm.nih.gov/24637808/)
35. Hopf TA, Scharfe CP, Rodrigues JP, Green AG, Kohlbacher O, Sander C, et al. (2014) Sequence co-evolution gives 3D contacts and structures of protein complexes. *Elife* 3.
36. Skerker JM, Perchuk BS, Siryaporn A, Lubin EA, Ashenberg O, Goulian M, et al. (2008) Rewiring the specificity of two-component signal transduction systems. *Cell* 133: 1043–1054. doi: [10.1016/j.cell.2008.04.040](https://doi.org/10.1016/j.cell.2008.04.040) PMID: [18555780](https://pubmed.ncbi.nlm.nih.gov/18555780/)
37. Weigt M, White RA, Szurmant H, Hoch JA, Hwa T (2009) Identification of direct residue contacts in protein-protein interaction by message passing. *Proceedings of the National Academy of Sciences of the United States of America* 106: 67–72. doi: [10.1073/pnas.0805923106](https://doi.org/10.1073/pnas.0805923106) PMID: [19116270](https://pubmed.ncbi.nlm.nih.gov/19116270/)
38. Ovchinnikov S, Kamisetty H, Baker D (2014) Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *Elife* 3.
39. Jeong CS, Kim D (2012) Reliable and robust detection of coevolving protein residues. *Protein Eng Des Sel* 25: 705–713. doi: [10.1093/protein/gzs081](https://doi.org/10.1093/protein/gzs081) PMID: [23077274](https://pubmed.ncbi.nlm.nih.gov/23077274/)
40. Pazos F, HelmerCitterich M, Ausiello G, Valencia A (1997) Correlated mutations contain information about protein-protein interaction. *Journal of Molecular Biology* 271: 511–523. PMID: [9281423](https://pubmed.ncbi.nlm.nih.gov/9281423/)
41. Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, et al. (2011) Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc Natl Acad Sci U S A* 108: E1293–1301. doi: [10.1073/pnas.1111471108](https://doi.org/10.1073/pnas.1111471108) PMID: [22106262](https://pubmed.ncbi.nlm.nih.gov/22106262/)
42. Brown CA, Brown KS (2010) Validation of coevolving residue algorithms via pipeline sensitivity analysis: ELSC and OMES and ZNMI, oh my! *PLoS One* 5: e10779. doi: [10.1371/journal.pone.0010779](https://doi.org/10.1371/journal.pone.0010779) PMID: [20531955](https://pubmed.ncbi.nlm.nih.gov/20531955/)
43. Ashenberg O, Laub MT (2013) Using analyses of amino acid coevolution to understand protein structure and function. *Methods Enzymol* 523: 191–212. doi: [10.1016/B978-0-12-394292-0.00009-6](https://doi.org/10.1016/B978-0-12-394292-0.00009-6) PMID: [23422431](https://pubmed.ncbi.nlm.nih.gov/23422431/)

44. Martin LC, Gloor GB, Dunn SD, Wahl LM (2005) Using information theory to search for co-evolving residues in proteins. *Bioinformatics* 21: 4116–4124. PMID: [16159918](#)
45. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7: 539. doi: [10.1038/msb.2011.75](#) PMID: [21988835](#)
46. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410. PMID: [2231712](#)
47. Di Lena P, Nagata K, Baldi P (2012) Deep architectures for protein contact map prediction. *Bioinformatics* 28: 2449–2457. PMID: [22847931](#)
48. Halabi N, Rivoire O, Leibler S, Ranganathan R (2009) Protein sectors: evolutionary units of three-dimensional structure. *Cell* 138: 774–786. doi: [10.1016/j.cell.2009.07.038](#) PMID: [19703402](#)
49. Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, et al. (1997) Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science* 276: 2045–2047. PMID: [9197268](#)
50. Pasanen P, Myllykangas L, Siitonen M, Raunio A, Kaakkola S, Lyytinen J, et al. (2014) A novel alpha-synuclein mutation A53E associated with atypical multiple system atrophy and Parkinson's disease-type pathology. *Neurobiol Aging* 35: 2180 e2181–2185.
51. Dvir H, Harel M, McCarthy AA, Toker L, Silman I, Futerman AH, et al. (2003) X-ray structure of human acid-beta-glucosidase, the defective enzyme in Gaucher disease. *EMBO Rep* 4: 704–709. PMID: [12792654](#)
52. Knaggs MH, Salsbury FR, Edgell MH, Fetrow JS (2007) Insights into correlated motions and long-range interactions in CheY derived from molecular dynamics simulations. *Biophysical Journal* 92: 2062–2079. PMID: [17172298](#)
53. Noivirt-Brik O, Unger R, Horovitz A (2009) Analysing the origin of long-range interactions in proteins using lattice models. *Bmc Structural Biology* 9.
54. Dunn SD, Wahl LM, Gloor GB (2008) Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. *Bioinformatics* 24: 333–340. PMID: [18057019](#)
55. Ashkenazy H, Kliger Y (2010) Reducing phylogenetic bias in correlated mutation analysis. *Protein Eng Des Sel* 23: 321–326. doi: [10.1093/protein/gzp078](#) PMID: [20067922](#)
56. Lee BC, Kim D (2009) A new method for revealing correlated mutations under the structural and functional constraints in proteins. *Bioinformatics* 25: 2506–2513. doi: [10.1093/bioinformatics/btp455](#) PMID: [19628501](#)
57. Wollenberg KR, Atchley WR (2000) Separation of phylogenetic and functional associations in biological sequences by using the parametric bootstrap. *Proc Natl Acad Sci U S A* 97: 3288–3291. PMID: [10725404](#)
58. Fares MA, Travers SA (2006) A novel method for detecting intramolecular coevolution: adding a further dimension to selective constraints analyses. *Genetics* 173: 9–23. PMID: [16547113](#)
59. Beutler E, Gelbart T, Balicki D, Demina A, Adusumalli J, Elsas L 2nd, et al. (1996) Gaucher disease: four families with previously undescribed mutations. *Proc Assoc Am Physicians* 108: 179–184. PMID: [8774051](#)
60. Caffrey DR, Somaroo S, Hughes JD, Mintseris J, Huang ES (2004) Are protein-protein interfaces more conserved in sequence than the rest of the protein surface? *Protein Sci* 13: 190–202. PMID: [14691234](#)
61. Cho MK, Nodet G, Kim HY, Jensen MR, Bernado P, Fernandez CO, et al. (2009) Structural characterization of alpha-synuclein in an aggregation prone state. *Protein Sci* 18: 1840–1846. doi: [10.1002/pro.194](#) PMID: [19554627](#)
62. Gruschus JM, Jiang Z, Yap TL, Hill SA, Grishaev A, Piszczek G, et al. (2015) Dissociation of glucocerebrosidase dimer in solution by its co-factor, saposin C. *Biochem Biophys Res Commun* 457: 561–566. doi: [10.1016/j.bbrc.2015.01.024](#) PMID: [25600808](#)
63. Yap TL, Velayati A, Sidransky E, Lee JC (2013) Membrane-bound alpha-synuclein interacts with glucocerebrosidase and inhibits enzyme activity. *Mol Genet Metab* 108: 56–64. doi: [10.1016/j.ymgme.2012.11.010](#) PMID: [23266198](#)
64. Pele J, Moreau M, Abdi H, Rodien P, Castel H, Chabbert M (2014) Comparative analysis of sequence covariation methods to mine evolutionary hubs: examples from selected GPCR families. *Proteins* 82: 2141–2156. doi: [10.1002/prot.24570](#) PMID: [24677372](#)
65. Wei RR, Hughes H, Boucher S, Bird JJ, Guziewicz N, Van Patten SM, et al. (2011) X-ray and biochemical analysis of N370S mutant human acid beta-glucosidase. *J Biol Chem* 286: 299–308. doi: [10.1074/jbc.M110.150433](#) PMID: [20980263](#)

66. Zarranz JJ, Alegre J, Gomez-Esteban JC, Lezcano E, Ros R, Ampuero I, et al. (2004) The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia. *Ann Neurol* 55: 164–173. PMID: [14755719](#)
67. Appel-Cresswell S, Vilarino-Guell C, Encarnacion M, Sherman H, Yu I, Shah B, et al. (2013) Alpha-synuclein p.H50Q, a novel pathogenic mutation for Parkinson's disease. *Mov Disord* 28: 811–813. doi: [10.1002/mds.25421](#) PMID: [23457019](#)
68. Kowarsch A, Fuchs A, Frishman D, Pagel P (2010) Correlated Mutations: A Hallmark of Phenotypic Amino Acid Substitutions. *Plos Computational Biology* 6.
69. Yap TL, Gruschus JM, Velayati A, Sidransky E, Lee JC (2013) Saposin C protects glucocerebrosidase against alpha-synuclein inhibition. *Biochemistry* 52: 7161–7163. doi: [10.1021/bi401191v](#) PMID: [24070323](#)
70. Brighina L, Okubadejo NU, Schneider NK, Lesnick TG, de Andrade M, Cunningham JM, et al. (2007) Beta-synuclein gene variants and Parkinson's disease: a preliminary case-control study. *Neurosci Lett* 420: 229–234. PMID: [17556099](#)